Lecture Notes
in Geoinformation and Cartography

LNG&C

Vasily Popovich
Manfred Schrenk
Jean-Claude Thill
Christophe Claramunt
Tianzhen Wang *Editors*

Information Fusion and Intelligent Geographic Information Systems (IF&IGIS'17)

New Frontiers in Information Fusion and Intelligent GIS: From Maritime to Land-based Research

Springer

Lecture Notes in Geoinformation and Cartography

The Lecture Notes in Geoinformation and Cartography series provides a contemporary view of current research and developments in Geoinformation and Cartography, including GIS and Geographic Information Science. Publications with associated electronic media examine areas of development and current technology. Editors from multiple continents, in association with national and international organizations and societies, bring together the most comprehensive forum for Geoinforma-tion and Cartography.

The scope of Lecture Notes in Geoinformation and Cartography spans the range of interdisciplinary topics in a variety of research and application fields. The type of material published traditionally includes:

- proceedings that are peer-reviewed and published in association with a conference;
- post-proceedings consisting of thoroughly revised final papers; and
- research monographs that may be based on individual research projects.

The Lecture Notes in Geoinformation and Cartography series also include various other publications, including:

- tutorials or collections of lectures for advanced courses;
- contemporary surveys that offer an objective summary of a current topic of interest; and
- emerging areas of research directed at a broad community of practitioners.

More information about this series at http://www.springer.com/series/7418

Vasily Popovich · Manfred Schrenk
Jean-Claude Thill · Christophe Claramunt
Tianzhen Wang
Editors

Information Fusion and Intelligent Geographic Information Systems (IF&IGIS'17)

New Frontiers in Information Fusion and Intelligent GIS: From Maritime to Land-based Research

Springer

Editors
Vasily Popovich
SPIIRAS Hi Tech Research
and Development Office Ltd.
St. Petersburg
Russia

Manfred Schrenk
Department for Urbanism, Transport,
Environment and Information Society
Central European Institute
of Technology—CEIT ALANOVA
gemeinnützige GmbH
Schwechat
Austria

Jean-Claude Thill
Department of Geography and Earth
Sciences
The University of North Carolina
Charlotte, NC
USA

Christophe Claramunt
Naval Academy Research Institute
Brest Naval, Finistère
France

Tianzhen Wang
Shanghai Maritime University
Shanghai
China

ISSN 1863-2246 ISSN 1863-2351 (electronic)
Lecture Notes in Geoinformation and Cartography
ISBN 978-3-319-59538-2 ISBN 978-3-319-59539-9 (eBook)
DOI 10.1007/978-3-319-59539-9

Library of Congress Control Number: 2017941069

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This book contains a number of original scientific papers that were selected after a peer-review process for presentation at the 8th International Symposium "Information Fusion and Intelligent Geographical Information Systems (IF&IGIS'2017)." The symposium was held from 10 to 12 May 2017 in Shanghai, China, with a specific focus this year on solving vital issues at the intersection of geoinformational and maritime research fields. This symposium was organized by the SPIIRAS Hi Tech Research and Development Office Ltd, St. Petersburg, Russia, with support of the Shanghai Maritime University, China.

The main goal of the IF&IGIS'2017 symposium was to bring together leading world experts in the field of spatial information integration and intelligent geographical information systems (IGIS) to exchange cutting-edge research ideas and experiences, to discuss perspectives on the fast-paced development of geospatial information theory, methods, and models, to demonstrate the latest achievements in IGIS, and for applying these research concepts to real-world use cases. The full papers, selected by the International Program Committee of IF&IGIS'2017, address fundamentals, models, technologies, and services of IGIS in the geoinformational and maritime research fields including underwater acoustics, radio-location, maritime navigation safety, marine energy, logistics, environmental management, as well as other modeling and data-driven matters critical to the effectiveness of information-fusion processes and intelligent geographical information systems.

The paper-submission process of the symposium attracted 44 abstracts from 20 countries; 26 papers were selected at the first step of a blind-review process for presentation at the conference. After the second step of the review process, the Program Committee accepted 19 full papers contributed by authors from 10 countries for presentation and publication in this book. In accordance with subjects of the accepted papers, 5 parts of the book were formed: (1) data modeling, integration, fusion, and analysis in IGIS; (2) maritime traffic-control methods; (3) IGIS integration with acoustic, remote-sensing, and radar systems; (4) ports and maritime transportation and logistics; and (5) IGIS for land-based research. Special guests of the symposium were two invited speakers this year who provided us with high-profile lectures on information fusion and intelligent geographical information

systems: Prof. George Vouros from the Department of Digital Systems, University of Piraeus, Greece, and Prof. Bo Huang from the Department of Geography and Resource Management, The Chinese University of Hong Kong.

The success of the symposium is, undoubtedly, a result of the combined and dedicated efforts of sponsors, organizers, reviewers, and participants. We are also grateful to several colleagues from the Shanghai Maritime University that have been very helpful in organizing the whole event, special thanks to Prof. Yongxing Jin, Xiong Hu, Qinyou Hu, Tianhao Tang and Huafeng Wu. We acknowledge the Program Committee members for their help with the review process. Our thanks go to all participants and authors of the submitted papers as well. We are also very grateful to our sponsors—International Maritime Lecturers Association (IMLA), Computer Applications Committee of CSNAME, Key Laboratory of Geographic Information Science Ministry of Education, Shanghai Power Supply Society, and Shanghai Federation of Pudong R&D Institutions—for their generous support. Finally, we also extend our deep gratitude to Springer's publishing team, managed by Dr. Christian Witschel and Anand Manimudi Chozhan, for their professionalism, help, and focus on the success of this collaborative effort.

St. Petersburg, Russia — Vasily Popovich
Schwechat, Austria — Manfred Schrenk
Charlotte, USA — Jean-Claude Thill
Brest, France — Christophe Claramunt
Shanghai, China — Tianzhen Wang
May 2017

Contents

Abbreviations

AI	Artificial intelligence
AIS	Automatic identification system
ATM	Air traffic management
CEM	Comprehensive emergency management
CHI	Carville hurricane index
DB	Data base
DSS	Decision support system
DTN	Delay-tolerant network
ECDIS	Electronic chart display and information system
ENC	Electronic nautical chart
FCD	Floating car data
FCFS	First-Come, First-Served
FDI	Fire danger index
FDR	Fire danger rating
G-BEP	GIS-based evacuation planning
GIS	Geographic information system
GPS	Global positioning system
HAC	Hydrological and acoustic conditions
IALA	International association of lighthouse authorities
IE	Inference engine
IGIS	Intelligent geographical information system
IHM	infinite homogeneous medium
IMO	International Maritime Organization
ITS	Intelligent transport system
JDL	Joint Directors of Laboratories
KB	Knowledge base
MAUP	Modifiable areal unit problem
MOGA	Multi-objective genetic algorithm
MSA	Maritime situational awareness
MSI	Maritime safety information

MSP	Marine service portfolio
MSP	Maritime service portfolios
MSRS	Maritime surveillance and recognition systems
NASA	National Aeronautics and Space Administration
OOS	Operational oceanology system
OSM	Open Street Map
PCD	Probability of correct detection
PIERS	Port import export reporting service
POI	Places of interest
PVSF	Port vessel schedule in fairway
RIAM	Randomised aggregated indices method
ROI	Regions of interest
SNA	Social network analysis
SNR	Signal-to-noise ratio
SOM	Self-organizing map
ST-PF	Spatio-temporal particle filter
TAIS	Tourist attraction information service
TEUs	Twenty-foot-equivalent units
TS	Tactical situation
UAV	Unmanned aerial vehicle
UGCoP	Uncertain geographic context problem
UI	User's interface
USV	Unmanned surface vessel
VGI	Volunteered geographic information
VSPV	Vertical sound propagation velocity
VTS	Vessel traffic service
WKT	Well known text

Part I
Data Modelling, Integration, Fusion, and Analysis in Intelligent GIS

Space Theory for Intelligent GIS

Vasily Popovich

Abstract The major purpose of this paper is to discuss two phenomena: (1) geographic information systems (GIS) and (2) space. The discussion is from a computer science point-of-view. As it is well known, every point in GIS can be presented in 1D, 2D, 3D or nD dimension. If we take a look on linear algebra, we can find definitions of "space" and "subspace." Therefore, the major point of our discussion is to determine space and subspace in GIS and to introduce some measures for dynamic space borders and other properties that can be calculated by different methods. The complexity of the proposed paper arrives from an original complexity of space definition and because of the fact that it is used in abstract algebra, philosophy, and GIS paradigms together.

Keywords GIS space theory · Space and subspace for GIS · Dynamic space

1 Introduction

In everyday awareness, the term "space" is one of the most mysterious. In philosophy, we can find many definitions of this term whilst it is regarded inseparably from the terms "time" and "existence." In philosophy and some classical mathematics sources, we can find that

> . . . the metaphysician Immanuel Kant said that the concepts of space and time are not empirical ones derived from experiences of the outside world—they are elements of an already given systematic framework that humans possess and use to structure all experiences. Kant referred to the experience of "space" in his *Critique of Pure Reason* [1] as being a subjective "pure *a priori* form of intuition". According to Kant, knowledge about space is *synthetic*, in that statements about space are not simply true by virtue of the meaning of the words in the statement. In his work, Kant rejected the view that space must be either a substance or relation. Instead he came to the conclusion that space and time are not discovered by humans to be objective features of the world, but imposed by us as part of a framework for organizing experience.

V. Popovich (✉)
SPIIRAS-HTR&DO Ltd, #No. 39, 14 Line, St. Petersburg, Russia
e-mail: popovich@oogis.ru

V. Popovich et al. (eds.), *Information Fusion and Intelligent Geographic Information Systems (IF&IGIS'17)*, Lecture Notes in Geoinformation and Cartography, DOI 10.1007/978-3-319-59539-9_1

Modern scientists usually fall under two opposing groups: overt opponents and adherers of Kant's idea. While not being experts in a field of philosophical inquiry, we however can note that we are closer to Kant's idea since in computer science, and especially in geoinformatics, this idea becomes a very good ideological basis for the design of actual technology and applications. Jean-Claude Thill [1] proposed very important idea that geographic space and semantic information associated with events and features are fully separable. For a common case, space exists independently of events and experiments. Taking in mind philosophical aspects of space, we should also note that the GIS has strong applied sense. According to this, two ideas should be investigated together: absolute space and relative space [1]. Following Jean-Claude Thill, we will study relative space in this paper. Considering such complex history of this term and its current ambiguity, we should address its mathematical definition. All the more so because in our field of study, i.e., the field of geoinformatics, anthropological nuances do not aid but rather obscure this truly important matter.

In [2] and other mathematical sources, the mathematical term of "space" can be defined as "[a] mathematical set that possesses structure defined by axiomatic properties of its elements (e.g., points in geometry, vectors in linear algebra, events in probability theory, etc.). Subset of space is called "subspace" if space structure initialises the structure of same type on this subset (the exact definition depends on space type)."

The term "space" for mathematics turned out to be extremely useful. Let us give a partial list of various types of space in mathematics:

- **Affine space** is a space that generalizes the properties of Euclidean spaces. It is mostly similar to a vector space; however, affine space is distinctive by the fact that all its points are equal (in particular, the concept of zero point is not defined in it).
- **Banach space** is a complete normed vector space.
- **Probability space** is a concept introduced by A. N. Kolmogorov in 1930s in order to formalise the concept of probability that originated the rapid development of probability theory as a strict mathematical discipline.
- **Hilbert space** is a generalisation of Euclidean space that allows an infinite number of dimensions.
- **Euclidean space**, in initial terms, is a space which properties are described by Euclidean geometry axioms. In this case, it is assumed that the space is three-dimensional. In modern understanding, in a more general sense, it can denote one of closely related objects: finite-dimensional real vector space with positively defined scalar product or metric space corresponding to such a vector space.
- **Normed space** is a vector space on which a norm is defined.
- **Vector space** is a mathematical structure that represents a set of elements (points), called "vectors," for which the operations of addition and multiplication by number (scalar) are defined. These operations are defined by eight axioms.

- **Metric space** is a set for which distance, with certain properties (metric) between any pair of elements, is defined.
- **Minkowski space** is a four-dimensional pseudo-Euclidean space with signature that has been suggested as an interpretation of space–time in the special theory of relativity. For every event, there is a corresponding point in Minkowski space in Galilean coordinates, three coordinates of which are Cartesian coordinates of three-dimensional Euclidean space, and the fourth is ct coordinate, where c denotes the speed of light, and t is time of event.
- **Topological space** is a set with an additional structure of a certain type (so called "topology"). It is one of the objects of study for one of branches of mathematics also called "topology." Historically, the term "topological space" originated as a generalisation of metric space.
- L^p Spaces (Lebesgue spaces) are spaces of measurable functions for which their p-th power is integrable for $p \geq 1$.

The classification of mathematical term "space" given above has formed historically, and it reflects the level of generalisation of fundamental concepts of "point," "measure," and some other qualities. A question emerges: Can GIS space belong to one of the listed types, or to a number of them? Of course, it can. For example, it can belong to a common for ordinary person Euclidean space, topological space (navigational format C57), and some others.

At the same time, we should note that, at the present time such traditional notions like "space," "set," and "point" are failing to satisfy not only theorists but practitioners as well. These contradictions can be clearly represented in GIS. The notion of "point" is so abstract that it becomes increasingly more difficult to create applicable, along with theoretical, interpretation and to refer it to one of the types of mathematical spaces described above. The case is that it is physically impossible to denote the term "point" in a traditional sense (as some abstraction indicating coordinates). We always deal with a certain multidimensional, at least with a dimension of three, neighbourhood of some point that cannot always be denoted as a centre of some local coordinate system. In addition, even if we do so, there is no positive gain from this abstraction. An attempt to solve practical tasks in GIS leads to the piling of large abstractions and relations between them. This complicates the development, support, and operating of such GIS. It is no coincidence that in Java programming language there are no simple abstract data types like point, line, etc., as there were in preceding languages. They used to be practically identical to the notion of "point" in Euclidean space. In Java, we initially have a notion of "object," and it is very right. In other words, we have a defined set with given structure. The term "object" cannot be referred to the term "set" in algebra. The latter is a new concept more closely related to the concept of "category" [3]. Therefore, the term "point" cannot be constricted to just some coordinate vector. Moreover, consequently, the term cannot be referred to any of the known abstract data types apart from "object." In this context, we have no simple analogue for known mathematical definitions of space.

At this rate, the question arises: what is "space" for GIS? If it is a point, the basis of practically any space, then, after generalisation of "point" concept, we have to

define what "space" denotes in GIS. The complexity of analysis of this concept also arises from the fact that GIS is simultaneously a theory, a technology, and a practice. At that, theory, technology, and practical application are closely connected and often change places in time in unnatural order (compared to the traditional concept of fundamental science, application-oriented technologies, and practice). In computer sciences, it has been long noted that application-oriented researches and technologies frequently outpace theoretical and fundamental ones. We can assume that GIS space is defined by such formats or sets of numbers that are needed for business analytics realisation (for user's convenience). Alternatively, this space is defined by such space (usually a mathematical concept) that represents GIS business analytics. In scientific work [4], it has been shown that GIS is specified on a multidimensional space. At that, all mathematical space paradigms turn out to be either trivial or useless due to interpretation complexity. However, we should clearly understand what we are dealing with and what interpretation capabilities we possess. Without a formal definition of GIS, success in the design and application of GIS is unlikely. Space is a foundation of all model systems specified on this space. Models are building blocks for application tasks and business logic that are created for simplified representation, study, and analysis of objective and/or abstract reality. Certainly, when GIS are applied for relatively simple and traditional tasks, like cartography and local monitoring systems, this problem is not so evident. Yet, when we attempt to design, produce, and then to support global monitoring and decision support systems based on GIS, the situation clearly spirals out of our control, and the system begins to live its own life, incomprehensible to both developers and users of this system.

If we turn our attention to such field of study as "data fusion in GIS," which can be regarded as one of the variants of theoretical justification of global monitoring system design methodology, it is easy to note that we are dealing with at least six global spaces (further development of the JDL model [5] idea can be found in work [4]). Considering the fact that according to Kant's hypothesis, such concepts as space and time are artificial, GIS has no objective foundations. The most objective data are observation data, yet they need to be interpreted, and herein lies the problem. That said, let us make a very important deduction that all types of spaces in GIS are artificial. In a practical sense, this means that we ought to have a nearly constant external data flow, preferably a flow of measurement data. We also have to understand that we have a very critical parameter: time of measurements or time of data reception. With the passage of time, some data lose their actuality or even meaning. Therefore, such terms as "time" and "space" are determinative in any GIS and in any system based on a GIS platform.

2 Space Definition

Having determined that such categories as "time" and "space" are key concepts in GIS, we still must give them a definition. The concept of time is universal for nearly all fields of study and is unlikely to have any particularities in GIS. Exclusively, we

can note that there can be several relative time scales. For example, while modelling, we can artificially slow down or speed up time and do time jumps forwards or backwards in time along the time line. Space is a different story. Considering the fact that, as stated above, we have at least six types of major spaces (scientific fields), but in truth there are much more of them, and we need to regard them separately.

As noted in the introduction, any recent graduate and even user understands that, the most simple and at the same time, the most general abstraction is a point. On its own, a point is a primitive concept with a sufficiently vast set of properties. Yet, the concept of point is not independent. Without definition of space, the concept of point is meaningless and vice versa. Considering the specificity of our field of study, GIS, the point is not simply a mathematical notion; but firstly it is both a coordinate but not only a coordinate. Depending on the context, the point has a whole set of properties and sometimes methods (functions).

Let us regard the concept of point for various situations from the point of view of mundane consciousness:

1. One-dimensional case (point): the point has one coordinate and a number of other parameters.
2. Two-dimensional case (Euclidean space): the point has two coordinates and a number of other parameters.
3. Three-dimensional case: the point has three coordinates and a number of other parameters.
4. Multidimensional case: the point has a number of ordinates plus a number of other parameters.
5. All previous cases plus time: the point has a time parameter added.

The point is an initial concept from which all other abstracts found in GIS can be formed, e.g., they are derivative. On the other hand, the abstract concept of point should differ. Turning back to the JDL model, every level has its own specified space and analytics. Regardless of the fact that we apply GIS for specific purposes and for one, in a particular case, subject area, these spaces should not overlap, else we will obtain a whole system of contradictions.

Let us make one small remark. We cannot build a GIS space system using an axiomatic approach [4]. It means that we cannot formulate a universal set of constraints and assumptions for all GIS space systems. It is more than likely that we should use an evolutionary approach, and perhaps in the future it will be possible to design an axiomatic system (theory). In this case, let us systematically consider a system of typical spaces using the JDL model as a basis:

1. the zero or the first level is a space of measurements and/or different signals in one or various environments (atmosphere, ocean);
2. the second level is a space of targets and objects, a set of tracks;
3. the third level is a space of tactical situation (TS) (a tactical situation is a time sample of state for several spaces: environment, manageable resources, obstacles and a goal to be reached; as a rule, TS space is formed beforehand, but it can be created dynamically);

4. the fourth level is a space of dangers and/or threats with a quantitative evaluation;
5. the fifth level is a space of resources at our disposal;
6. the space of solutions on all previously mentioned spaces.

We have a very complex combination of spaces, rather closely interrelated, that however have a principal difference in the fundamental concept of "space point." Shortly, these abstract points can be defined as a measurement (signal), an object (physical or abstract), a tactical situation, a threat, a resource, and a solution. For every space, the measure should be defined as a mean of specifying analytics on space in favour of practical and/or abstract task solving.

3 Measure Definition for Space-Type Determination

In a philosophical sense, a measure is a philosophical category denoting a unity of the qualitative and quantitative qualities of some object. According to Ogurtsov [6], this category generalises the means and results of measuring objects. Measure analysis derives from the importance of variation intervals of quantitative values, in terms of which we can talk of the object's quality preservation. The measure category is closely related to a number of philosophic notions along with those falling into the fields of ethics and aesthetics. In mathematics, "measure" is a common term for different types of generalisation of notions of Euclidean length, area, and n-dimensional volume. There are various specifications to the notion of measure:

- **Jordan measure** is an example of a finitely additive measure or one of the ways of formalising notions of length, area, and n-dimensional volume in Euclidean space.
- **Lebesgue measure** is an example of denumerable additive measure, and it is a continuation of Jordan measure on a more vast class of sets.
- **Riemann measure** (Riemann integral) is an area of region under a curve (a figure between a graph of function and the abscissa).
- **Hausdorff measure** is a special mathematical measure.

Regarding Hausdorff measure, the necessity of the introduction of such a measure derived from the need to calculate the length, area, and volume of nonspecific figures that can not be specified analytically. The application of such a generalisation in GIS reveals new opportunities for the realisation of different kinds of business logics in complex, multidimensional, and implicitly specified spaces. In particular, this measure can be used to calculate the volume of an acoustic field. This is practically impossible, or at least vey difficult, to do analytically (Fig. 1). Approximately the same level of difficulty is common for the task of calculating the volume of available networks (3G, 4G, Wi-Fi, etc.) outdoors for a particular user, or

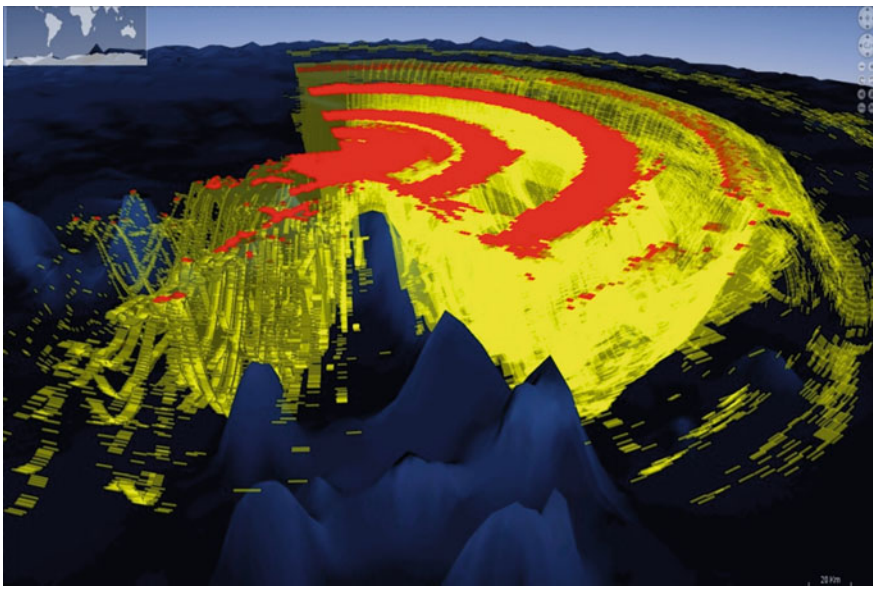


Fig. 1 3-D Model of acoustic sound field propagation in the ocean

for a given type of users, taking into account the complex surrounding space (buildings, metal fences, etc.). The numerical values of such space can be calculated directly in the process of field-task solving (by specifying the Hausdorff measure and step-by-step calculation) or by using an imitation modelling method when the field is already given.

4 Detection Common Problem for Space

The task of detection in GIS holds the same fundamental meaning as the concept of point. If the point is the basis, the building brick on which the space abstracts (line, polyline, and their derivatives) are formed, then the detection task is the basis of nearly all business analytics in GIS. The concept of "detection" has a number of synonyms such as "identification," "classification," etc. and depending on the field of study, one notion can be used, or a whole number of notions can be applied for the core clarification of the detection concept. For example, in acoustics and radio location, the following notions can be used: signal detection, object detection, and object classification. In pattern recognition, a slightly different system of notions, in one way or another related to detection, is used. In literature, there is no abstract concept of "detection." As a rule, it is already connected to a specific process (signal detection, object detection, detection of entries in database, detection of a word [phrase] in text, etc.). In a well-known search of moving objects theory,

detection is a registration of physical fields of the searched object by a detection system. Classification is an attribution of registered physical fields to the known object class (specified beforehand). Tracking is a registration of the coordinates of the source of physical fields in time. Other notions of higher abstract class are also used: track, tactical situation, etc. The concept of detection is a metric concept of space of the first level. Detection takes place when some searched object (its physical field) enters the interval or area of detection of some detection system. For most practical cases, we can be satisfied with Riemann measure. For cases of line or plane, business analytics can base itself on existing detection theory [7]. For three or more dimensional spaces, there is no such ready-made theory. For the case in Fig. 1, no abstract detection process is applicable. Detection field physics, shown in Fig. 1, is common to almost any physical detection in nature. Approximation in the form of a circle, a sector, or a sphere is unacceptable because it distorts the physical sense. Hence, for the case given in Fig. 1, it is impossible to form the classical workable and probabilistic characteristics of a detector because the detection is closely linked to the spatial channel in which it occurs. In addition, there can be quite a lot of those channels, and they can significantly differ from one another. A classical case of detection can happen only in degenerate cases and rarely occurs in practice. Given that, we should note that the dominating majority of detectors these days is operating based on the principal of "averaged patients' tests in hospital." This is what modern statistical detection theory and optimal signal processing are in essence.

5 Search Common Problem for Space

"Search" is an active process, the result of which (with certain probability) can be a detection of a searched object or of an object (set of objects) of a given class. Objects can be very different, e.g., a point on map or on a screen, an entry in a database, a file on a hard-drive a fragment of information, or physical objects such as cars, plains, ships, satellites, people, groups of people, etc. An attempt to create a unified search theory has turned out to be a failure, although just after the Second World War the progress of operations research groups from England and the USA looked promising. As of today, unfortunately, we can discuss only a particular search in a particular subject area whether physical (ocean, atmosphere, cosmos, etc.) or abstract such as the Internet, computer, RAM, hard-drive, etc.

A classical example of a search in space is the well-known search theory [8] or search theory for moving objects [7]. The main theorem of search is defined on a bounded space L_1 with specified Riemann measure. As a search task, we will consider those that include four objects:

1. A search system is a mean of detection, a monitoring system, before which the task of detecting some physical or abstract object (entity) according to one or several parameters (characteristics, methods) is set.

2. A search target is a physical or abstract object that needs to be detected according to one or several parameters (characteristics, methods).
3. Beforehand given space (known) on characteristics of which search target parameters have influence (can be identified).
4. The search operation is an aggregation of actions correlated in time and space or an aggregation of search systems used to optimise the process of target search under given spacial conditions at a given point in time (it is assumed that space can change its parameters in relation to a specified set of coordinates and time).

According to new concepts of space and point, search is a verification of affiliation of one or many points of space to a searched object. In other words, search is an oriented process in space and time. Orientation is an algorithm or a scenario according to which sorting is performed. This algorithm can be specified by a particular relation or a function on space. The search process is estimated by given criteria and efficiency factors tied to a specified measure. In practical cases, in the search for physical objects of a different nature, the following criteria are used: minimal search time, maximal number of detected objects, maximal detection probability, etc. Efficiency factors can be detected objects mean, time detection mean, etc.

In some cases, when a searched object is deliberately hiding or resisting detection, counteractions to goal achievement and other factors should be included in the operation.

The following relations and/or notions for search operation execution are specified on the search space:

1. The goal of operation is a searched object (physical or abstract) that should be found on a given space class.
2. Aggregation of efforts (resources) of operation is, as a rule, the means and systems of search and detection.
3. Algorithm or scenario of search and resources management in a search operation.
4. Estimation of operation efficiency according to given criteria and factors.
5. Result analysis tools and operation optimisation methods in case the operation is recurring or cycled.

6 Tactical Situation Understanding

Under the term "**tactical situation**," we understand a co-relation and a state of efforts (resources) directed to searching, of searched object (objects), and of environmental parameters on a given (usually but not necessarily bounded) subset of points of a given space at a given time. In a practical sense, apart from georeferencing (area coordinates, target's and observer's coordinates) and apart from all tactical situation components (course, speed, bearing, bearing variation,

environmental field characteristics, etc.), we should perform classification of the tactical situation (TS) itself. The reason for this is that besides knowing the quantitative parameters and georeference, it is important to estimate what is happening in a given area at a given time and what can happen in the next instant of time. On the other hand keeping in mind qualitative or simulation analysis, we should anchor TS to given tools for its analysis and decision-making.

Therefore, the task of first priority is selection of the basis of classification of TS and the execution of classification. Selection of the basis of classification of TS allows to obtain several classes of TS to which a particular TS can be attributed. As a result, we have another meta-space of tactical situations.

In practice, the formation of initial basic TS can be performed in several ways:

- by using experts;
- by applying some kind of theoretical apparatus;
- by choosing some basis of classification.

That said, we need to consider that the existence of basic TS implies existence of its adequate interpretation or solution of this situation, seeing that otherwise, in practical cases, it loses meaning.

7 Conclusion

The introduction of an abstract concept of "space" in theoretical GIS research derives from a number of factors that recently appeared as a result of intensive research and development of GIS theory itself and, in particular, different GIS applications. Currently it is easier to list those subject areas where GIS are not applied then to list the whole spectrum of research and practice where GIS are successfully applied. By analogy, as once happened with operation research methods [7], we cannot talk of some unified GIS theory because there are many different theoretical researches and technologies. At the same time, theorists and practitioners should understand what they are dealing with. From our point of view, the concept of "space" is suitable enough for a generalised view of GIS as well as of applicable business analytics realised in GIS. Through this concept, many GIS classes and their applications can be theoretically sorted and specified. It especially concerns big and distributed systems and big data problems related to them. Both the user and developer should understand what they are dealing with.

References

1. Kant I (1781) Critique of pure reason (*Kritik der reinen Vernunft*) (trans. Smith NK). St. Martins, 1965, New York, pp A 51/B 75
2. Shraer O, Shperner G Introduction to linear algebra in geometrical interpretation (*Einfuhrung in die analytische geometrie und algebra*) (trans. Olshansky G). M–L: ONTI, 1934, pp 210

3. MacLane S (1971) Categories for the working mathematician. Springer
4. Intelligent GIS (2013). In V. Popovich (ed). Moscow, Nauka, pp 324 (in Russian)
5. Blasch E (2002) Fundamentals of information fusion and applications. Tutorial, TD2, Fusion
6. Ogurtsov AP (1988) Disciplinary Structure of Science. Genesis and Argumentation, Moscow (in Russian)
7. Detection and Search Theory for Moving Objects (2016). In Popovich V. (ed), Moscow, Nauka, pp 423 (in Russian)
8. Koopman BO (1956) Theory of search: 2. Target detection. Operations Research. Vol 4, No 5
9. Eilenberg S, MacLane S (1945) A general theory of natural equivalences. Trans. Amer. Math. Soc. 58:231–294

Taming Big Maritime Data to Support Analytics

George A. Vouros, Christos Doulkeridis, Georgios Santipantakis and Akrivi Vlachou

Abstract This article presents important challenges and progress toward the management of data regarding the maritime domain for supporting analysis tasks. The article introduces our objectives for big data–analysis tasks, thus motivating our efforts toward advanced data-management solutions for mobility data in the maritime domain. The article introduces data sources to support specific maritime situation–awareness scenarios that are addressed in the datAcron [The datAcron project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 687591 (http://datacron-project.eu).] project, presents the overall infrastructure designed for managing and exploiting data for analysis tasks, and presents a representation framework for integrating data from different sources revolving around the notion of semantic trajectories: the datAcron ontology.

Keywords Big data · Mobility data · Semantic trajectories · Data management

1 Introduction: Overall Objectives to Manage Mobility Data

The datAcron project aims to advance the management and integrated exploitation of voluminous and heterogeneous data-at-rest (archival data) and data-in-motion (streaming data) sources so as to significantly improve the capacities of surveillance systems to promote safety and effectiveness of critical operations for large numbers of moving entities in large geographical areas. Challenges throughout the Big-Data ecosystem of systems concern effective detection and forecasting of moving entities' trajectories as well as the recognition and prediction of events due to entities' trajectories and contextual data.

G.A. Vouros (✉) · C. Doulkeridis · G. Santipantakis · A. Vlachou
Department of Digital Systems, University of Piraeus, Piraeus, Greece
e-mail: georgev@unipi.gr

V. Popovich et al. (eds.), *Information Fusion and Intelligent Geographic Information Systems (IF&IGIS'17)*, Lecture Notes in Geoinformation and Cartography, DOI 10.1007/978-3-319-59539-9_2

Challenges emerge as the number of moving entities and related operations increase at unprecedented scale. This, in conjunction with the demand for increasingly more frequent data from many different sources and for each of these entities, results in generating vast data volumes of a heterogeneous nature, at extremely high rates, whose intertwined exploitation calls for novel big-data techniques and algorithms that lead to advanced data analytics. This is a core research issue that we address in the datAcron project. More concretely, core research challenges in datAcron include the following:

- distributed management and querying of spatiotemporal RDF data-at-rest (archival) and data-in-motion (streaming) following an integrated approach;
- reconstruction and forecasting of moving entities' trajectories in the challenging maritime (2D space with temporal dimension) and aviation (3D space with temporal dimension) domains;
- recognition and forecasting of complex events due to the movement of entities (e.g., the prediction of potential collision, capacity demand, hot spots/paths); and
- interactive visual analytics for supporting human exploration and interpretation of the above-mentioned challenges.

Technological developments are validated and evaluated in user-defined challenges that aim at increasing the safety, efficiency, and economy of operations concerning moving entities in the aviation and maritime domains. The main benefit arising from improved trajectory prediction in the aviation use–case lies in the accurate prediction of complex events, or hot spots, leading to benefits to the overall efficiency of an air traffic–management (ATM) system. Similarly, discovering and characterizing the activities of vessels at sea are key tasks to Maritime Situational Awareness (MSA) indicators and constitute the basis for detecting/predicting vessel activities toward enhancing safety, detecting anomalous behaviors, and enabling an effective and quick response to maritime threats and risks.

In both domains, semantic trajectories are turned into "first-class citizens." In practice, this forms a paradigm shift toward operations that are built and revolve around the notion of trajectory. For instance, in the MSA world, trajectories are essential for tracking vessels' routes, detecting and analyzing anomalous behavior, and supporting critical decision-making. datAcron considers trajectories as first class citizens and aims to build solutions toward managing data that are connected by way of, and contribute to, enriched views of trajectories. In doing so, datAcron revisits the notion of semantic trajectory and builds on it. Specifically, it is expected that meaningful moving patterns will be computed and exploited to recognizing and predicting the behavior and states of moving objects, taking advantage of the wealth of information available in disparate and heterogeneous data sources, and integrated in a representation in which trajectories are the main entities.

The objective of this section is to review the challenges of and recent progress toward managing big data for supporting analysis tasks regarding moving objects at sea (e.g., for predicting vessels' trajectories, events, and/or support visual-analytics tasks). Such data may be surveillance data but also data regarding vessels'

characteristics, past events, areas of interest, patters of movement, etc. These are data from disparate and heterogeneous sources that should be integrated, together with the automatic computations of indicators that contribute to support maritime experts' awareness of situations.

The article presents the datAcron maritime-use case. It then presents the overall datAcron infrastructure to manage big mobility data focusing on data-management issues. It then presents the datAcron ontology for the representation of maritime data towards providing integrated views of data for disparate sources focusing on the notion of semantic trajectory.

2 Taming Big Data in the Maritime-Use Case: Motivation and Challenges

The maritime environment has a huge impact on the global economy and our everyday lives. Specifically, surveillance systems of moving entities at sea have been attracting increasing attention due to their importance for the safety and efficiency of maritime operations. For instance, preventing ship accidents by monitoring vessel activity represents substantial savings in financial cost for shipping companies (e.g., oil-spill cleanup) and averts irrevocable damages to maritime ecosystems (e.g., fishery closure). The past few years have seen a rapid increase in the research and development of information-oriented infrastructures and systems addressing many aspects of data management and data analytics related to movement at sea (e.g., maritime navigation, marine life). In fact, the correlated exploitation of heterogeneous and large-data sources offering voluminous historical and streaming data is considered as an emergent necessity given the (a) wealth of existing data, (b) the opportunity to exploit such data toward building models of entities' movement patterns, and (c) understanding the occurrence of important maritime events.

It is indeed true that reaching appropriate MSA for the decision-maker requires processing in real-time of a high volume of information of different nature, originating from a variety of sources (sensors and humans) that lack veracity and comes at high velocity. Different types of data are available, which can provide useful knowledge only if properly combined and integrated. However, the correlated exploitation of data from disparate and heterogeneous data sources is a crucial computational issue.

The growing number of sensors (in coastal and satellite networks) makes the sea one of the most challenging environments to be effectively monitored; the need for methods for processing of vessel-motion data, which are scalable in time and space, is highly critical for maritime security and safety. For instance, approximately 12,000 ships/day are tracked in EU waters, and approximately 100,000,000 AIS positions/month are recorded in EU waters (EMSA 2012). Beyond the volume of

data concerning ships' positions obtained from AIS, these trackings might not be always sufficient for the purposes of detection and prediction algorithms. Only if properly combined and integrated with other data acquired from other data/information sources (not only AIS) can they provide useful information and knowledge for achieving the maritime situational awareness in support to the datAcron maritime-use case.

The Maritime-Use Case for datAcron focuses on the control of fishing activities because it fulfills many of the requirements for validating the technology to be developed in datAcron: It addresses challenging problems deemed of interest for the maritime operational community in general; it is aligned with the European Union maritime policy and needs in particular; and it relies on available datasets (unclassified, shareable) among the teams and others of interest in the research community (e.g., AIS data, radar datasets, databases of past events, intelligence reports, etc.). Moreover, it is of considerable complexity because it encompasses several maritime risks and environmental issues such as environmental destruction and degradation as well as maritime accidents, illegal, unreported, and unregulated (IUU) fishing; and trafficking problems.

The support for processing, analyzing, and visualizing fishing vessels at the European scale, although not worldwide, along with the capability of predicting the movement of maritime objects and the identification of patterns of movement and navigational events, shall improve existing solutions to monitor compliance with the European common fisheries policy. In addition to the control of fishing activities, another core issue is safety. Fishing, even under peace conditions, is known as one of most dangerous activities and is regularly ranked among the top five dangerous activities depending on the years being considered. Safety does not concern only fishing vessels themselves but also the surrounding traffic and more generally all other human activities at sea.

The data to be used in datAcron comprise real and quasi-real data streams as well as archival (or historical) European datasets supporting the fishing scenarios specified. These usually need to be cleaned up from inconsistencies, converted into standard formats, harmonized, and summarized.

The following list briefly summarizes typical datasets that are relevant to the datAcron scenarios:

- automatic Identification System (AIS) messages broadcasted by ships for collision avoidance;
- marine protected/closed areas where fishing and sea traffic may be (temporarily) forbidden;
- traffic-separation schemes and nautical charts useful to define vessel routes;
- vessel routes and fishing areas estimated from historical traffic data;
- registry data on vessels and ports;
- records of past events such as incidents and illegal-activities reports; and
- meteorological and oceanographic (METOC) data on atmospheric and sea-state conditions and currents.

Despite the urgent need for the development of maritime data infrastructures, current information and database systems are not completely appropriate to manage this wealth of data, thus also supporting the analytics tasks targeted in datAcron. To address these limitations, we at datAcron put forward two major requirements. First, the very large data volumes generated require the development of pre-filtering data-integration process that should deliver data synopses in real-time while maintaining the main spatio-temporal and semantic properties. Next, additional ocean and atmospheric data, in conjunction to other data sources at the global and local scales are often necessary to evaluate events and patterns at sea in the most appropriate way, thus leading to additional data-integration issues [15].

In addition to the above-mentioned points, data measurements have an intrinsic uncertainty, which may be addressed by proper data-fusion algorithms and clustering in the preparation/preprocessing phase (by assessing the quality of data themselves) and by combining measurements from complementary sources [15].

3 Big-Data Management Challenges in datAcron

As already said, we at datAcron aim at recognizing and forecasting complex events and trajectories from a wealth of input data, both data-at-rest and data-in-motion, by applying appropriate techniques for Big-Data analysis. The technical challenges associated with Big-Data analysis are manifold and are perhaps better illustrated in [1, 2] where the Big Data–Analysis Pipeline is presented. As depicted in Fig. 1, five major phases (or steps) are identified in the processing pipeline:

- data acquisition and recording;
- information extraction and cleaning;
- data integration, aggregation, and representation;
- query processing, data modeling, and analysis; and
- data interpretation.

Data Acquisition

As already said, large volumes of high-velocity data are created in a streaming fashion, including surveillance data and weather forecasts that must be consumed in datAcron. One major challenge is to perform online filtering of this data in order to keep only the necessary data that contain the useful information. To this end, we apply data-summarization techniques on surveillance data, thus keeping only the

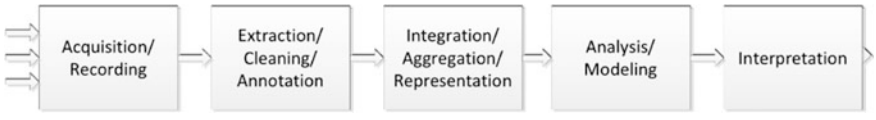


Fig. 1 Major steps in the analysis of big data (from [1, 2])

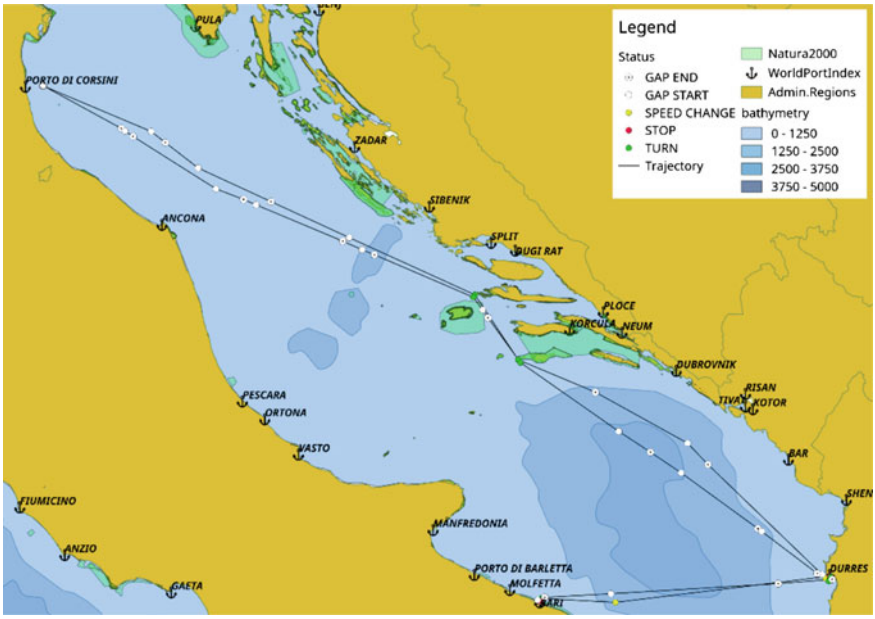


Fig. 2 Summarized trajectory: Critical points with specific meaning indicate important low-level events at specific points

"critical points" of a moving object's trajectory, which signify changes in the mobility of the moving object. Such a summarized trajectory is shown in Fig. 2 comprising the low-level events detected as critical trajectory points. A research challenge for datAcron is to achieve a data-reduction rate >90% without compromising the quality of the compressed trajectories and, of course, the quality of trajectories' and events' analysis tasks [14].

Another challenge in the data-acquisition phase is to push computation to the edges of the Big Data–management system. To achieve this, we perform online data summarization of surveillance data on the input stream directly as soon as it enters the system. Moreover, we employ in-situ processing techniques, near to the streaming data sources, in order to identify additional low-level events of interest such as the entrance/leave of moving objects in specific areas of interest (such as protected marine areas) and events requiring cross-streaming processing.

Information Extraction and Cleaning

Given the disparity of data sources exploited in datAcron, with miscellaneous data in various formats for processing and analysis, a basic prerequisite for the subsequent analysis tasks is to extract the useful data and transform it into a form that is suitable for processing. As a concrete example, weather forecasts are provided as large binary files (GRIB format), which cannot be effectively analyzed. Therefore, we extract the useful meteorological variables from these files, together with their

spatio-temporal information, so that they can be later associated with mobility data. These should be done in operational time (i.e., in milliseconds), enriching the stream(s) of surveillance data.

In addition, surveillance data are typically noisy, contain errors, and are associated with uncertainty. Data-cleaning techniques are applied in the streams of surveillance data in order to reconstruct trajectories with minimum errors, which will lead to more accurate analysis results with higher probability. Indicative examples of challenges addressed in this respect include handling delayed surveillance data and dealing with intentional erroneous data (spoofing) or hardware/equipment errors, etc.

Data Integration, Aggregation, and Representation

Having addressed data cleaning, the next challenge is to integrate the heterogeneous data coming from various data sources in order to provide a unified and combined view. Our approach is to transform and represent all input data in RDF following a common representation (i.e., the datAcron ontology), which was designed purposefully to accommodate the different data sources. However, data transformation alone does not suffice. To achieve data integration, we apply online link-discovery techniques in order to interlink streaming data from different sources, a task of major significance in datAcron.

In particular, the types of discovered links belong to different categories with the most representative ones being (a) moving object with static spatial area, (b) moving object with spatio-temporal variables, and (c) moving object with moving objects. In the first case, we monitor different relations (enter, exit, nearby) between a moving object and areas of interest such as protected natural areas or fishing zones. In the second case, we enrich the points of a trajectory with weather information coming from weather forecasts. Finally, in the last case, we identify relations between moving objects, e.g., two vessels approaching each other or staying in the same place for unusually long period. By means of link discovery, we derive enriched-data representations across different data sources, thereby providing richer information to the higher-level analysis tasks in datAcron.

Query Processing, Data Modeling, and Analysis

Another Big-Data challenge addressed in datAcron relates to the scalable processing of vast-sized RDF graphs that encompass spatio-temporal information. Toward this goal, we designed and developed a parallel spatio-temporal RDF processing engine on top of Apache Spark. Individual challenges that need to be solved in this context include RDF-graph partitioning, implementing parallel query operators that shall be used by the processing engine, and exploiting the capabilities of Spark in the context of trajectory data.

Complex event detection is also performed in datAcron where the objective is to detect events related to the movement of objects in real-time.

Last, but not least, particular attention is set toward predictive analytics, namely, trajectory prediction and event forecasting. Both short- and long-term predictions are useful depending on the domain and in particular for maritime: A difficult

problem is to perform long-term prediction. For instance, as far as trajectory prediction is concerned, we may distinguish location prediction (where a moving object will be after X number of hours) and trajectory prediction (what path will a moving object follow in order to reach position P).

Interpretation

To assist the task of human-based interpretation of analysis results, as well as the detection of patterns that may further guide the detection of interesting events—tasks that are fundamental for any Big Data–analysis platform—datAcron relies on visual analytics. By means of visual-analysis tools, it is possible to perform visual and interactive exploration of moving objects and their trajectories, visualize aggregates or data summaries, and ultimately identify trends or validate analysis results that would be hard to find automatically.

4 Semantic Trajectories Revisited: An Ontology for Maritime Data to Support Movement Analysis

Given the significance of trajectories, analysis methods (e.g., for the detection and prediction of trajectories and events), in combination with visual analytics methods, require trajectories to be (a) available at multiple levels of spatio-temporal analysis, (b) easily transformed into spatio-temporal constructs/forms suitable for analysis tasks, and (c) provide anchors for linking contextual information and events related to the movement of any object. In doing so, representation of trajectories at the semantic level aim to provide semantically meaningful integrated views of data regarding the mobility of vessels at different levels of analysis.

The term "contextual information" denotes any type of information about entities that affect the behavior of an object (e.g., weather conditions or events of special interest) as well as information about entities that are being affected by the behavior of an object (e.g., a fishing or protected area). Moreover, the context of an objects' trajectory may include the trajectories of other objects in its vicinity. As already said, surrounding traffic may entail safety concerns. The association of trajectories to contextual information and events results in enhanced semantic trajectories of moving objects.

Existing approaches for the representation of semantic trajectories suffer from at least one of the following limitations: (a) there is use of plain textual annotations instead of semantic links to other entities [3–5, 8, 10–12]; (b) only limited types of events can be represented as resources [3–7]; (c) assumptions are made of the structure of trajectories, thus restricting the levels of analysis and representations supported [6, 9]; and (d) semantic links between entities are mostly application-specific rather than generic [6, 7].

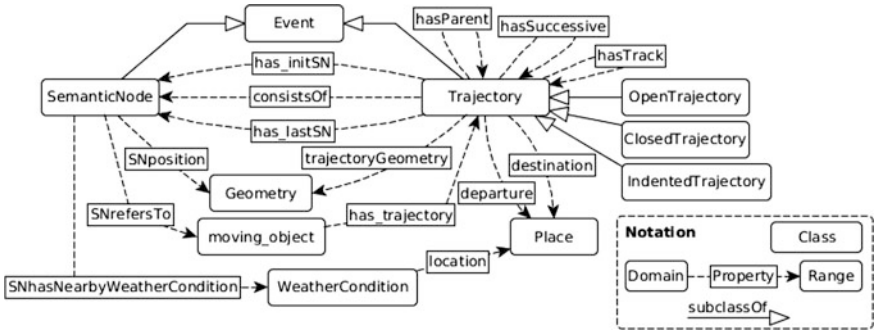


Fig. 3 Core concepts and properties

Motivated by real-life emerging needs in MSA, we aim at providing a coherent and generic scheme supporting the representation of semantic trajectories at different levels of spatial and temporal analysis: Trajectories may be seen as single geometries, as arbitrary sequences of moving objects' positions over time, as sequences of events, or as sequences of trajectories' segments each linked with important semantic information. These different levels of representation must be supported by the datAcron ontology.

The datAcron ontology is expressed in RDFS. The main concepts and properties in this ontology are depicted in Fig. 3 and are presented in the next paragraphs.

Places: The concept of "place" is instantiated by the static spatial resources representing places and regions of special interest. Places are related to any type of trajectory (or segment), weather conditions, and events (by relations "within" or "nearby"). "Place" is a generalization of Places of Interest (POIs) related to trajectories and Regions of Interest (ROIs) [13] associated with a stop event of a moving object.

A place is always related to a geometry with the property "hasGeometry."

Semantic Nodes: For the representation of moving objects' behavior at varying levels of analysis in space and time, and in order to associate trajectories with contextual information, we use the concept of "SemanticNode." A semantic node specifies the position of a moving object in a time period or instant or a specific set of spatio-temporal positions of a single moving object. In the latter case, it specifies an abstraction/aggregation of movement track positions (e.g., the centroid for a set of positions) and can be associated with a place and a temporal interval when this movement occurred. In both cases, a semantic node may represent the occurrence of an event type.

More importantly, any instance of SemanticNode can be associated with contextual information known for the specific spatio-temporal position or ROI.

In addition, the semantic node may be associated with weather information regarding the node's spatio-temporal extent.

Trajectories: A trajectory is a temporal sequence of semantic nodes or trajectory segments. The main properties relating trajectories to semantic nodes are

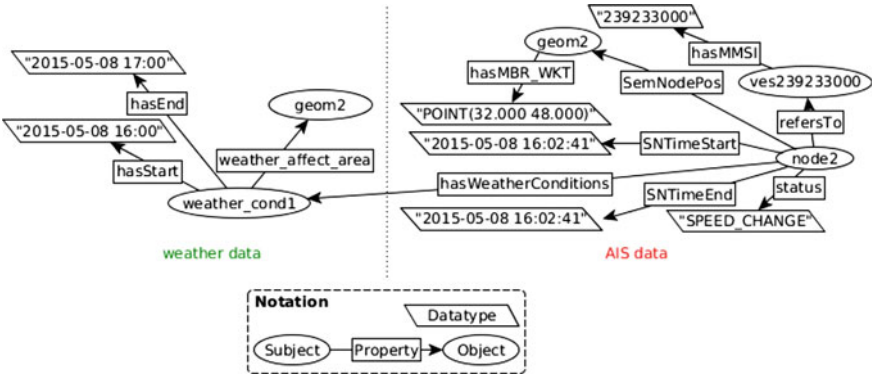


Fig. 4 A semantic-node instance linked to weather data

"hasInitNode," "hasLastNode" (representing the trajectory initial and last semantic nodes, respectively) and "consistsOf" (for relating trajectories to intermediate semantic nodes). The property "hasNext" relates consecutive semantic nodes in a trajectory to maintain the temporal sequence of nodes (Fig. 4).

Various types of trajectories are supported such as "OpenTrajectory" where the last semantic (terminal) node is not yet reached and "ClosedTrajectory," in which the last node is specified. Having said this, it should be noticed that criteria for determining terminal positions are application and domain specific. A trajectory can also be classified as "intendedTrajectory" specifying a planned or predicted trajectory. Thus, each moving object, at a specific time, may be related to multiple trajectories, actual or intended/predicted ones, and semantic nodes can be reused in different types of trajectories, w.r.t. spatial and temporal granularity.

The properties "hasParent" and "hasSuccessive" relate a trajectory with other trajectories, thus forming a structured trajectory. Specifically, the first property relates a trajectory to its parent (the whole), and the second one relates the successive trajectories (the parts).

For instance, Fig. 5 illustrates the trajectory of a vessel through ports Porto Di Corsini, Durres, and Bari, and Fig. 6 demonstrates the corresponding structured trajectory with its trajectory segments.

Trajectory segments have a starting and an ending semantic node and are associated with a time interval and geometry.

Event: The "Event" concept is instantiated by spatio-temporal entities representing specific, aggregated, or abstracted positions instantiating a specific event pattern. The instantiation of any such event pattern can be part of a preprocessing task on raw data, or it can be done by a function applied to the RDF data, thus resulting in the generation of new triples representing the recognized events. Thus, an event is associated with a set of semantic nodes, which may be in temporal sequence. Each event may be associated with one or more moving objects, and it has spatial, temporal and domain-specific properties according to the properties of

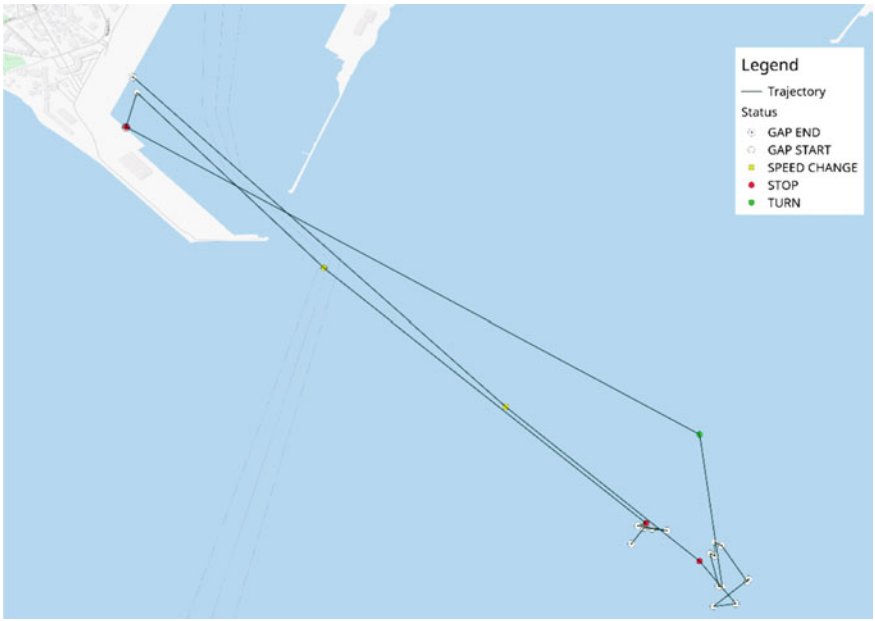


Fig. 5 An example trajectory: The figure shows the activity of a vessel near Durres port. The multiple stop positions with intermediate gap-start and gap-end points could indicate a suspicious behavior

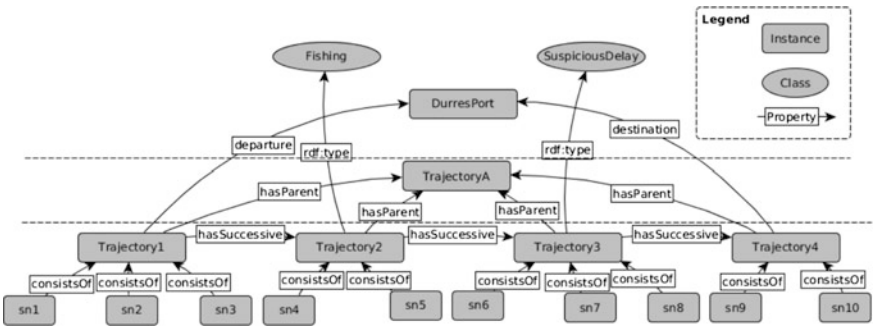


Fig. 6 A structured trajectory

semantic nodes and trajectories. It must be pointed out that a semantic node or a trajectory can be specified to be associated to more than one event types (e.g., "Rendezvous" and "PackagePicking").

Events are distinguished as to low-level and high-level events: The former are those detected from raw trajectory data or from time-variant properties of a single moving object disregarding contextual data. For instance, a "Turning" or an "Accelerating" event is a low-level event because it concerns a specific object and

can be detected directly from raw trajectory data. Figures 2 and 5 depict such low-level events as trajectory "critical" points. High-level (or complex) events are detected or predicted by means of features specifying movement and/or time-variant properties, in addition to contextual ones, of moving objects. For example, the detection of a "Fishing" event needs consideration of the type of the vessel and the known fishing regions in addition to the vessel's raw trajectory.

5 Concluding Remarks

The datAcron project aims to advance the management and integrated exploitation of voluminous and heterogeneous data-at-rest (archival data) and data-in-motion (streaming data) sources so as to address important challenges in time critical domains, such as the maritime domain, for supporting analysis tasks. It is indeed true that vast data volumes of heterogeneous nature, flowing at extremely high rates —whose intertwined exploitation for supporting analysis tasks in the maritime domain, is an emergent necessity—calls for novel big-data techniques and algorithms that lead to advanced data analytics.

Toward achieving its objectives, datAcron considers semantic trajectories to be "first-class citizens" following the paradigm shift towards operations that are built and revolve around the notion of trajectory. Thus, datAcron revisits the notion of semantic trajectory and builds on it. Specifically, it is expected that meaningful moving patterns will be computed and be exploited to recognizing and predicting the behavior and states of moving objects taking advantage of the wealth of information available in disparate and heterogeneous data sources.

Given the significance of trajectories, analysis methods (e.g., for the detection and prediction of trajectories and events), in combination with visual analytics methods, require trajectories to be (a) available at multiple levels of spatio-temporal analysis, (b) easily transformed to spatio-temporal constructs/forms suitable for analysis tasks, and (c) able to provide anchors for linking contextual information and events related to the movement of any object. Toward these objectives, datAcron has devised a representation for trajectories at the semantic level, providing semantically meaningful integrated views of data regarding the mobility of vessels at different levels of analysis.

Finally, as already mentioned, we address issues of all five major phases (or steps) identified in the processing pipeline of a big-data architecture:

- data acquisition and recording;
- information extraction and cleaning;
- data integration, aggregation, and representation;
- query processing, data modeling, and analysis; and
- interpretation.

The design of the datAcron overall architecture reflects these issues and is in close connection with requirements of the maritime domain.

Our current work focuses on data integration, aggregation, and representation as well as on query processing, data modeling, and analysis. Methods aim toward providing big-data solutions to these processing phases even during operational times.

References

1. Agrawal D et al (2012) Challenges and opportunities with big data: A white paper prepared for the computing community consortium committee of the computing research association. http://cra.org/ccc/resources/ccc-led-whitepapers/
2. Jagadish HV, Gehrke J, Labrinidis A, Papakonstantinou Y, Patel JM, Ramakrishnan R, Shahabi C (2014) Big data and its technical challenges. Commun. ACM 57(7): 86–94
3. Alvares LO, Bogorny V, Kuijpers B, de Macêdo JAF, Moelans B, Vaisman AA (2007) A model for enriching trajectories with semantic geographical information. In: Proc. of GIS, p 22
4. Baglioni M, de Macêdo JAF, Renso C, Trasarti R, Wachowicz M (2009) Towards semantic interpretation of movement behavior. In: Advances in GIScience, Springer, pp 271–288
5. Bogorny V, Kuijpers B, Alvares LO (2009) ST-DMQL: A semantic trajectory data mining query language. Int J Geogr Inf Sci 23(10):1245–1276
6. Bogorny V, Renso C, de Aquino AR, de Lucca Siqueira F, Alvares LO (2014) Constant - A conceptual data model for semantic trajectories of moving objects. Trans GIS 18(1):66–88
7. Fileto R, May C, Renso C, Pelekis N, Klein D, Theodoridis Y (2015) The Baquara2 knowledge-based framework for semantic enrichment and analysis of movement data. Data Knowl Eng 98:104–122
8. van Hage WR, Malaisé V, de Vries G, Schreiber G, van Someren M (2009) Combining ship trajectories and semantics with the simple event model (SEM). In: Proc. of EIMM., pp 73–80
9. Nogueira TP, Martin H (2015) Querying semantic trajectory episodes. In: Proc. Of MobiGIS., pp 23–30
10. Spaccapietra S, Parent C, Damiani ML, de Macêdo JAF, Porto F, Vangenot C (2008) A conceptual view on trajectories. Data Knowl Eng 65(1):126–146
11. Yan Z, Chakraborty D, Parent C, Spaccapietra S, Aberer K (2013) Semantic trajectories: Mobility data computation and annotation. ACM TIST 4(3):49
12. Yan Z, Macedo J, Parent C, Spaccapietra S (2008) Trajectory ontologies and queries. Trans GIS 12(s1):75–91
13. Baglioni M, de Macêdo JAF, Renso C, Trasarti R, Wachowicz M (2009) Towards semantic interpretation of movement behavior. In: Advances in GIScience, Springer, pp 271–288
14. Patroumpas K, Alevizos E, Artikis A, Vodas M, Pelekis N, Theodoridis Y (2017) Online event recognition from moving vessel trajectories. GeoInformatica 21(2):389–427
15. Claramunt C, et al (2017) Maritime data integration and analysis: recent progress and research challenges, EDBT 2017, 192–197

Fuzzy-Vector Structures for Transient-Phenomenon Representation

Enguerran Grandchamp

Abstract This paper deals with data structures within GIS. Continuous phenomena are usually represented by raster structures for simplicity reasons. With such structures, spatial repartitions of the data are not easily interpretable. Moreover, in an overlapping/clustering context, these structures remove the links between the data and the algorithms. We propose a vector representation of such data based on non-regular multi-ring polygons. The structure requires multi-part nested polygons and new set operations. We present the formalism based on belief theory and uncertainty reasoning. We also detail the implementation of the structures and the set operations. The structures and the set operations are illustrated in the context of forest classification having diffuse transitions.

Keywords Data structures · Transient phenomenon · Forest classification · Fuzzy

1 Introduction

The content of paper lies at the crossroads of knowledge representation [16], fuzzy modeling [1, 6] and geographic information systems. Indeed, we propose a new way to represent uncertain knowledge within a GIS [2, 7, 8].

The fuzzy concept is used to represent uncertainty [3, 5]. A fuzzy set is based on a belief function f (see Fig. 1) [10–13], which gives for each point P of the space a confidence coefficient ($C \in [0, 1]$) regarding the membership of P to a given hypothesis or characteristics.

Within a GIS there are two ways to represent data [15]: a vector layer [4] or a raster one. Without loss of generality, let us consider spatial data in the rest of the article [i.e., two dimensions $P(x, y)$]. Spatial data can be classified in two categories: discrete and continuous. Discrete data—such as buildings, roads, or administrative

E. Grandchamp (✉)
LAMIA, Université des Antilles, Campus de Fouillole, 97157 Pointe-à-Pitre, Guadeloupe, France
e-mail: enguerran.grandchamp@univ-antilles.fr

V. Popovich et al. (eds.), *Information Fusion and Intelligent Geographic Information Systems (IF&IGIS'17)*, Lecture Notes in Geoinformation and Cartography, DOI 10.1007/978-3-319-59539-9_3

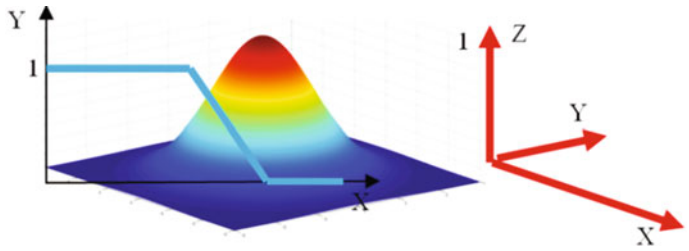


Fig. 1 One-dimensional (confidence along the y-axis) and 2D (confidence along the z-axis) Belief function

limits—naturally find a representation with the vector formalism because they have, by definition, a well-defined and localized border. In contrast, continuous data, often resulting from a phenomenon (such as temperature, humidity, ecosystem ground occupation, forests, etc.), find a better representation in a raster. The space is then divided in regular cells using a grid having a resolution depending on the required precision, the sources, or any other criterion such as memory limits to load or store the data.

In the GIS context, fuzzy structures are represented in a raster way (Figs. 1 and 2). We propose in this paper a vector structure for fuzzy sets as well as implementation and coding inspired by the standard notation WKT as defined by the OGC.

The rest of this article is organized as follows: Sect. 2 presents the standard notation for classical structures as defined by the OGC. Section 3 gives our formalism based on previous ones. Section 4 presents the constraints on the topology when building the fuzzy polygons, the fuzzy set operation, and the fusion of two fuzzy sets. Section 5 gives the conclusion and perspectives of the study.

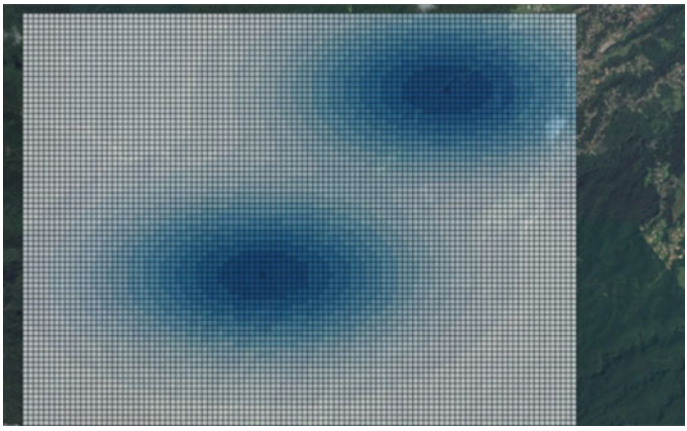


Fig. 2 Raster representation of fuzzy sets

2 OGC-SFS Formalism

In this section, we give a brief and partial description of WKT encoding, which is more detailed for polygons in order to prepare fuzzy description. For a complete description of WKT notation, readers are referred to [14].

2.1 Wkt

WKT (Well-Known Text) encoding allows the description of geometries with a simple string. Described points, lines and polygons can be the following:

simple or complex: A polygon is complex if it has an interior ring.
single or multiple (MULTI): composed of distinct objects
in two or three (Z) dimensions
with or without a measure (M) for each coordinate.

Table 1 gives some encoding examples, and Fig. 3 gives some illustrations.

3 Fuzzy Formalism

In this paper, we only deal with simple polygons (i.e., those without interior rings). The extension to complex polygons will we presented in another article. The main particularities with such polygons is the notion of interior and exterior, which is inverted for fuzzy rings starting from the interior ring of the fuzzy polygon.

Table 1 WKT examples

Geometry type	WKT example
XY POINT	POINT(850.12 578.25) One point
XYZ LINESTRING	LINESTRINGZ(10.0 15.0 30.2, 20.5 30.5 15.2, 25.90 45.12 75.6) Three summits
POLYGON	POLYGON((1.03 150.20, 401.72 65.5, 215.7 201.953, 101.23 171.82)) exterior ring, no interior ring POLYGON((10 10, 20 10, 20 20, 10 20, 10 10), (13 13, 17 13, 17 17, 13 17, 13 13)) exterior ring and interior ring
XY + M MULTILINESTRING	MULTILINESTRINGM((1 2 100.1, 3 4 200.0), (5 6 150.3, 7 8 155.4, 9 10 185.23), (11 12 14.6, 13 14 18.9))
MULTIPOLYGON	MULTIPOLYGON(((0 0,10 20,30 40,0 0), (1 1,2 2,3 3,1 1)), ((100 100,110 110,120 120,100 100))) Two polygons; the first one has an interior ring

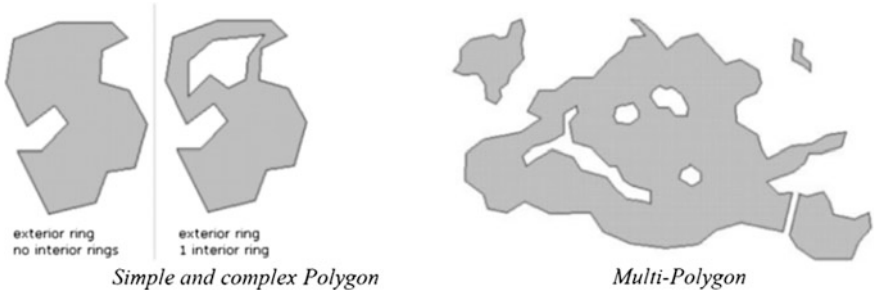


Fig. 3 Polygon and multi-polygon illustration

Fuzzy Ring

A fuzzy ring is a closed LineString representing a constant confidence value (C). The given notation in the WKT extension is as follows:

FuzzyRingC = LINESTRINGF(C, X1 Y1, X2 Y2, …, Xn Yn, X1 Y1)

With C in [0,1]

Fuzzy Polygon

A fuzzy polygon splits the space into two subspaces:

1. the exterior (Ext) of the fuzzy polygon is the subspace having a confidence ($Conf$) equals to 0: $Ext(FP) = \{(X, Y) | Conf(X, Y) = 0\}$
2. the interior (Int) of the fuzzy polygon is the subspace having a positive confidence $(Conf)$: $Int(FP) = \{(X, Y) | Conf(X, Y) > 0\}$

FuzzyPolygon((ExteriorBorderFuzzyRing), (FuzzyRing_{C1}), … (FuzzyRing_{cn}))

The Exterior ring has a confidence value = 0.

The ExteriorBorderFuzzyRing = FuzzyRing0.

The rings are listed in an ascendant order of confidence value. It is possible to have several rings with the same confidence (see Fig. 4).

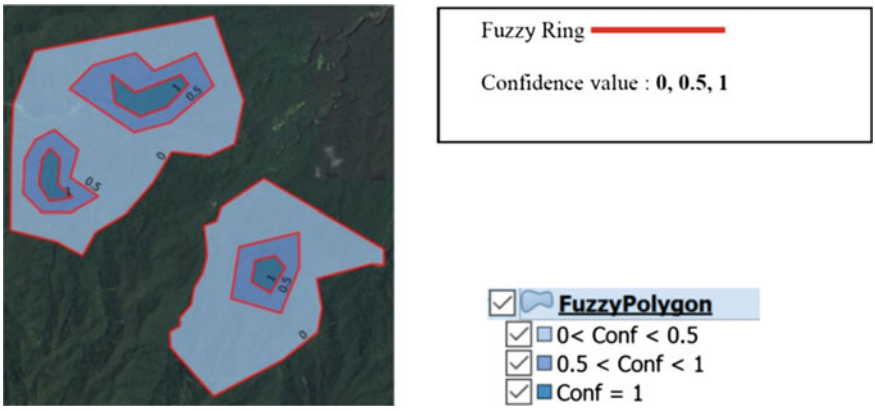


Fig. 4 Multi fuzzy polygon illustration

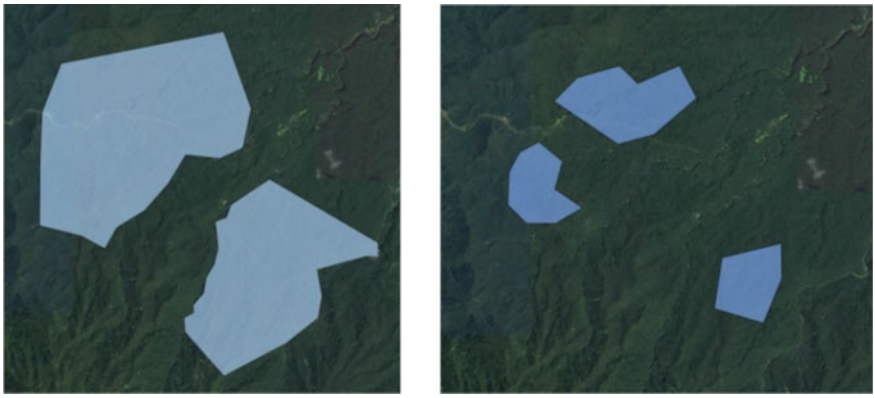


Fig. 5 Fuzzy polygon for confidence >0 and >0.5

By similarity with the previous notation, each fuzzy ring with confidence C taken separately can be viewed as a fuzzy polygon with the same confidence (FP_C) with one ring separating the space into two subspaces (see Fig. 5).

1. The exterior (*Ext*) of the fuzzy polygon is the subspace having a confidence (*Conf*) lower than C: $Ext(FP_C) = \{(X, Y) | Conf(X, Y) \leq C\}$
2. The interior (*Int*) of the fuzzy polygon is the subspace having a confidence (*Conf*) greater than C: $(Conf) : Int(FP_C) = \{(X, Y) | Conf(X, Y) > C\}$

A fuzzy multi-polygon is based on the WKT multi-polygon, and is a set of fuzzy polygon.

FuzzyMultiPolygon = MULTIPOLYGONF((FuzzyPolygon$_1$),…, (Fuzzy Polygon$_m$).

The extended WKT notation corresponding to Fig. 4 is as follows:

```
MULTIPOLYGONF(  (  (0,  636509.52544101  1790082.24725461,  …,
638329.97928085    1788083.94830689),    (0.5,    636530.7885287
1788728.42395832,  …,  636640.38068263  1787639.0679111),  (0.5,
637406.21786877    1789870.69902808,     …,     638042.35004102
1788857.08355125),  (1,  636635.43399854  1788435.28033697,  …,
636590.60850521    1788064.5607564),     (1,     637689.83884426
1789731.97180041,  …,   637743.94207098  1789302.2811451)),  ((0,
639381.8755922     1787737.17118929,     …,     640929.70958006
1786226.83502128),  (0.5,  639655.63021098  1787134.25010224,  …,
639655.63021098    1787134.25010224),    (1,    639901.14677616
1786931.41452703, …, 639901.14677616 1786931.41452703) ) )
```

Fig. 6 Fuzzy set operations

4 Constraints on Topology and Fuzzy Set Operations

Constraints

The building of a fuzzy set is governed by some rules and constraints on the topology in order to guarantee the well-formed structure.

The main constraint is that there is no intersection (neither point nor lines) between the fuzzy rings.

This leads to a succession of rings (or polygons) inclusion. They represent a followed suit with the following rule:

$$C_1 > C_2 \Rightarrow FP_{C1} \subset FP_{C2} \text{ Then } FP_1 \subset \ldots \subset FP_0$$

Fuzzy Set Operation

A fuzzy set is a set of regions labeled with a confidence. The regions represent a partition of the space from the highest confidence (1) to the lowest (0). The region with confidence between C_1 and C_2 is obtained with the set difference between FP_{C1} and FP_{C2} (Fig. 6).

Fuzzy polygon fusion

When two or more fuzzy polygons split the same geographical area, they must be merged. The set operation sequence allowing merging the fuzzy polygons is as follows:

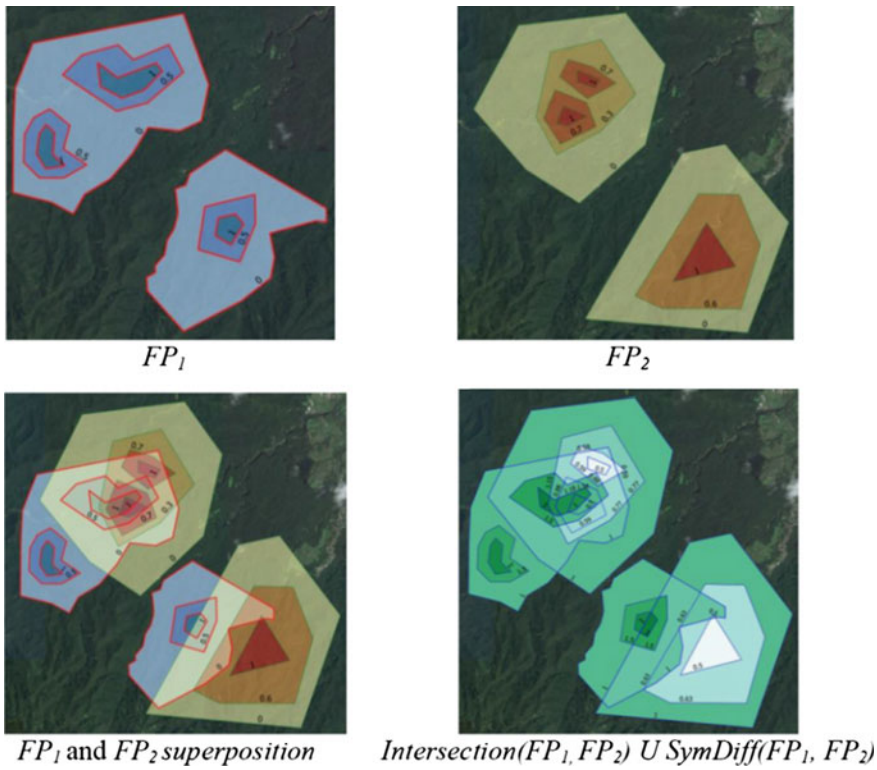


Fig. 7 Fuzzy polygons merge

1. I = Intersection between FP_1 and FP_2
2. SD = Symmetrical difference between FP_1 and FP_2
3. Merge Set = Union of I and SD

Figure 7 illustrates this principle.

The resulting fuzzy polygons have two confidence values. Depending on the nature of FP_1 and FP_2, the meaning of the confidence of these two values can be ordered in a two-objectives way or within a single expression.

In Fig. 7, the expression used to display the resulting set is $(C_1 + 1)/(C_2 + 1)$.

5 Conclusion and Perspectives

We present in this paper an overview of fuzzy-vector structures within GIS based on the OGC standard WKT encoding. It also presents constraints and basic operation such as sub-region extraction and fuzzy-polygon fusion. A complete description of the

encoding of complex fuzzy polygons, as well as all of the operators, will be given in a future paper.

Management scripts are developed in JAVA and Python for a better integration within a GIS tool. Currently, we are working on a better representation of fuzzy polygons. Indeed, when a two–fuzzy polygon model shows two structures evolving close to each other (such as two types of forest or other ecosystems), and having a transition that spreads over a long distance, the diffuse border involves an overlap of the corresponding fuzzy polygons. We must develop a tool allowing us to automatically visualize the two fuzzy polygons at the same time.

Finally, in a classification perspective based on the fuzzy point of view, this approach raises the question of such a transition [9]. Belonging to one of the original classes or to another new one. This question will be discussed in another issue and will deal with multi-label classification, emerging class, and overlap classification.

References

1. Altman D (1994) Fuzzy set theoretic approaches for handling imprecision in spatial analysis. Int J Geogr Inf Syst 8(3):271–289
2. Benz Ursula C et al (2004) Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. ISPRS J Photogrammetry Remote Sens 58:239–258
3. Bjorke JT (2004) Topological relations between fuzzy regions: derivation of verbal terms. Fuzzy Sets Syst 141:449–467
4. Coros S, Ni JingBo, Matsakis Pascal (2006) Object localization based on directional information: Case of 2D vector data. In: 14th Int. Symposium on Advances in Geographic Information Systems (ACM-GIS)
5. Cross VV (2001) Fuzzy extensions for relationships in a generalized object model. International Journal on Intelligent Systems 16:843–861
6. Fisher P (2000) Sorites paradox and vague geographies. Fuzzy Sets Syst 113:7–18
7. Grandchamp E (2012) Raster vector integration within GIS. Chap. 2, The geographical information sciences, Intech, http://dx.doi.org/10.5772/51413 pp 22–36, ISBN: 978-953-51-0824-5
8. Grandchamp E (2012) Data structures for fuzzy objects in Geographic Information Systems. SELPER, Cayenne, Guyane, 19–23 Nov 2012
9. Grandchamp E, Régis S, Rousteau A (2012) Vector transition classes generation from fuzzy overlapping classes. In: Alvarez L, Mejail M, Gomez L, Jacobo J (eds) Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. 17th Iberoamerican Congress, CIARP 2012, Buenos Aires, Argentina, 3–6 Sept 2012, Springer, pp 204–211, ISBN: 978-3-642-33274-6 (Print) 978-3-642-33275-3 (Online)
10. Guesgen HW, Albrecht J (2000) Imprecise reasoning in geographic information systems. Fuzzy Sets Syst 113:121–131
11. Guo D, Guo R, Thiart C (2004) Integrating GIS with Fuzzy Logic and Geostatistics: Predicting air pollutant PM10 for California, Using Fuzzy Kriging
12. Kainz (2007) Chap 1. Fuzzy Logic and GIS. University of Vienna (Vienna)
13. Karimi M, Menhaj MB, Mesgari MS (2008) Preparing Mineral Potential Map Using Fuzzy Logic In GIS Environment. In: ISPRS, vol XXXVII
14. OGC, WKT http://www.opengeospatial.org/standards/wkt-crs
15. Sawatzky D, Raines GL, Bonham-Carter G (2008) Spatial data modeller. Technical Report
16. Schneider M (1997) Spatial data types for database systems finite resolution geometry for geographic information systems. Series: Lecture Notes in computer science, vol 1288. p 275, ISBN: 978-3-540-63454-6

Detecting Attribute-Based Homogeneous Patches Using Spatial Clustering: A Comparison Test

Thi Hong Diep Dao and Jean-Claude Thill

Abstract In spatial and urban sciences, it is customary to partition study areas into sub-areas—so-called regions, zones, neighborhoods, or communities—to carry out analyses of the spatial patterns and processes of socio-spatial phenomena. The purposeful delineation of these sub-areas is critical because of the potential biases associated with MAUP and UGCoP. If one agrees to characterize these sub-areas based on their homogeneity on a certain attribute, these homogeneous areas (patches) can be detected by data-driven algorithms. This study aims to evaluate the effectiveness and performance of five popular spatial clustering and regionalization algorithms (AZP, ARISeL, max-p-regions, AMOEBA, and SOM) for detecting attribute-based homogeneous patches of different sizes, shapes, and those with homogeneous values. The evaluation follows a quasi-experimental approach: It is based on 68 simulated data sets where the true distribution of patches is known and focuses purely on the capability of algorithms to successfully detect patches rather than computational costs. It is the most comprehensive assessment to-date thanks to a systematic control of various conditions so that a true baseline is available for comparison purposes. Among the tested algorithms, SOM and AMOEBA were found to perform very well in detecting patches of different sizes, different shapes, including those with holes, and different homogeneous values.

Keywords Spatial clustering algorithms · Comparison · Homogeneous patches

T.H.D. Dao (✉)
Department of Geography and Environmental Studies, University of Colorado-Colorado Springs, 1420 Austin Bluffs Pkwy, Colorado Springs, CO 81918, USA
e-mail: tdao@uccs.edu

J.-C. Thill
Department of Geography and Earth Sciences, University of North Carolina-Charlotte, 9201 University City Blvd, Charlotte, NC 28223, USA
e-mail: Jean-Claude.Thill@uncc.edu

V. Popovich et al. (eds.), *Information Fusion and Intelligent Geographic Information Systems (IF&IGIS'17)*, Lecture Notes in Geoinformation and Cartography, DOI 10.1007/978-3-319-59539-9_4

1 Introduction

Many studies in spatial and urban sciences assume the partitioning of study areas into sub-areas—so called regions, zones, neighborhoods, or communities—to carry out analyses to understand spatial patterns and processes of certain phenomena. The assumption underlying the delineation of these sub-areas is frequently driven either by data availability (e.g., census administrative boundaries) or by one's own spatial cognition and a priori belief of the spatial realities being modelling as part of the geospatial modelling effort. This potentially leads to the well-known modifiable areal unit problem (MAUP) [1] as well as the uncertain geographic context problem (UGCoP) [2].

On the premise that these areas should be delineated on the basis of their homogeneity on a certain attribute, they can be detected by means of data-driven algorithms. Thus, these homogeneous areas can be referred to as "homogeneous attribute-based *patches*." Within the family of spatially constrained clustering and regionalization methods, a wide range of algorithms have been advanced for this purpose. Regionalization is a special kind of spatially constrained clustering where the condition of spatial contiguity among spatial objects plays a priority role. Generally, spatially constrained clustering and regionalization are classified into four groups. The first group includes a partitioning-based approach aiming to divide the data set into several clusters, which together cover the whole data space exhaustively. An analysis unit is assigned to the "closest" cluster based on a proximity or similarity measure. Some examples of this class of methods include the K-means [3], K-medoids [4], expectation maximization [5], AZP [6], ARISeL [7], and max-p-regions [8]. The second group is hierarchical clustering, which aims at decomposing the data set with a sequence of nested partitions from fine to coarse resolution. Examples under this group include ICEAGE [9] and RedCAP [10]. The third group is spatial statistical-based clustering to detect hot and cold concentrations and includes LISA statistics [11], local G statistics [12], and AMOEBA [13]. The fourth group is self-organizing neural networks, e.g., SOM [14], GeoSOM [15], and contextual neural gas (CNG) [16]. A thorough review of these algorithms can be found in [9, 10, 17–19].

This chapter aims to evaluate and compare the performance of spatial clustering and regionalization algorithms of geospatial analytics. Although the number and variety of algorithms is considerable, this study focuses particularly on five popular spatially constrained clustering techniques for detecting homogeneous attribute-based patches in geospatial data. The five selected algorithms represent three different groups of spatially constrained clustering methods including AZP, ARISeL, max-p-regions (partitioning-based), AMOEBA (spatial statistical-based), and SOM (self-organizing neural networks). Comparison tests are based on a significant number of simulated geospatial data sets whose true distribution of patches is known and designed to enable various scenarios for different cluster sizes, shapes, and homogeneous values. A homogeneous attribute-based patch here refers to a patch

(i.e., aggregated area) characterized either by an absolute homogeneous value or a range of non-absolute homogeneous values regarding a certain attribute. Comparison tests focus purely on the capability of algorithms to successfully detect the patches rather than their computational costs.

There is a significant body of work on the performance evaluation of spatial clustering for attribute-based homogeneous patches. However, these efforts are restricted to specific case studies with real-world data sets (e.g., [20, 21]) where the true distribution of patches is not available for comparison, which calls into questions the robustness of their conclusions. Other studies based on synthetic data are rather simple exercises with a limited number of scenarios [22]. Reference [23] presented a comparison test with a significant numbers of synthetic data sets simulating patches of high and low values. However, their focus was limited to different approaches implemented within the RedCAP package. Another major limitation with existing studies is that discussions on homogeneity is often restricted to within-patch similarity and neglects performance testing for different homogeneous values as well as different levels of homogeneity.

This study stands out by enhancing existing research through the comparison of algorithms from different groups of spatial clustering methods with a carefully constructed set of synthetic data for clusters of different sizes, shapes, and homogeneous values. This quasi-experimental approach sets this research apart from previous research.

The rest of the paper is organized as follows. Section 2 reviews the clustering algorithms used for comparison. Section 3 describes the synthetic data sets and different testing scenarios. Section 4 presents the evaluation approach, and Sect. 5 shows the test results. Section 6 includes a discussion of the results and the conclusions of our research.

2 Algorithms in Comparison

2.1 AZP: Automatic Zoning Procedure

The AZP algorithm was first introduced in [6] to handle zone-design process while dealing with fine-resolution census data. AZP works as a simple local-boundary optimizer to aggregate N areas into M regions with the following steps: (1) It generates a random-zoning system of N small zones into M regions with $M < N$; (2) it makes a list of the M regions; (3) it selects and removes any region K at random from this list; (4) it identifies a set of zones bordering on members of region K that could be moved into region K without destroying the internal contiguity of the donor region; and (5) it randomly selects zones from this list until either there is a local improvement in the current value of the objective function or a there is a move that is equivalently as good as the current best. Then, it makes the move, updates the list of candidate zones, and returns to step 4 or else repeats step 5 until the list is exhausted; (6) when the list for region K is exhausted, it returns to step 3,

selects another region, and repeats steps 4 through 6; (7) then it repeats steps 2 through 6 until no further improving moves are made.

AZP was improved through three subsequent versions including AZP Simulated Annealing, AZP-Tabu, and AZP-R-Tabu [24]. This study uses a modified version of AZP implemented in the clusterPy package by [25] where the objective function is to minimize F(Z) where Z is the within-cluster sum of distance similarity squares from each area to the attribute centroid of its cluster. In addition, the initial feasible solution is not random but is selected using the K-means++ algorithm [26] because this strategy has proven to generate better results [25].

2.2 *ARISeL—Automatic Rationalization with Initial Seed Location*

The ARiSeL algorithm proposed by [7] is in fact a modification of AZP-tabu algorithm [24], which itself is a modification of the AZP [6]. It too aims at aggregating N areas into P regions (P < N) while minimizing intra-regional heterogeneity, which is measured as the within-cluster sum of distance similarity squares from each area to the attribute centroid of its cluster. ARISeL pays particular attention to generating a good initial solution before running a Tabu Search algorithm for the most optimized regional boundaries. In ARISeL, a set of feasible solutions are first generated by running the K-means++ algorithm [26] several times (inits). Information on how the aggregation criterion changes through the assignation process is used to make changes in the initial set of seeds. The initial feasible solution yielding the best result will then be used to run the Tabu Search algorithm. The Tabu Search algorithm in short is a heuristic approach for global optimization [27]. The argument is that using a good feasible solution as an input to the second stage will decrease the possibility of being trapped into a local optimal solution; by the same token, the number of moves performed during the second stage is also minimized. The present study uses the ARISeL algorithm implemented in the clusterPy package [25].

2.3 *Max-p-Regions*

The max-p-regions algorithm introduced in [8, 28] aggregates a set of geographic areas into the maximum number of homogeneous regions such that the value of a spatially extensive regional attribute is above a predefined threshold value. Heterogeneity is measured as the within-cluster sum of distance similarity squares from each area to the attribute centroid of its cluster, and the regional attribute is the number of areas (spatial analysis units) in the homogeneous patch.

Being similar to AZP and ARISeL, the max-p-regions algorithm is composed of two main computational blocks: (1) construction of an initial feasible solution and (2) local improvement. There are three methods for local improvement: Greedy (execMaxpGreedy), Tabu (execMaxpTabu), and Simulated Annealing (execMaxpSa) [28]. For this study, the Tabu Search algorithm implemented in Duque et al. (2011) is used. At variance with AZP and ARISeL, the max-p-regions algorithm determines the number of regions (p) endogenously on the basis of a set of areas, with a matrix of attributes on each area and a floor constraint, instead of using a predefined number of regions. The floor constraint used in this study is the minimum number of spatial analysis units in a patch. As an algorithm for regionalization, max-p-regions, further enforces a spatial contiguity constraint on the areas within regions.

2.4 SOM—Self-organizing Maps

SOM [14] is an unsupervised neural network that adjusts its weights to represent a data-set distribution on a regular lattice of low dimensionality. It consists of an arbitrary number of neurons connected to adjacent neurons by a neighborhood relation, thus defining the topology of the output map. Associated with each of these neurons is a prototype vector of the same dimension as the input space. During the training, input vectors are presented to the SOM, and the neuron with the smallest distance to the input vector, referred to as the "best-matching unit" (BMU), is identified. Then the prototype vector of the BMU and the prototype vectors within a certain neighborhood on the map are moved in the direction of the input vector. The magnitude of the displacement depends on the distance of the neurons to the BMU on the map as well as the actual learning rate. Both the size of the neighborhood and the learning rate decrease monotonically during the learning process. Thus, in the beginning of the learning phase, the arrangement of neurons on the map can be altered significantly, whereas at the end of the training phase, only small changes are made to fine-tune the map. The trained SOM represents a low-dimensionality map of the input space where each neuron represents some portion of the input space and where the distance relationships of the input space are mostly preserved [19]. In this study, a two-dimensional rectangular regular lattice, along with a rook contiguity-based neighbourhood for the weights-updating process, are used.

2.5 AMOEBA—A Multi-directional Optimum Ecotope-based Algorithm

Developed by [13, 29], AMOEBA is a statistics-based approach to detect spatial concentrations of hot and cold attribute values by conducting multi-directional

searches for irregular clusters at the finest spatial granularity in an iterative procedure. Clusters are detected based on the local Getis-Ord G statistic. Functionally, the algorithm starts from one spatial analysis unit (i.e., area), to which neighboring areas are iteratively attached until the addition of any neighboring area fails to increase the magnitude of the local G statistic [30]. The resulting region is considered an ecotope. This procedure is executed for all areas, and the final ecotopes are defined after resolving overlaps and asserting non-randomness. AMOEBA is regarded as the most robust, unsupervised, statistics-based algorithm to detect hot and cold concentrations. Reference [31] enhanced the algorithm to achieve lower computation cost in the clusterPy package, which is used in the present study.

3 Simulated Data

Following a quasi-experimental approach, the comparison tests in this study were performed on simulated data. This allowed for generating the true distribution of patches for algorithm performance evaluation. A simulated normally distributed 30 × 30 cell data set with mean equal to 100 and SD equal to 5 was employed for this purpose. Two patches of *non-absolute* homogeneous low values and two patches of *non-absolute* homogeneous high values were introduced by taking the 5% left and right tails, respectively, of a normal distribution with mean = 100 and SD = 25 (Fig. 1). Reference [29] used the same simulated data and illustrated the performance of the AMOEBA algorithm to be very good as far as its ability to detect these distinct high- and low-value patches.

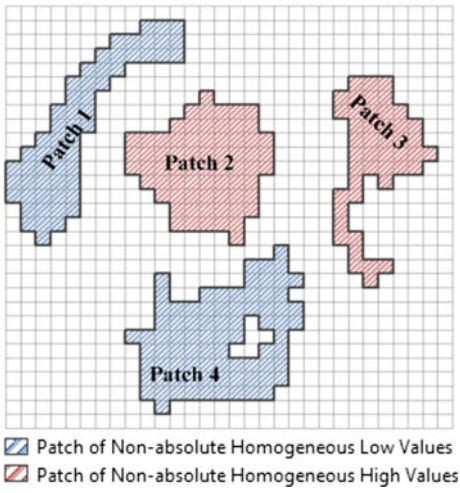


Fig. 1 Simulated data with patches of distinct high or low values

To conduct comparison tests for the five algorithms identified previously, this base data set was altered as follows to produce multiple alternative data sets: in each data set, the base value of each patch had its *absolute* homogeneous value modified, one by one, in the range of 10–170 by increment of 10. This means that each patch, in turn, was modified to have an absolute homogeneous value of 10, 20, 30, …, 150, 160, 170 while keeping the other values as in the base data set. This required the generation of 68 (17 × 4) simulated data sets in total, thus allowing for value changes on each of the four patches (Fig. 2). The clustering algorithms were then run on each of these data sets.

With these data sets, our purpose was to test algorithms for detecting homogeneous patches of different sizes, different shapes, as well as absolute homogeneous (i.e., characterized by a single value) versus non-absolute homogeneous (i.e., characterized by a range of homogeneous high or low values as captured in the base data set) patches. Such arrangement of homogeneous patches with mix types will create different scenarios of varying data distributions for testing.

4 Algorithm Evaluation

Algorithms are evaluated based on the success of detecting the absolute and non-absolute homogeneous patches. Performance evaluation of an algorithm in detecting a patch c is estimated according a standard approach through a simple fit score [23]:

$$\text{score}_c = \frac{TP}{FN + FP} \tag{1}$$

where true positive (*TP*) is the number of cells overlapping with the true simulated patch being correctly detected; false negative (*FP*) is the number of cells overlapping with the true simulated patch being not detected; and false positive (*FP*) is the number of cells not belonging to the true simulated patch but detected as such (Fig. 3). The higher the score, the better the performance of the algorithm.

Performance evaluation of an algorithm in detecting all *P* clusters can be similarly estimated based on the sum of *TP*, the sum of *FP*, and the sum of *FN* over all clusters:

$$\text{score}_{all\_c} = \frac{\sum_{c=1}^{P} TP_c}{\sum_{c=1}^{P} FP_c + \sum_{c=1}^{P} FN_c} \tag{2}$$

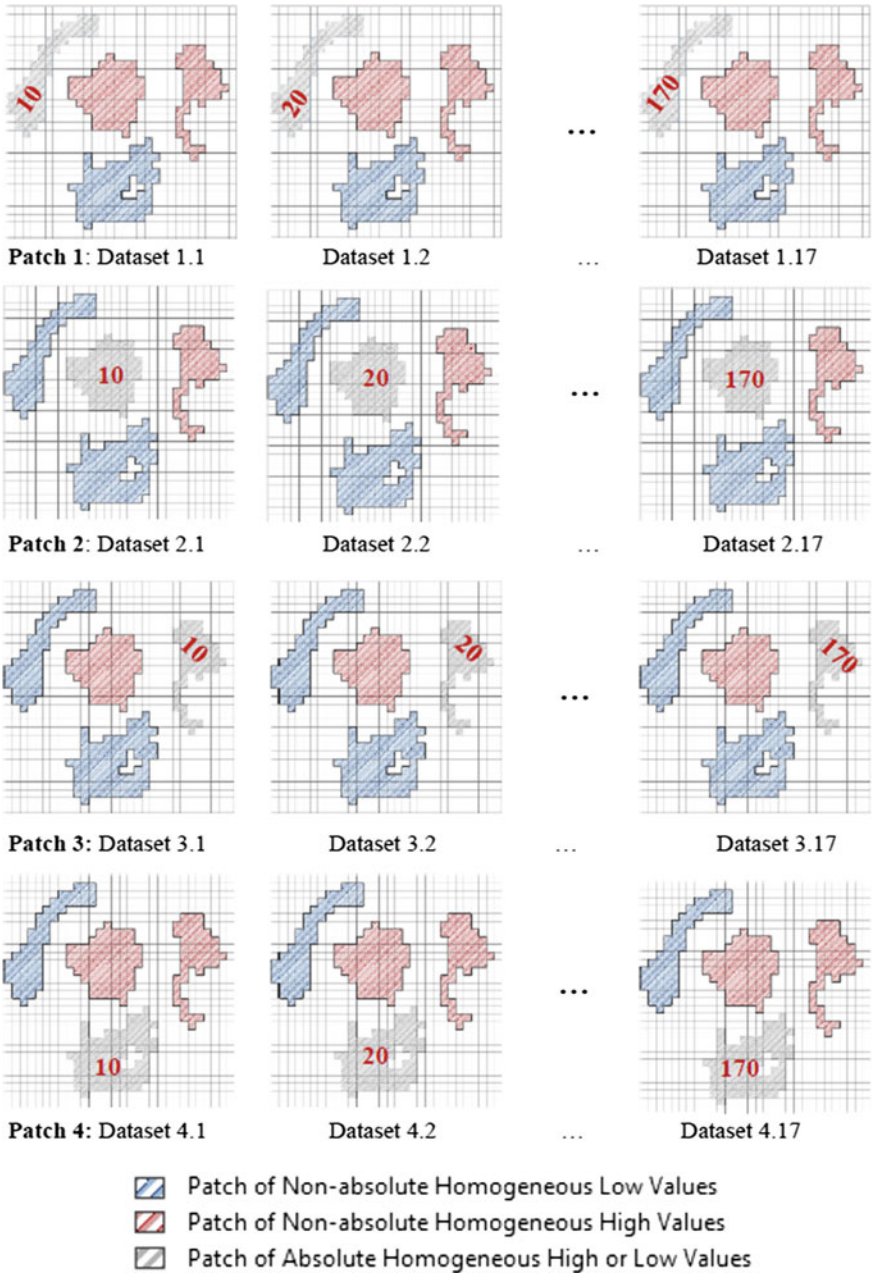


Fig. 2 Simulated data sets with patches of absolute and marginal homogeneity. The absolute homogeneous values are reported in *red* typeface on the patch

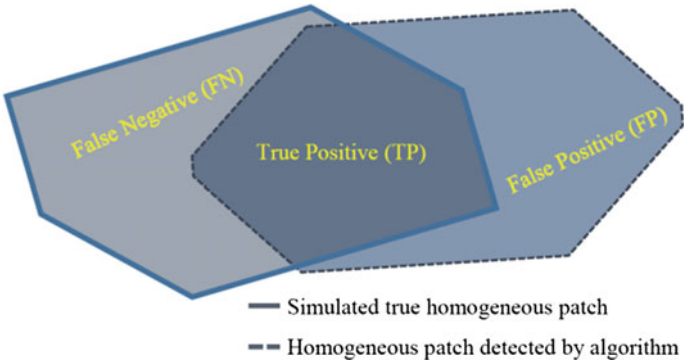


Fig. 3 Performance evaluation scheme in detecting a homogeneous patch based on true positive, false positive, and false negative

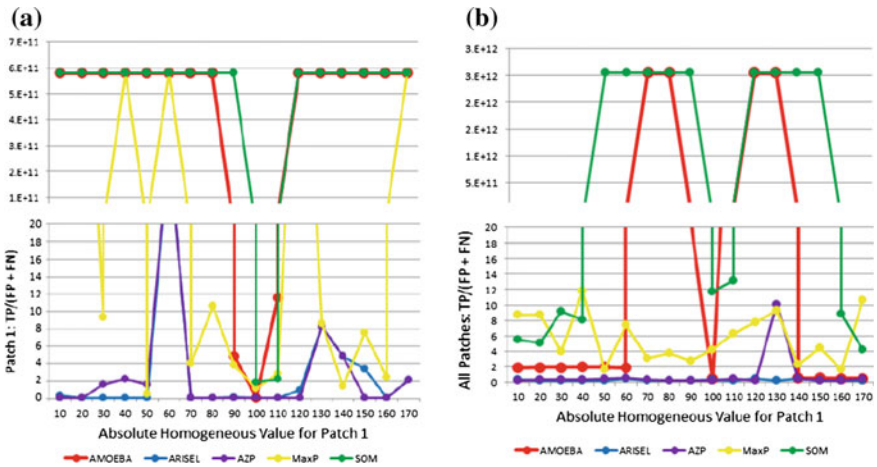


Fig. 4 Performance comparison of five algorithms in detecting **a** patch no. 1 and **b** all patches when patch no. 1 takes an absolute homogeneous value from 10 to 170

5 Comparison Results

For algorithms that require the number of patches to be prespecified, the simulated true value of 5 patches was used. When running max-p-regions, the floor constraint on the minimum number of analysis units within patches is set to 45, which is very close to the simulated true value of 46.

Figures 4, 5, 6, 7 shows the comparison results of the five algorithms when patches 1 through 4, respectively, take absolute homogeneous value ranging from 10 to 170. Each figure contains two sub-figures: The one on the left-hand side shows the performance comparison in detecting the patch whose uniform value was

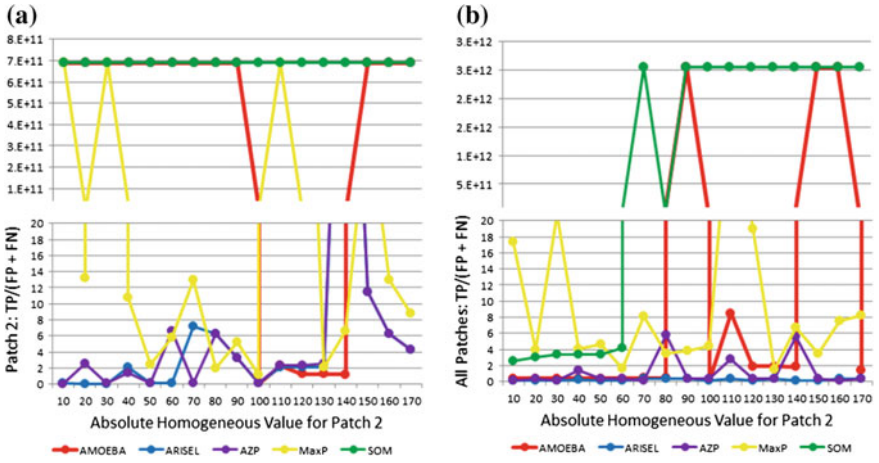


Fig. 5 Performance comparison of five algorithms in detecting **a** patch no. 2 and **b** all patches when patch no. 2 takes an absolute homogeneous value from 10 to 170

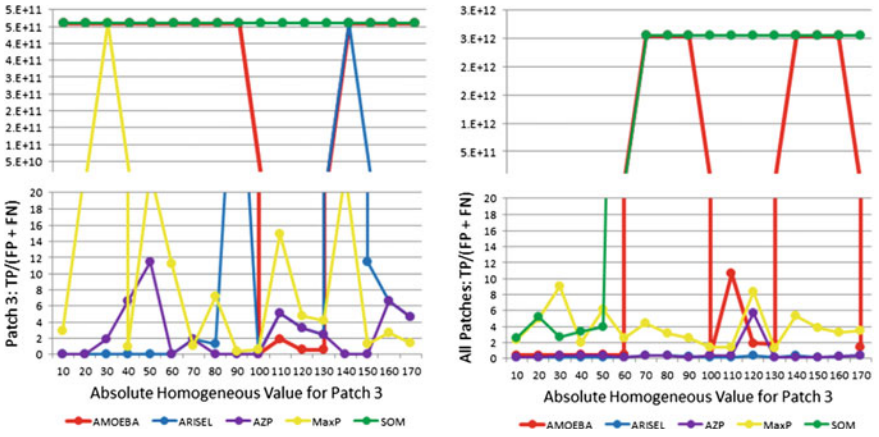


Fig. 6 Performance comparison of five algorithms in detecting **a** patch no. 3 and **b** all patches when patch no. 3 takes an absolute homogeneous value from 10 to 170

adjusted (target patch); the one on the right-hand side shows the performance comparison in detecting all patches. These figures are broken along their vertical axis to allow for the extremely large accuracy scores as well as the low values.

Generally, we find that SOM outperforms other algorithms in detecting the target patch as well as all patches. The performance of SOM in detecting all patches decreases when there exists a patch having low (i.e., 10–70) absolute homogeneous value. SOM performance also decreased when patch 1 or patch 4 took a high absolute homogeneous value in the range of 150–170. This is examined and shown in Figs. 8 and 9. The cell labels are simulated attribute values used for clustering.

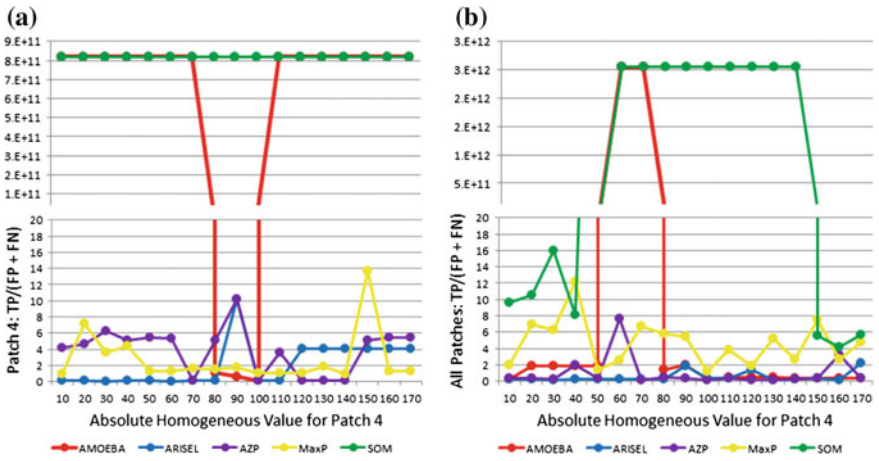


Fig. 7 Performance comparison of five algorithms in detecting **a** patch no. 4 and **b** all patches when patch no. 4 takes an absolute homogeneous value from 10 to 170

When a patch takes a very low absolute homogeneous value, the data distribution over the whole area is changed in such a way that the simulated non-absolute homogeneous of low values (patches no. 1 and 4 in Fig. 8a or patch no. 1 in Fig. 8b) have now become non-significant and are not detected. When a patch takes very high absolute homogeneous values, it also changes the data distribution and further divides the non-absolute homogeneous patch of high values (patches no. 2 and 4 in Fig. 9). At the performance dipping points, SOM however still outperforms other algorithms in detecting homogeneous patches of different sizes, shapes, and homogeneous attribute values.

Compared with SOM, AMOEBA also performs very well in detecting the target patch of varying absolute homogeneous values if the value is statistically high or low with respect to the mean of the whole data space. It is interesting to see dips in the performance of AMOEBA to detect a patch (regardless of size and shape) having absolute homogeneous value around the mean value of 100 (Fig. 10d, l). In such cases, the target patch having an absolute homogeneous value will be miss-detected and classified as being outside clusters. This is reasonable considering that the mechanism of AMOEBA is designed to detect statistically hot or cold concentrations. When being evaluated for detecting all patches, AMOEBA's performance as shown in Figs. 4, 5, 6 and 7 shows very similar patterns, i.e., a peak when the absolute homogeneous value is right below (being cool) or right above (being warm) the mean of 100. When the target patch is characterized by cool or warm values, AMOEBA's performance is generally better compared with its performance when the target patch contains an absolute homogeneous low or high value. Figure 10 depicts the performance of AMOEBA on different cases, specifically when patches no. 1 and 3 in turn take an absolute homogeneous value equal 10 to 170. When patch no. 1's absolute homogeneous value is adjusted (Fig. 10a

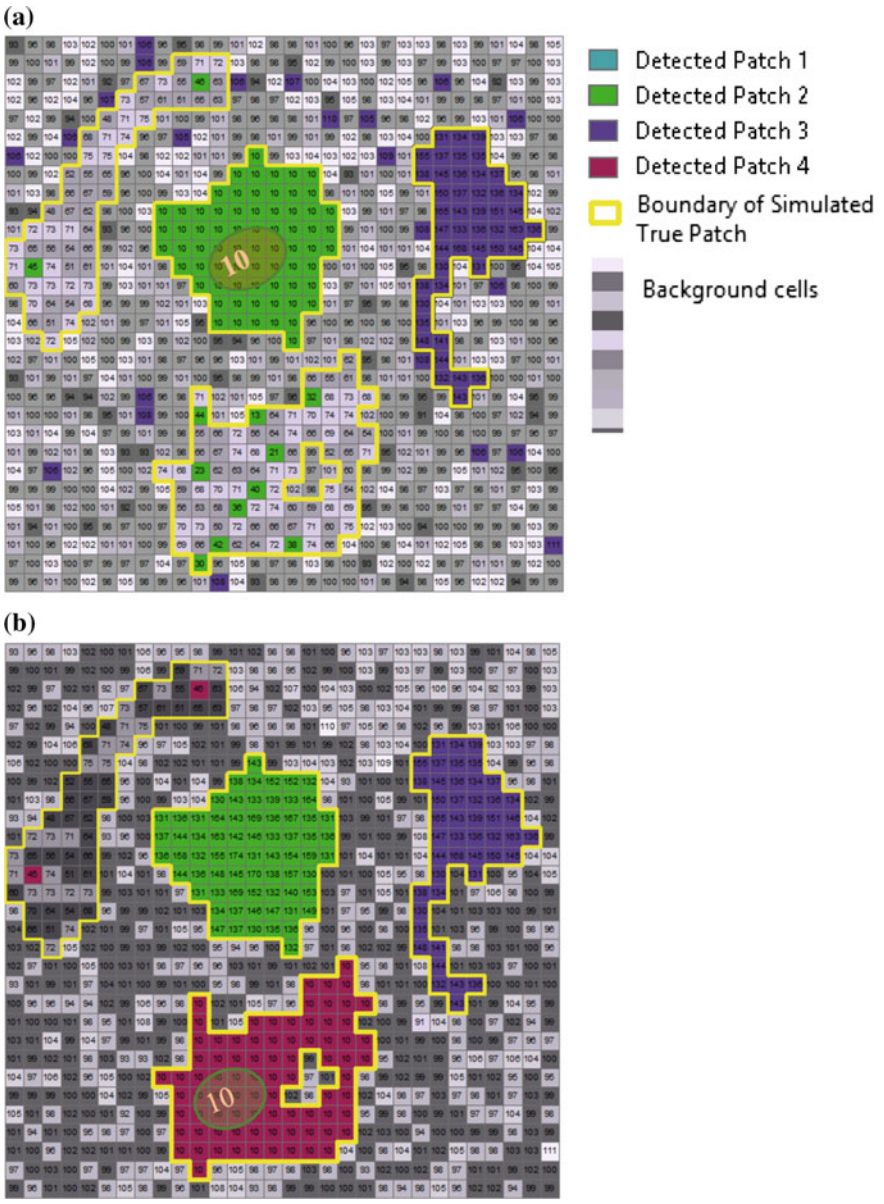


Fig. 8 Performance of SOM when **a** patch no. 2 takes 10 as the absolute homogeneous value and **b** patch no. 4 takes 10 as the absolute homogeneous value (the cell labels denote simulated attribute values)

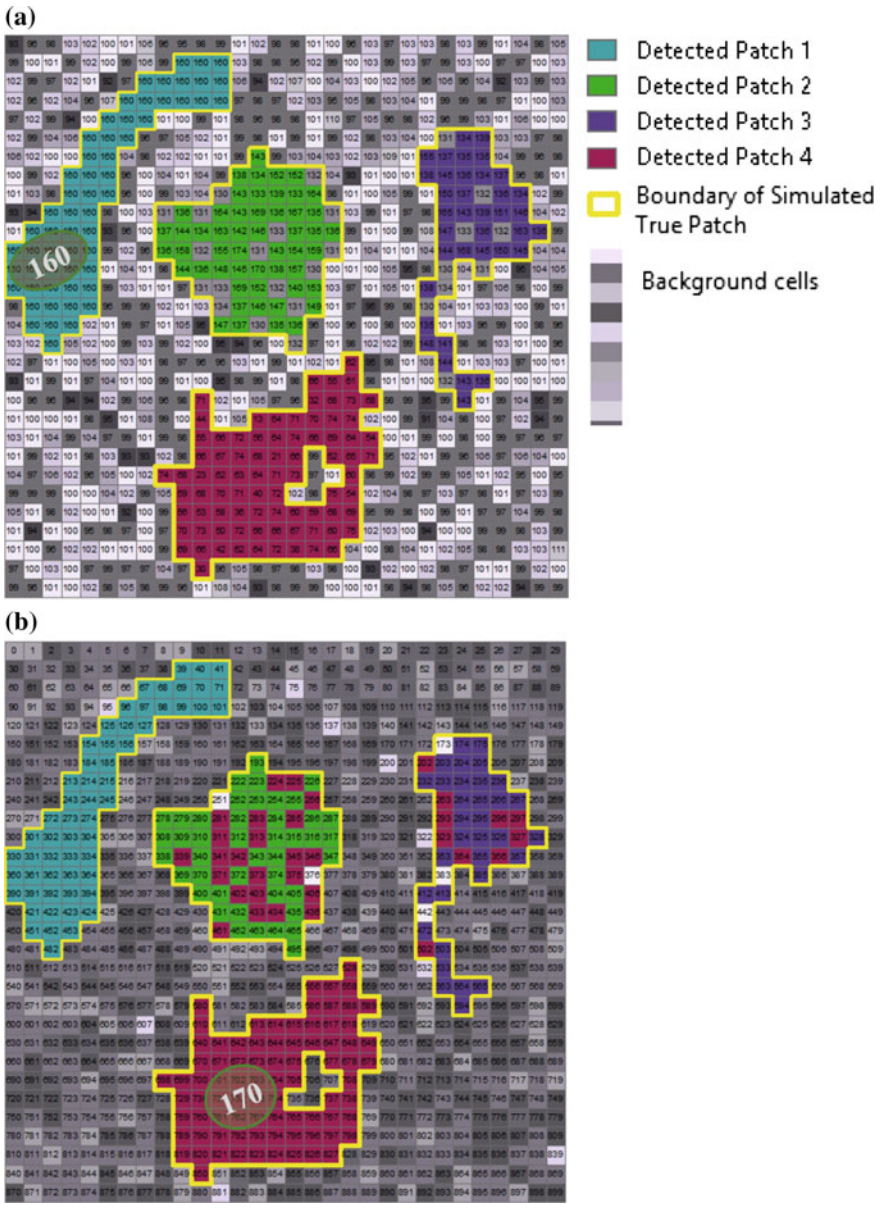


Fig. 9 Performance of SOM when **a** patch no. 1 takes 160 as the absolute homogeneous value and **b** patch no. 4 takes 170 as the absolute homogeneous value (the cell labels denote simulated attribute values)

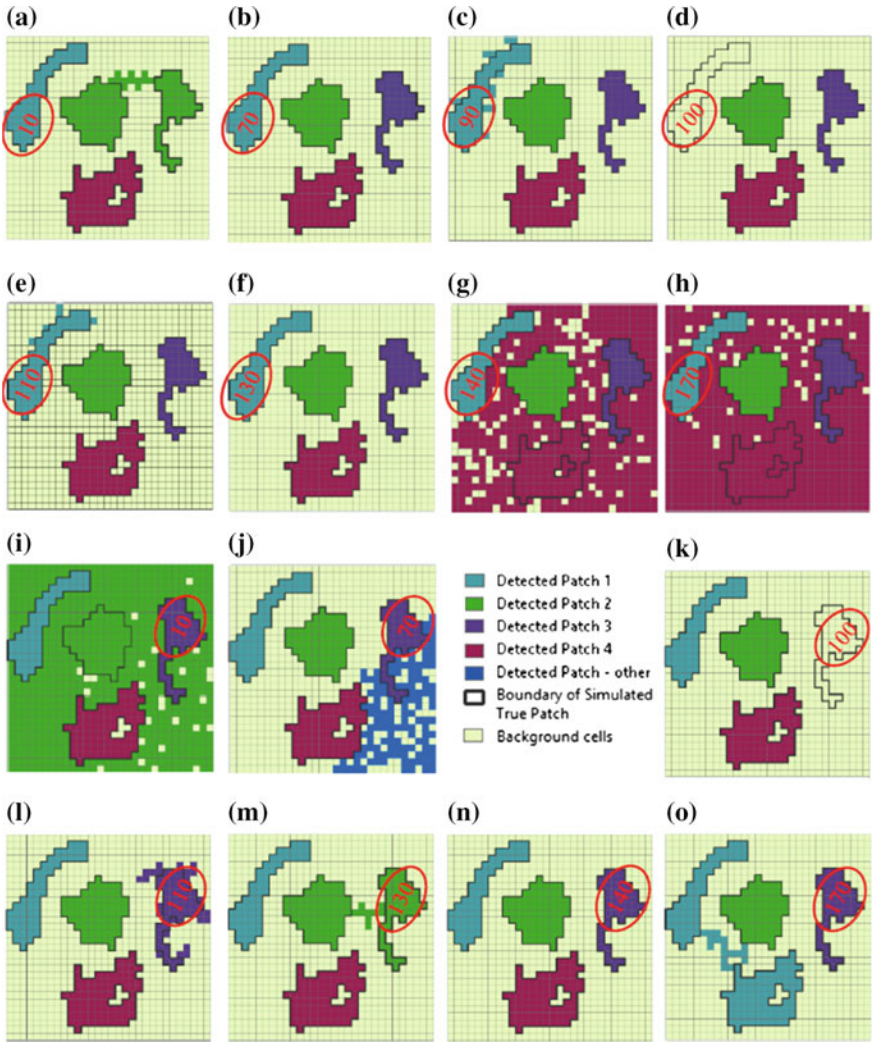


Fig. 10 Performance of AMOEBA when patch no. 1 (**a** through **h**) and patch no. 3 (**i** through **p**) take an absolute homogeneous value from 10 to 170

through h), the data distribution over the whole area is changing considering that patches no. 2 and 3 contain non-absolute homogeneous high values, and patch no. 4 contains non-absolute homogeneous low values. It is observed that when patch no. 1 of very low absolute homogeneous value is introduced, AMOEBA does not have difficulty detecting patch 1. However, the patches of high values are not detected as distinct patches. On the other cases, when patch 1 of very high absolute homogeneous value is introduced, AMOEBA does not have issue detecting patch no. 1 again, but the patch of low value (patch no. 4) is now detected merely with

background cells to become statistically a low patch. A reverse pattern is observed when introducing varying absolute homogeneous values to patch no. 3 while patches no. 1 and 4 remain with non-absolute homogeneous low, and patch no. 2 remains with non-absolute homogeneous high (Figs. 10i through p). It is confirmed by these results that AMOEBA's performance is very sensitive to the data distribution of the whole data space.

Figure 11 presents the performance of AZP (Fig. 11a through d), ARISeL (Fig. 11e through h), and max-p-regions (Fig. 11i through m) for instances when patch no. 1 takes absolute homogeneous values of 10, 100, 130, and 170. Given that

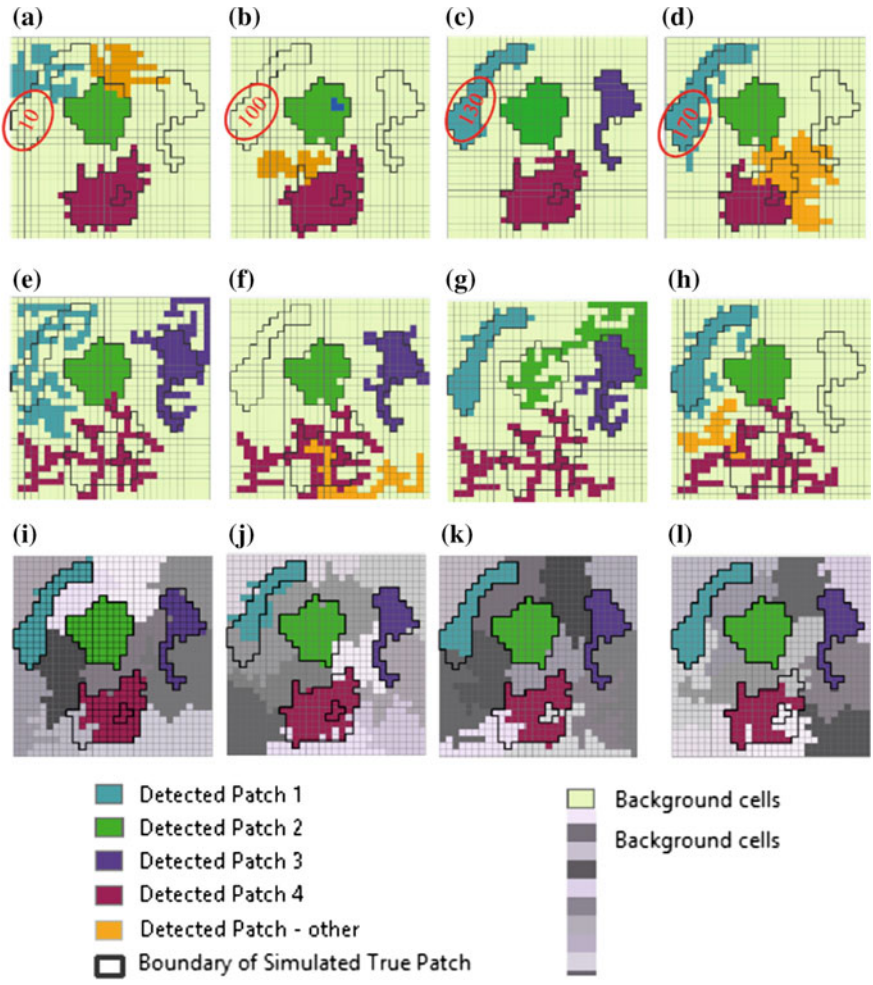


Fig. 11 Algorithm performance when patch no. 1 (**a** through **h**) takes the absolute homogeneous value of 10, 100, 130, and 170: (Fig. 10a through d) AZP, (Fig. 10e through h) ARISeL, and (**i** through **m**) max-p-regions

AZP and ARISeL are very similar algorithms, as reviewed in Sect. 2, their performances are very similar. Working based on a heuristic optimization search for the best boundary adjustment of the initial regions generated by K-mean++, they both require an accurate set-up parameter (the number of clusters was set to five in this study for these two algorithms). Yet they have the least competitive performance in this study as shown in Figs. 4, 5, 6 and 7. The max-p-regions algorithm, in contrast, performs much better than AZP and ARISeL thanks to its more sophisticated optimization technique. However, max-p-regions still misses the exact boundaries of patches in several instances. Finally, of the five algorithms, only SOM and AMOEBA are able to successfully detect clusters with holes, namely patch no. 4.

6 Conclusions

Five popular algorithms for spatial clustering including heuristic-based optimization approaches (AZP, ARISeL, and max-p-regions), a spatial statistical-based approach (AMOEBA), and a machine-learning method (SOM) were compared in this study for detecting homogeneous patches of different sizes, shapes, and values of homogeneity. Sixty-eight synthesis data sets were generated for testing. Tested data distributions included 3 distinct patches of high or low non-absolute homogeneous values and 1 patch of absolute homogeneous value. The absolute homogeneous value ranged from the minimum to the maximum value in the simulated data space. The algorithm performance measure was a function of the true positive, false positive, and false negative portions for respective clusters.

Among the tested algorithms, SOM and AMOEBA performed very well in detecting patches of different sizes, different shapes, including those with hole, and different homogeneous values. However, the test results suggest that performance of these two algorithms is very sensitive to the data distribution of the whole data space and to the homogeneous nature of patches including the values and the levels of homogeneity.

Three heuristic-based optimization approaches—including AZP, ARISeL, and max-p-regions—achieved a lower level of performance compared with SOM and AMOEBA in this study. They were very sensitive to the input parameters such as number of patches (for AZP and ARISeL) and floor constraint (for max-p-regions). With the parameters being set to their true simulated values, AZP and ARISeL performed noticeably worse than max-p-regions.

This study presents a prospective testing structure for all spatial-clustering algorithms aimed at detecting attribute-based homogeneous patches. The accurate detection of attribute-based homogeneous patches in geospatial databases is of significance because regions and patches are important constructs of geospatial analysis. They generalize raw and disaggregated data and establish meta-structures that may help in searching through large geospatial databases faster; they also constitute important building blocks in other methods of knowledge discovery and

machine learning such as spatial-association rule mining [32]. Thus, while the systematic evaluation reported here is important, more research is needed to fully gauge the effectiveness of various algorithmic methods across a more complete range of control conditions such as variance in levels of homogeneity, i.e., different internal-data distribution of homogeneous patches.

References

1. Openshaw S (1983) The modifiable areal unit problem. Geo Books, Norwich
2. Kwan M-P (2012) The uncertain geographic context problem. Ann Assoc Am Geogr 102:958–968
3. MacQueen JB (1967) Some methods for classification and analysis of multivariate observations. In: Proceedings of 5th Berkeley symposium on mathematical statistics and probability, University of California Press, pp 281–297
4. Kaufman L, Rousseeuw P (1990) Finding groups in data: an introduction to cluster analysis. Wiley, Hoboken
5. Yu D, Chatterjee S, Sheikholeslami G, Zhang A (1998) Efficiently detecting arbitrary-shaped clusters in very large datasets with high dimensions. Technical Report 98-8, State University of New York at Buffalo, Department of Computer Science and Engineering
6. Openshaw S (1977) Algorithm 3: a procedure to generate pseudo random aggregations of N zones into M zones where M is less than N. Environ Plan A 9:1423–1428
7. Duque JC, Church RL (2004) A new heuristic model for designing analytical regions. In: North American meeting of the regional science association international, Seattle, WA, November
8. Duque JC, Church RL, Middleton RS (2011) The p-regions problem. Geogr Anal 43(1): 104–126
9. Guo D, Peuquet D, Gahegan M (2003) ICEAGE: interactive clustering and exploration of large and high-dimensional geodata. Geoinformatica 7:229–253
10. Guo D (2008) Regionalization with dynamically constrained agglomerative clustering and partitioning (REDCAP). Int J Geogr Inf Sci 22:801–823
11. Anselin L (1995) Local indicators of spatial association—LISA. Geogr Anal 27:93–115
12. Getis A, Ord J (1992) The analysis of spatial association by use of distance statistics. Geogr Anal 24(3):189–206
13. Getis A, Aldstadt J (2004) Constructing the spatial weights matrix using a local statistic. Geogr Anal 36:90–104
14. Kohonen T (2001) Self-organizing maps, 3rd edn. Springer, Berlin
15. Bação F, Lobo V, Painho M (2004) Geo-self-organizing map (Geo-SOM) for building and exploring homogeneous regions. In: Egenhofer M, Miller H, Freksa C (eds) Geographic information science. Lecture Notes in Computer Science. Springer, Berlin, pp 22–37
16. Hagenauer J, Helbich M (2013) Contextual neural gas for spatial clustering and analysis. Int J Geogr Inf Sci 27:251–266
17. Han J, Kamber M, Tung AKH (2001) Spatial clustering methods in data mining: a survey. In: Miller HJ, Han J (eds) Geographic data mining and knowledge discovery. Taylor & Francis, London, pp 188–217
18. Duque J, Ramos R, Surinach J (2007) Supervised regionalization methods: a survey. Int Reg Sci Rev 30:195–220
19. Grubesic TH, Wei R, Murray AT (2014) Spatial clustering overview and comparison: accuracy, sensitivity, and computational expense. Ann Assoc Am Geogr 104(6):1134–1156
20. Rogerson P, Yamada I (2009) Statistical detection and surveillance of geographic clusters. Taylor & Francis Group, London and New York

21. Craglia M, Haining R, Wiles P (2000) A comparative evaluation of approaches to urban crime pattern analysis. Urban Stud 37(4):711–729
22. Murray AT, Grubesic TH (2013) Exploring spatial patterns of crime using non-hierarchical cluster analysis. In: Leitner M (ed) Crime modeling and mapping using geospatial technologies. Springer, Berlin, pp 105–124
23. Guo D, Wang H (2011) Automatic region building for spatial analysis. Trans GIS 15:29–45
24. Openshaw S, Rao L (1995) Algorithms for reengineering 1991 census geography. Environ Plan A 27:425–446
25. Duque JC, Dev B, Betancourt A, Franco JL (2011) ClusterPy: library of spatially constrained clustering algorithms, Version 0.9.9. RiSE-group (Research in Spatial Economics). EAFIT University. http://www.rise-group.org
26. Arthur D, Vassilvitskii S (2007) k-means++: the advantages of careful seeding. Proceedings of the eighteenth annual ACM-SIAM symposium on discrete algorithms. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, pp 1027–1035
27. Glover F (1977) Heuristic for integer programming using surrogate constraints. Decis Sci 8:156–166
28. Duque JC, Anselin L, Rey S (2010) The max-p region problem. Working paper, GeoDa Center for Geospatial Analysis and Computation, Arizona State University, Tempe, AZ
29. Aldstadt J, Getis A (2006) Using AMOEBA to create a spatial weights matrix and identify spatial clusters. Geogr Anal 38:327–343
30. Ord J, Getis A (1995) Local spatial autocorrelation statistics: distributional issues and application. Geogr Anal 27(4):286–306
31. Duque JC, Aldstadt J, Velasquez E, Franco JL, Betancourt A (2011) A computationally efficient method for delineating irregularly shaped spatial clusters. J Geogr Syst 13:355–372
32. Dao THD, Thill J-C (2016) The SpatialARMED framework: handling complex spatial components in spatial association rule mining. Geogr Anal 48:248–274

Part II
Maritime Traffic-Control Methods

Vessel Scheduling Optimization in Different Types of Waterway

Xinyu Zhang, Xiang Chen, Changbin Xu and Ruijie Li

Abstract With the continuous development of maritime-hardware systems, intelligent decision making in port transport becomes increasingly more important. There are three types of waterways in China, which is specific about one-way waterways, two-way waterways, and compound waterways. Modelling of vessel-scheduling optimization in ports aims at giving a concrete contribution to efficient, safe, and environmentally friendly maritime transport. The model will service to innovative the e-navigational shipping industry, which can lay the groundwork for a future vessel traffic service (VTS). These studies assess the effects of channel and berth resources to improve the operation efficiency of ports using a multi-objective vessel scheduling–optimization model. Taking the minimum total scheduling and waiting times of vessels, the model was established based on these waterways by considering safety, efficiency, and fairness. To solve the proposed model, a multi-objective genetic algorithm (MOGA) was used in this paper. Compared with the first-come, first-served (FCFS) scheduling method, the capacity-conversion times of vessels entering and leaving a port and the total scheduling time are decreased efficiently. These works contribute to improving the safety and optimization of ship scheduling catering to the intelligence trends of future maritime service portfolios (MSP).

Keywords Vessel scheduling · Optimization method · Genetic algorithm · E-navigation · Maritime service portfolios (MSP) · Waterways

1 Introduction

With the development of maritime-hardware systems, navigation technology is rapidly developing. It provides plenty of navigational aids and decreases the risk of marine accidents when the VTS, global positioning system (GPS), electronic

X. Zhang (✉) · X. Chen · C. Xu · R. Li
The Key Laboratory of Nautical Dynamic Simulation and Control of Ministry of Transportation, Dalian Maritime University, Linghai Rd.1, 116026 Dalian, China
e-mail: zhang.xinyu@sohu.com

V. Popovich et al. (eds.), *Information Fusion and Intelligent Geographic Information Systems (IF&IGIS'17)*, Lecture Notes in Geoinformation and Cartography, DOI 10.1007/978-3-319-59539-9_5

nautical chart (ENC), electronic chart display and information system (ECDIS), and automatic identification system (AIS) emerge nearby. With the development of an integrated assistant-navigation technology such as VTS, the duty recently changes from marine-regulator model to marine-information provider. The International Association of Lighthouse Authorities (IALA) first and officially promulgated the e-navigation concept. The concept means that e-navigation is the harmonized collection, integration, exchange, presentation, and analysis of marine information on board and ashore by electronic means to enhance berth-to-berth navigation and related services for safety and security at sea and protection of the marine environment. We can understand from this concept that e-navigation emphasizes maritime service. At the same time, the main function of VTS provides the marine-information service. Therefore, it is necessary to research the relevant function of VTS. At present, vessel quantities are huge and vessels should essentially wait in anchorage for a long time. The berth and fairway resources, however, may be hard to increase. However, the schedule strategy adopted by the VTS and the Port-Schedule Department is a simple schedule rule. The relevant schedule rules are the first-come, first-serve (FCFS) rule, the large-vessel first rule, and then to arrange the vessel to be inbound or outbound. Therefore, the Port-Schedule Department and the VTS need solutions to arrange the number of waiting vessels to decreases the waiting time and increase the berth or channel efficiency. At the same time, with large-scale dredging of the channel to increase the channel capacity and allow large draught vessels to transit the channel easily, the main seaports in China are now in the category of two-way waterways. The infrastructure is developing rapidly, but the management of using the one-way or two-way waterways is relatively lagging. This chapter, from the perspective of inbound or outbound ships, mainly considers how to improve the current situation and enhance the usage of the berth and channel. This research also laid a foundation about the next VTS and the Marine Service Portfolios (MSPs). Next, the literature is presented, and then the vessel scheduling in one-way and two-way waterways are detailed taken on, and the last part is the conclusion and the future work to be done.

2 Literature Review

Byeong-Deok [1] states that the different communication protocol and the data-exchange format of VTS can cause some problems in the overlap area. Therefore, they created a standard communication protocol and some supplement information. Seung-hee and Daehee [2] concluded that e-navigation focuses on information exchange and the marine-information service, not the connection and coordination of the security issue between vessels. Thus, they believe that e-navigation should have more emphasis on security. They researched the security-communication protocol and added the security-information service into the marine-service portfolios (MSPs). Byunggil Lee and Namje Park [3] researched the security architecture of the inter-VTS exchange-format protocol. They designed

a secure inter-VTS network security structure and security protocol based on the sharing of sensitive maritime data. Zhang [4] focused on e-navigation dynamic-information services such as the dynamic changes of basic geographic information, hydrological and meteorological information, water-transportation environmental information, and ship dynamic information. Huang and Chang [5] believe that maritime-safety information (MSI) with navigation display is a high-priority user need in the e-navigation survey, and they built a software module to send MSI messages through the ship's navigation sensor. Through the literature review, we found that one research hot point concerns how to use pertinent information more efficient. Therefore, this paper reports our studies under the MSPs framework to explore how to best use maritime-traffic information by the VTS. At present, the ACCSEAS4, the MONALISA5, and the Marine Electronic Highway in the Straits of Malacca and Singapore [6] are the most important test beds in the world, and these projects each have a certain forward-looking characteristics, to which the which the IMO (International Maritime Organization) is paying attention. The MONALISA maritime project is a comprehensive project to test some new technologies in the Baltic Sea. The project involves not only the research and application of advanced technologies, it also involves the research and trial of new methods as well as new models of maritime supervision and services. Regarding mathematical models, studies involve channel-usage efficiency, berth allocation, quay crane, and other resource scheduling. HSU [7] investigated berth distance, vessel size, vessel type, vessel draught, etc. as weighting factors and used these to propose an optimal sequence model for vessels entering and leaving port on a one-way channel. Based on the fundamental rules of the Strait of Istanbul, Ozgeca [8] developed a scheduling algorithm involving the relevant specific considerations to aid decisions on sequencing vessel entrances and giving way to vessel traffic in either direction. Jannes [9] defined the navigation lock–scheduling problem as a parallel-machine scheduling problem and adopted mixed integer linear programming methods as a solution. Berth, where vessels charge or discharge, is the most important resource in port operation. However, Jannes omitted the berth operation, which will decrease the berth-usage efficiency. In contrast, some scholars have performed studies on berth operation. For example, Imai and Nishimura [10] researched berth allocation in container ports. In addition, the efficiency of operations and processes on the ship-berth link has been analyzed using the basic operating parameters for Pusan East Container Terminal [11]. Jin [12] studied the quay-crane dynamic-scheduling problem based on berth schedules in a container terminal. The focus of these studies, however, was on berth operation; they did not consider channel-throughput time, which may lead to congestion on the channel. Reinhardt et al. [13] presented a model to minimize fuel consumption while retaining customer-transit times including the trans-shipment times per port-berth time. Song et al. [14] considered a joint tactical-planning problem for the number of ships, the planned maximum sailing speed, and the liner-service schedule in order to simultaneously optimize the expected cost, the service reliability, and the shipping emission in the presence of port-time uncertainty.

3 Scheduling Modeling of Vessel in One-Way and Two-Way Waterways

When one vessel leaves the last port to navigate to the next port, the vessel will send an estimated arrival report to the destination port. The Port-Schedule Department will receive the report to make a berth-work plan, and then the schedule department will send the plan to the vessel-traffic service (VTS) center to verify it. When the vessel approaches to the port boundary, it will send the arrival report to the VTS center. The VTS center will judge the vessel to enter the channel or anchor in the anchorage per the current channel situation. At the same time, for the vessel alongside the berth, the vessel will apply for leaving the port to the VTS center. The VTS center will plan the channel situation. The whole process is shown in Fig. 1. Considering the number of vessels, the tonnage of ships, and the special demand for the berth, it is difficult to make a perfect plan through the manual schedule. Thus, we need to build a model to reflect the real schedule-working.

3.1 Mathematical Model of Vessel-Scheduling Optimization in a One-Way Waterway

The one-way vessel scheduling problem in a fairway aims to determine the optimal sequence of arranging the fairway and the berth resource to decrease the total vessel

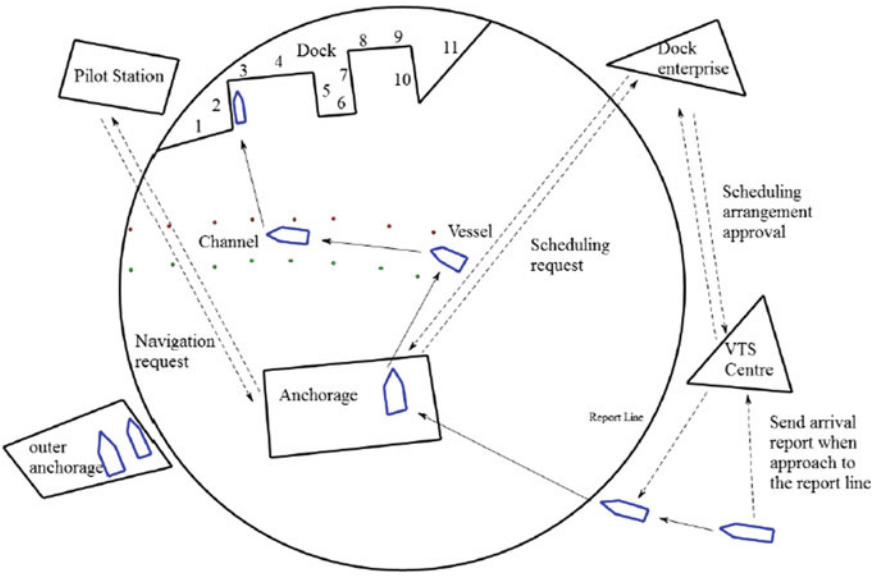


Fig. 1 Vessel-operation flow in a seaport

time in port while at the same time improving the port-service quality and efficiency. First, the optimization objectives are the total scheduling time and the total waiting time. The objective is that the total schedule time will represents vessels to complete inbound or outbound. It means the waiting time of each vessel to inbound or outbound. The port vessel schedule in the fairway (PVSF) can be formulated as follows:

$$\min[F, W] \tag{1}$$

$$F = \sum_{i=1}^{n} (f_i - f_{i-1}) + f_1 - b_0 - h_0 \tag{2}$$

$$W = \sum_{i=1}^{n} \left(b_i + t0_i - \frac{s + s_i}{v_i} \right) + \max(b_i - t0_i) - \min(b_i - t0_i) \tag{3}$$

Subject to:

$$b_i - t0_i - \frac{s_i}{v_i} > 0 \tag{4}$$

$$\left(b_i - b_0 - h - B_{ij}h_1\right) \times \left(1 - (IO_i - IO_0)^2\right) \geq 0 \tag{5}$$

$$\left(f_i - f_0 - h - B_{ij}h_1\right) \times \left(1 - (IO_i - IO_0)^2\right) \geq 0 \tag{6}$$

$$\left(b_i - f_0 - \tilde{h}\right) \times (IO_i - IO_0)^2 \geq 0 \tag{7}$$

$$f_i - b_i - p_i \geq 0 \tag{8}$$

$$p_i - \frac{s_i}{v_i} \geq 0 \tag{9}$$

$$\left(b_i - b_j + M \times R_{ij} - p_i - \tilde{h}\right) \times \left(IO_i - IO_j\right)^2 \geq 0 \tag{10}$$

$$\left(-b_i + b_j + M \times \left(1 - R_{ij}\right) - p_i - \tilde{h}\right) \times \left(IO_i - IO_j\right)^2 \geq 0 \tag{11}$$

$$\left(b_i - b_j + M \times R_{ij} - h - B_{ij}h_1\right) \times \left(1 - \left(IO_i - IO_j\right)^2\right) \geq 0 \tag{12}$$

$$\left(-b_i + b_j + M \times \left(1 - R_{ij}\right) - h - B_{ij}h_1\right) \times \left(1 - \left(IO_i - IO_j\right)^2\right) \geq 0 \tag{13}$$

$$\left(f_i - f_j + M \times R_{ij} - h - B_{ij}h_1\right) \times \left(1 - \left(IO_i - IO_j\right)^2\right) \geq 0 \tag{14}$$

$$\left(-f_i + f_j + M \times \left(1 - R_{ij}\right) - h - B_{ij}h_1\right) \times \left(1 - \left(IO_i - IO_j\right)^2\right) \geq 0 \tag{15}$$

where $i \in N^*$, and $i < j$.

The objective function (1) minimizes the total schedule time and the total waiting time in port. Constraint (4) states that the vessel-schedule time should be later than the time of the vessel arriving at the entrance of the fairway. Constraints (5) to (7) keep the safety time interval with each vessel. Constraints (8) and (9) state that the actual time is no less than the theoretical time for any vessel. Constraints (10) and (11) limit the safety time interval at the start time for two vessels such that it is no less than the theoretical time interval in different directions. Constraints (13)–(15) require that a vessel in the same direction must have a time-slot allocation. Constraint (16) constrains the problem when the berth is occupied.

Variables:

F the total time required by all vessels to complete the inbound and outbound processes;
W the total waiting time of all vessels to be inbound or outbound after correction;
b_i the starting time of vessel i to inbound or outbound;
f_i the ending time of vessel i to inbound or outbound;
b_0 the starting time of the last vessel in the previous stage to inbound or outbound;
f_0 the ending time of the last vessel in the previous stage to inbound or outbound;
p_i the theoretical time needed for a vessel to pass the fairway;
s_i the theoretical distance of vessel i from the entrance of the fairway to the berth;
v_i the average speed of vessel i from the entrance of the fairway to the berth;
$t0_i$ the time needed by vessel i to cross the report line;
IO_i a binary that = 1 only if the ship enters the port; otherwise, it = 0;
R_{ij} a binary that = 1 only if ship j is later than ship i inbound or outbound; otherwise, it = 0;
B_{ij} a binary that = 1 only if the ship moors the near distance berth when ship i is scheduled earlier than vessel j; otherwise, it = 0; and
$Berth_{im}$ a binary that = 1 if berth m is occupied; otherwise, it = 0

Parameters

s the distance from the report line to the entrance of the fairway;
h the safety time interval of vessels in the same direction;

$\tilde{h}$ the safety time interval of vessels in different directions;
M a fixed constant number to ensure that the inequality is valid in the case where the conditions meet the requirement.

3.2 Vessel Scheduling in a Two-Way Waterway

The modeling of a two-way waterway is similar to that of a one-way channel. The big difference is that the two-way channel can be transformed into one-way channel under some irregular conditions such as visibility, wind force, and vessel properties. Considering these special requirements, the traffic-transformation model should be added into the two-way waterway scheduling. When under the following circumstances, the vessel only sails using a one-way navigation pattern:

- The vessel's beam is >50 m;
- The vessel's beam is >25 m, and the vessel is carrying these materials:
  - Carrying 100-tonnage explosive materials;
  - Carrying A or B chemical materials of the International Convention for the Prevention of Pollution From Ships (MARPOL) annex II;
  - Carrying liquefied gas.
- The total ship length of the inbound and outbound vessels is >86 m; if one vessel carries goods in second items, the total length of vessels is >45 m;
- The wind force is >7 °C;
- The visibility is <3000 m; and
- The ice conditions are serious.

Under these conditions, the two-way waterway should be transformed into one-way waterway. These are the distinguished features and special requirements in the two-way waterway.

The other difference of a two-way channel is that the safety requirements have a lateral interval compared with the one-way waterway as shown in Fig. 2.

Figure 3 shows four navigational pattern converts. Except for the one-way/one-way navigational pattern and the two-way/two-way navigational pattern, there are two navigational-pattern converts, which are the one-way/two-way navigational pattern and the two-way/one-way navigational pattern. In addition, the inbound or outbound direction of the navigation model switch and the berth distance (near or far the ending point of fairway) must also be considered.

3.3 Vessel Scheduling in a Compound Channel

As of January 1, 2014, the compound channel has been working in Tianjin Seaport, which is the largest dredged waterway in the north of China. With the increasing

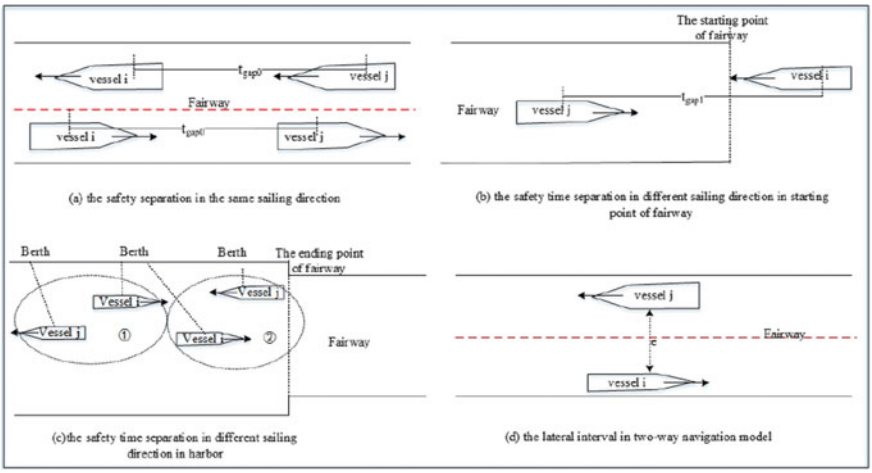


Fig. 2 Safety restriction in a two-way waterway

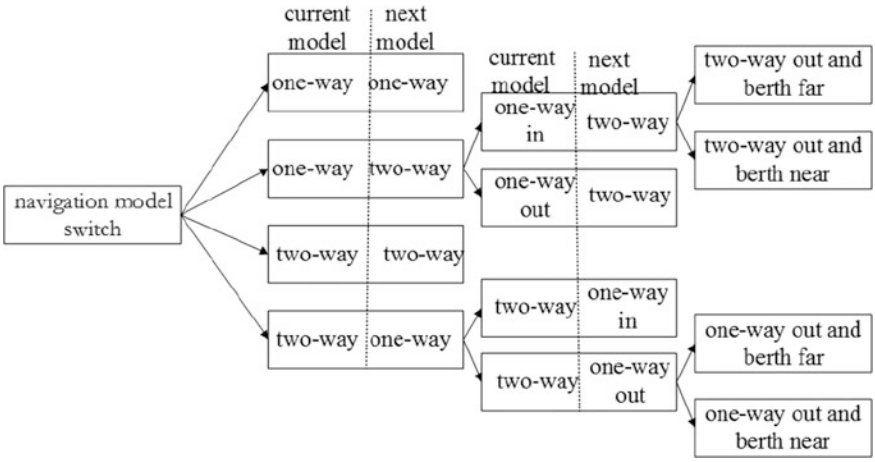


Fig. 3 Navigational-model convert

traffic flow, the two-way waterway in Tianjin has not been able to satisfy the demanding needs. Through field researches, investigators have found that a main channel (two-way waterway) is not in high-level use due to the large number of non-shipping vessels (such as patrol vessels, working boats, and others) occupying the channel.

Therefore, two waterways along the main channel have been established shown in Fig. 4. The compound channel is more complex than the two-way waterway, and there are two main traffic flows when vessels enter the port. One traffic flow comprises the vessels entering the Dongjiang port area (Fig. 5 [left]), and the other

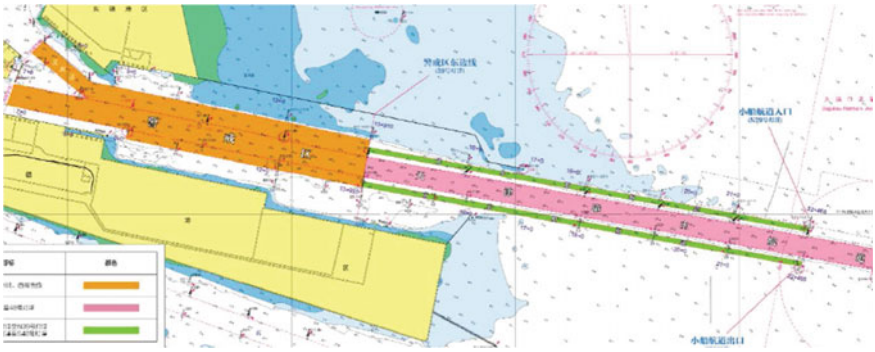


Fig. 4 Compound channel

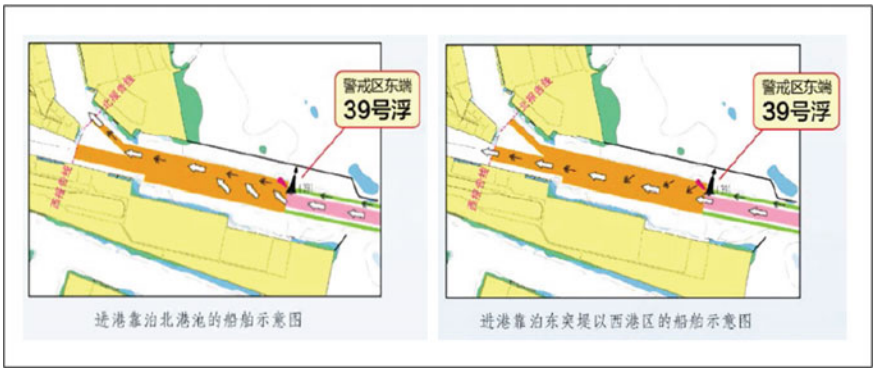


Fig. 5 Traffic flow entering the Dongjiang Port area

traffic flow comprises vessels entering the west end of the east pier area (Fig. 5 [right]). When vessels sail into the no. 39 buoy, the traffic flow of small vessels as well as that in the main channel should cross into the precaution area, for which the details are in Fig. 5. The scheduling challenge is knowing how to manage and control traffic flows to decrease cross-over frequency. This is now our main difficult task.

4 Case Study of One-Way Waterway Scheduling

The vessel information from the VTS center is listed in Table 1. The eight Pareto optimal solutions are listed in Table 3. With respect to total scheduling time, the optimal solution is 2.6186 h. At present, the fitness value of the total waiting time is

Table 1 Vessel information

Vessel	Direction	Length (m)	Berth no.	Speed (kn)	Sail (no. of miles)	Start time
1	In	148	3	8	1.5	7:08
2	In	142.7	6	14	1.8	7:01
3	In	84.4	1	8.8	1.3	7:07
4	Out	144.8	5	9	1.7	7:18
5	Out	178.7	14	8.3	2.6	7:04
6	In	241.3	12	7.1	2.4	7:40
7	In	115.8	9	11.2	2.1	7:56
8	Out	97.2	13	9.5	2.5	7:49
9	Out	150.4	6	15.5	1.8	7:36
10	In	88.8	13	7.7	2.5	7:23
11	In	130	17	10.7	2.9	7:24
12	Out	108.4	10	11.7	2.2	8:29
13	In	142.7	5	11.5	1.7	7:12
14	In	72	2	9.6	1.4	7:47
15	In	166.2	7	12.3	1.9	7:38
16	In	147.5	11	10.9	2.3	7:09
17	In	334.1	14	12.9	2.6	7:03
18	In	85.5	10	8.6	2.2	7:01
19	Out	145.1	3	16.5	1.5	7:11
20	In	95.2	4	7.8	1.6	7:00

24.7768 h. Regarding total waiting time, the optimal solution is 23.0124 h. Presently, the total scheduling time is 2.0168 h. This is in compliance with the features of the MOGA results, i.e., when one objective is improved, another objective will be correspondingly weakened. Because the vessel preparation–scheduling time is more concentrated, it can increase the time period for vessel scheduling but also prolong vessel-waiting time at the same time. It is known that both vessels no. 1 and 19 are docking at berth no. 3. According to the berth conflict–resolution constraint, vessel no. 19, which is going to leave the port, must be scheduled to depart before vessel no. 1 enters the port. From Table 2, among the eight Pareto optimal solutions, scheduling schemes obtained after berth-conflict resolution all schedule the outbound vessel before scheduling the entering vessel, which is a reasonable solution.

With respect to optimal-scheduling efficiency for a one-way fairway, because vessels on the fairway cannot encounter each other, it must be ensured that there is no vessel present before making the flow conversion, thus resulting in more time consumed. Therefore, theoretically, for a long one-way fairway, the fewest flow conversions will result in, the shortest time to complete the vessel scheduling.

Table 2 The Pareto optimal solutions

No.	First objective value	Second objective value	Scheduling schemes
1	2.6186	24.7767	5 4 13 15 6 17 11 19 9 8 12 2 7 18 10 3 14 1 20 16
2	2.6398	23.8463	5 4 13 15 16 17 11 3 19 9 8 12 2 6 10 14 1 20 7 18
3	2.6589	23.5049	5 4 13 15 16 17 11 7 19 9 8 12 2 18 10 3 14 1 20 6
4	2.6349	23.9836	5 4 13 15 16 17 11 20 19 9 8 12 2 6 10 3 14 1 7 18
5	2.6675	23.1771	5 4 13 15 16 17 11 20 19 9 8 12 2 18 10 3 14 1 7 6
6	2.6765	23.1661	5 4 19 15 16 17 11 14 1 9 8 12 2 18 10 3 20 13 7 6
7	2.6455	23.8244	5 4 19 15 16 17 11 14 13 9 8 12 2 6 10 3 1 20 7 18
8	2.6985	23.0124	5 4 19 15 16 17 11 14 20 9 8 12 2 18 10 3 1 13 7 6

Table 3 Vessel-scheduling solution 1

Pareto 1	Pareto 8						
Schemes	Direction	Start time	Finish time	Schemes	Direction	Start time	Finish time
5	Out	7:07	7:30	5	Out	7:07	7:30
4	Out	7:19	7:34	4	Out	7:19	7:34
13	In	7:36	7:45	19	Out	7:25	7:38
15	In	7:39	7:49	15	In	7:40	7:49
6	In	7:43	8:03	16	In	7:43	7:55
17	In	7:52	8:14	14	In	7:48	8:03
11	In	7:56	8:20	11	In	7:51	8:07
19	Out	8:22	8:27	14	In	8:07	8:16
9	Out	8:24	8:31	20	In	8:09	8:21
8	Out	8:25	8:41	9	Out	8:23	8:30
12	Out	8:29	8:43	8	Out	8:24	8:40
2	In	8:45	8:53	12	Out	8:29	8:42
7	In	8:46	8:58	2	In	8:44	8:52
18	In	8:48	9:03	18	In	8:46	9:01
10	In	8:50	9:09	10	In	8:48	9:07
3	In	9:07	9:17	3	In	9:05	9:15
14	In	9:09	9:20	1	In	9:09	9:20
1	In	9:13	9:24	13	In	9:12	9:25
20	In	9:15	9:27	7	In	9:15	9:31
16	In	9:19	9:37	6	In	9:21	9:41

Therefore, this model decreases the vessel-flow conversions nine-fold and thereby greatly improves the efficiency of vessel scheduling. The scheduling schemes of Pareto 1 and Pareto 8 are listed in Table 3. Each Pareto-optimal solution represents a scheduling scheme. When vessels with different directions encounter each other, the later scheduled vessel is scheduled after the earlier scheduled vessel with a different direction when it has completed its travel to ensure the safety of vessels in the scheduling operation. As seen in Table 3, among the Pareto 1 optimal solutions, the completion time of the no. 4 outbound vessel is 7:34 am and the start time of the no. 13 entering vessel is 7:36 am, thus ensuring the safety of the vessels and maintaining compliance with the previously mentioned requirements. Eight Pareto optimal solutions meet the previous requirements. Additionally, according to the FCFS rule, the total scheduling time of 20 vessels is 4.967 h, and the total waiting time is 49.5 h. Compared with the FCFS rule, the eight Pareto optimal solutions obtained in this model can decrease, at the most, 47.2% (Pareto 1) of the total scheduling time and at least 45.6% (Pareto 8) of the total scheduling time. In terms of practical applications, if the port pursues the shortest overall scheduling time, it can select the scheduling scheme value of Pareto 1 with the first smaller objective value. If the waiting time for each vessel is taken into account, it can select the scheduling scheme of Pareto 8 with the second smaller objective value. Given these analysis, the decision-makers select the scheduling schemes according to their needs.

5 Conclusions and Future Work

This chapter analyzed the mechanism of the one-way channel to build a schedule model and calculate the model by the genetic algorithm to obtain the schedule information. The case study shows that the results can fit the real demand. The model can also integrate with the e-navigation aspect in the MSP framework to service the harbor and the vessel. In the future, we will add the berth-allocation model to make the port-vessel schedule more reasonable and practical. We may consider that some vessels may wait the tide to satisfy the draft to pass the fairway, and a tug-boat allocation may be considered. We will research the two-way fairway and compound fairway to build a related mathematic model to develop scheduling software.

Acknowledgements This work was financially supported by the National Natural Science Foundation of China (Grant No. 51309043), the Fundamental Research Funds for the Central Universities (Grant No. 3132016315), the Outstanding Young Scholars Growth Plan of LiaoNing Province (Grant No. LJQ201405), a basic research project of Key Laboratory of Liaoning Provincial Education Department (Grant No. LZ2015009), and the Natural Science Foundation of Liaoning Province (Grant No. 2015020626).

References

1. Byeong-Deok YEA (2009) Introduction of the Standardization in VTS data exchange. In: Proceedings of Asia navigation conference
2. Oh S, Seo D, Lee B (2015) S3(Secure Ship to Ship) Information sharing scheme using ship authentication in the e-navigation. Int J Secur Appl 9:97–110
3. Lee B, Park N (2012) Seurity architecture of inter VTS exchange format protocol for secure e-navigation. Embedded Multimedia Comput Tech Serv 181:229–236
4. Zhang A (2013) Research on some key techniques of dynamic information services in e-Navigation. Wuhan
5. Huang C, Chang S (2013) Onboard integration of maritime safety information. In: 12th International Conference on ITS Telecommunications
6. IALA Gideline NO. 1107 on The Reporting of Results of e-Navigation Testbeds, Dec 2013
7. Xu G, Guo T, Wu Z (2008) Optimum scheduling model for ship in/outbound harbor in one-way traffic fairway. J Dalian Marit Univ 24(4):150–153,157
8. Ozgecan S, Tutun U (2008) Performance Modeling and risk analysis of transit vessel traffic in the Istanbul strait: Studies on queues with multiple types of interruptions. The State University Of New Jersey, Rutgers
9. Verstichel J, De Causmaecker P, Berghe GV (2011) Scheduling algorithms for the lock scheduling problem. Procedia-Soc Behav Sci 20:806–815
10. Dragovic B, Park NK, Radmilovic Z (2006) Ship-berth link performance evaluation: simulation and analytical approaches. Marit Policy Manage: Flagship J Int Shipping Port Res 33(3):281–299
11. Pachakis D, Kiremidjian AS (2003) Ship traffic modeling methodology for ports. J Waterw Port Coast Ocean Eng 129(5):193–202
12. Li P, Sun J, Han H (2006) The algorithm for the berth scheduling problem by the hybrid optimization strategy GASA. J Tianjin Univ Tech 22(4):58–61
13. Reinhardt LB et al (2016) The liner shipping berth scheduling problem with transit times. Transp Res Part E Logistics Transp Rev 86:116–128
14. Song DP, Li D, Drake P (2015) Multi-objective optimization for planning liner shipping service with uncertain port times. Transp Res Part E Logistics Transp Rev 84:1–22

An Intelligent GIS–Based Approach to Vessel-Route Planning in the Arctic Seas

Misha Tsvetkov and Dmitriy Rubanov

Abstract The Northern Sea Route along the Arctic coast of Russia offers a good alternative to the Suez Canal (due to the huge reduction in distance from Western Europe to East Asia) or Panama Canal and can also be used for supporting oil and gas industry in the Arctic. However, navigation in the Arctic seas is sufficiently complicated by extreme climate and ice conditions and the general lack of the marine infrastructure throughout the Arctic Ocean. This chapter presents an approach to vessel-route planning in the Arctic seas by means of special system, intelligent GIS. Ontology, expert subsystem, and some other Artificial-Intelligence technologies are proposed as basic ones for the system implementation. The application of the technique is illustrated by way of case study of situational route planning for a vessel navigating along the Northern Sea Route.

Keywords Northern sea route · Geographic information system · Vessel-route planning · Ontology · Situation assessment

1 Introduction

During the last decade, there has been increasing interest in the use of the Northern Sea Route for marine transportation. The route along the Arctic coast of Russia not only offers an alternative route to the Suez Canal (due to the huge reduction in distance from Western Europe to East Asia) or Panama Canal, it can also be used for supporting the oil and gas industries in the Arctic.

However, the opportunities afforded by Arctic navigation can be greatly lessened by some specific aspects. The region above the Arctic Circle is characterized by an

M. Tsvetkov (✉) · D. Rubanov
SPIIRAS Hi Tech Research and Development Office Ltd (SPIIRAS-HTR&DO Ltd),
St. Petersburg, Russia
e-mail: tmv@oogis.ru

D. Rubanov
e-mail: rubanov@oogis.ru

V. Popovich et al. (eds.), *Information Fusion and Intelligent Geographic Information Systems (IF&IGIS'17)*, Lecture Notes in Geoinformation and Cartography, DOI 10.1007/978-3-319-59539-9_6

extreme climate with short summers and long polar nights and extensive snow and ice cover. Average summer temperatures range from approximately −10 °C to +10 °C, and winter temperatures can drop below −40 °C over large parts of the Arctic. The average amount of storms in the Arctic Ocean can reach 22 with a duration ≤ 9 days. Some regions of the Arctic Ocean remain ice-covered for much of the year, and nearly all parts of the Arctic experience long periods with some form of ice on the surface. For example, some parts of the Northern Sea Route are covered with ice for >7 months a year. The thickness of ice can reach 2 meters, and the thickness of ice hummock can reach 6–8 m.

Trans-Arctic navigation is also complicated by the general lack of the marine infrastructure throughout the Arctic Ocean. Comparison of the Arctic and mid-latitude routes shows that existing Atlantic and Pacific routes have a rich historical experience of navigation, detailed and constantly updated navigation nautical charts, and a large number of navigational aids (approved, fairways, lighthouses, lanterns, buoy systems, etc.). The situation in the Arctic is different. Although the Russian Federation has an extensive system of fixed and floating aids for navigation, the possibilities of the system are not sufficient for safe marine navigation in the region.

At present, many government and commercial organizations use a large number of marine traffic–monitoring systems, decision-support systems, and navigation systems to solve the common safety tasks of marine navigation. However, such systems are not always able to solve specific problems related to safe navigation in the Arctic.

This chapter describes an approach to situation awareness, which allows supporting safe navigation in the Arctic. The proposed approach is based on GIS technology, intelligent analysis tools, and special-purpose mathematical models (including forecasting). The authors argue that the use of intelligent GIS technologies allows for flexibly adapting to the conditions of the Arctic seas and solving a wide range of specific tasks.

The outline of the chapter is as follows. The next section discusses the related works. This is followed by the basic approach for vessel-route planning in the Arctic seas in Sect. 3. Section 4 presents a description of intelligent GIS overview and intelligent GIS architecture. In Sect. 5, we introduce an ontology-based approach to situation awareness for vessel-route planning in the Arctic. The next section describes a case study of intelligent GIS–based vessel-route planning along the Northern Sea Route. Conclusions are presented in the final section of this chapter.

2 Related Works

In this section, we present a brief overview of the existing works concerning safe marine navigation, vessel-route planning, and situation awareness, especially in the Arctic region.

In papers [1, 2], the authors discuss various aspects of safe navigation in the Arctic seas. Although these authors give the main attention to changing ice situation (due to climate-warming trends over the past century), the challenges for ships navigating in Arctic waters are discussed. However, the authors [1, 2] do not consider safe marine-navigation issues in the Arctic.

Paper [3] gives an example of a safe navigation system based on GIS technology. It is necessary to note that the paper is dedicated to the problem of ship-borne system development. The main attention is given to the issues of navigation monitoring in a vessel's location area and the system's functioning in real-time mode. An integration of the solution suggested in [3] with an added intelligent navigation system is presented as one of the new lines of the research.

Paper [4] describes the capabilities of modern GIS for "intelligent decision making" in information systems associated with the safety of maritime activities. The authors pay close attention to the perspectives of cooperation between different organizations that possess important information about various fields of maritime activities. In addition, aspects related to training personnel to operate monitoring and safe navigation systems are considered in this work.

Paper [5] presents the development of intelligent GIS for maritime-situation monitoring in detail. Also, this work focuses on issues of integration, harmonization, and fusion of data obtained from various sources (sensors). In the paper, an ontology approach is offered as a basic one for IGIS architecture.

Although papers [4, 5] present a general approach to intelligent GIS for maritime-situation monitoring, the works do not address the important issues of safe navigating and vessel-route planning in the Arctic seas.

In paper [6], the authors discuss how intelligent GIS may be used for scenario-based geospatial processes description of dangerous situations in different subject domains.

In paper [7], the authors focus on the description of the ontology-based situation-awareness approach based on a situation-theory concept of situation and ontology-based situation awareness. The piece of information according to this theory is called an "infon." Infons are written as follows:

$$\ll R, a_1, \ldots, a_n, 0/1 \gg,$$

where R is n-place relation; $a_1, \ldots, a_n$ are objects peculiar to R; and 0/1 is the polarity of the infon. The polarity value 1 means that, in the situation s, objects $a_1, \ldots, a_n$ stand in the relation R. The value 0 means that, in the situation s, objects $a_1, \ldots, a_n$ do not stand in the relation R.

Infons of a subject domain consist of objects, their properties, and relations existing in the given subject domain. Relation between situations and infons write as follows:

$$s \models \sigma,$$

which means that infon σ is made by the situation s. The official terminology is that s supports σ. In paper [7], the authors define a situation in terms of infons and also suggest using rule-based inference for situation assessment.

3 A General Approach to Vessel-Route Planning in the Arctic Seas

In the general case, the procedure of vessel-route planning can be divided into the following steps:

Step 1: Choice of the all possible safe-navigation points in the region.

The first step is usually assisted with properly updated navigation nautical charts and supposes that the sea region in the interest has to be "covered" with so-called "possible route points" or approved, already checked sea lanes (or fairways). Each point has coordinates (latitude, longitude) and may be used as a node for some route; two points are chosen as start and end points. Then the vessel route can be defined as a path in directed or undirected geospatial graph where points are nodes connected with arcs (Fig. 1). In Fig. 1, points that cannot be used as route nodes, are marked with black fill. In practice, node positioning and binding can be performed with geographical information system (GIS) and digital navigation chart.

Step 2: Choice of possible safe navigation points in the region with regard to weather and ice information.

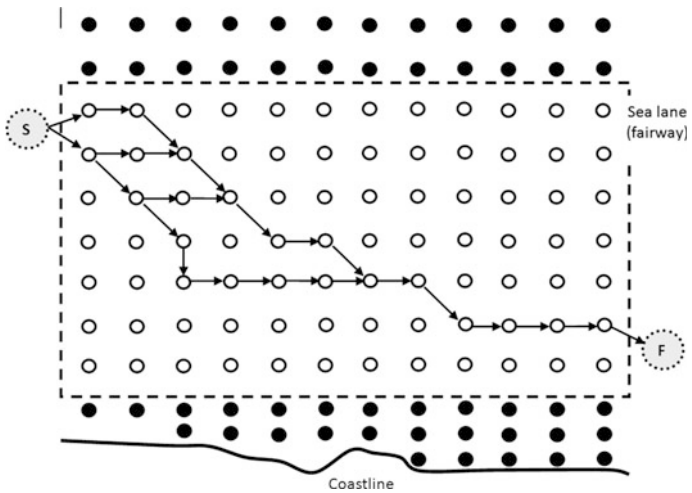


Fig. 1 Vessel route and geospatial graph (safe navigation points)

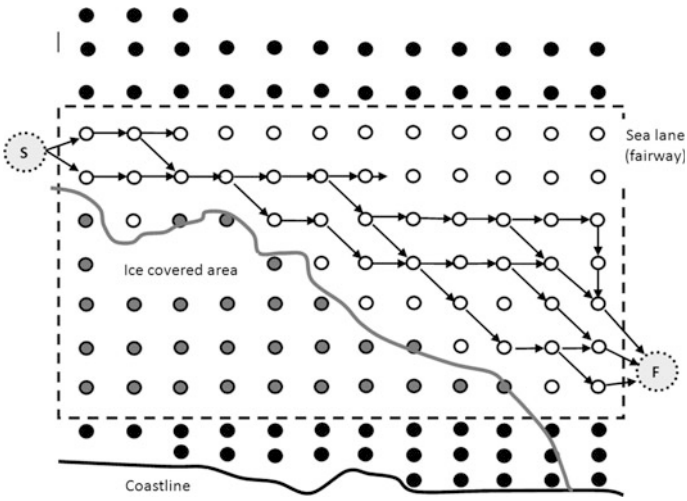


Fig. 2 Vessel route and geospatial graph (with regard to ice situation)

Safe marine navigation in the Arctic depends on accurate and timely weather and ice information to help to minimize the risk of accidents in this challenging environment.

Some of the possible vessel route points can be navigation dangerous ones because of the weather or ice situation in the region. For example, according to a navigation map a point can be used for vessel-route planning, but due to wind and wave forecasts the using of the point as a route node is not reasonable. In Figs. 2 and 3, these points are marked with gray fill.

Step 3: Choice of optimized vessel route.

The vessel owner or captain often has the problem of determining the optimized vessel route based on different criteria. It is clear that the above-mentioned graph may be considered as weighted, so each arc has its own positive weight (Fig. 4). The cost of the route's arcs may have the following meanings:

- distance between nodes (the simplest case),
- fuel cost, and
- cost of the passage through some route parts, etc.

Furthermore, the problem may be reduced to the shortest path problem (or the linear programming problem), and the optimized vessel route can be found in polynomial time (for example, by means of Dijkstra's algorithm).

It is readily seen that all of the three steps are easy in theory but difficult in practice. The first problem is in the choice of the safe navigation points for our future route. In mid- and low-latitudes, ships often go on standard and proven routes, and route planning can be performed simply by selecting one of the well-known routes (or sea lanes along the most part of the vessel's route). For some

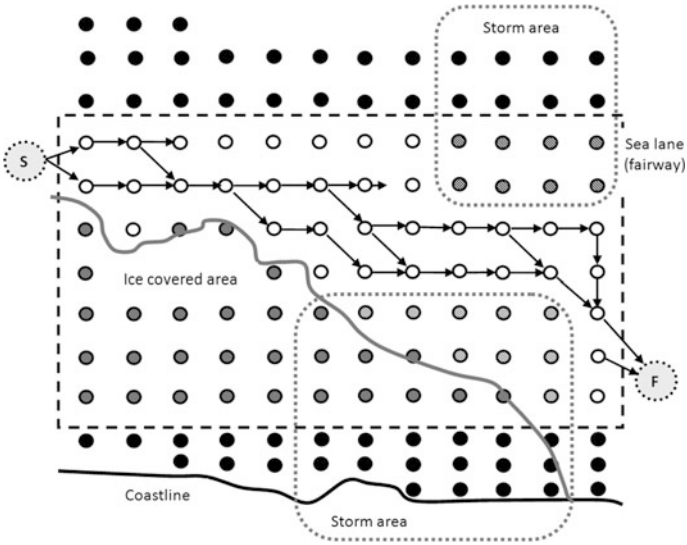


Fig. 3 Vessel route and geospatial graph (with regard to weather information)

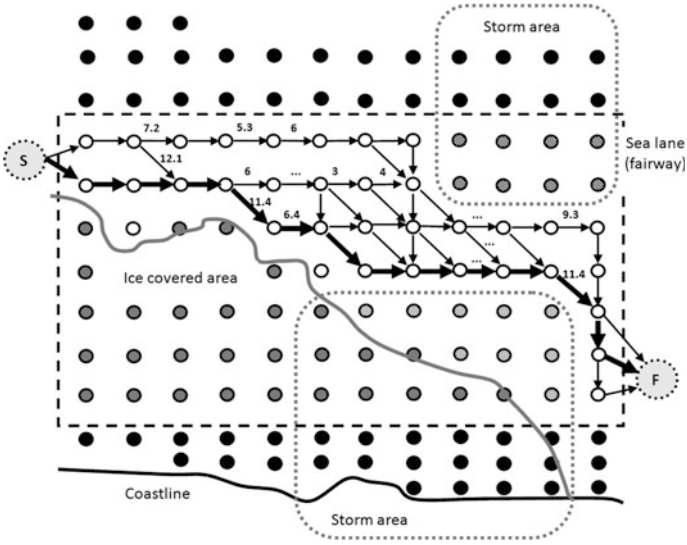


Fig. 4 Choice of the optimized vessel route

parts of the Northern Sea Route, there exists no standard routes; planning of the route also depends on current ice and weather conditions. Moreover, many high-risk areas in the Arctic are inadequately surveyed and charted, which means that some charts available to mariners may not be current or reliable.

The other problem consists of the necessity to correct a planned route when our vessel is already on the route. The necessity can be explained by fast-changing ice and weather conditions in the Arctic Ocean, the need to use icebreakers for the parts of the route, etc.

The authors suppose that in order to handle the problems, new techniques must be used. One of the most perspective approaches to solving the problem is to use GIS technologies with Artificial intelligence and situational-planning methods.

4 Intelligent GIS

GIS provides a powerful set of tools for geospatial data analyzing in different subject domains [5, 8]. Through the use of GIS technology we can solve various complex problems such as the following:

- monitoring of different situations (maritime, air, on-ground situations) in real time,
- intelligent-decision support and situation assessment,
- spatial process–trends prediction,
- complex geospatial-processes modeling, and
- visualization of modeling results.

However, a great number of problems cannot be resolved with traditional approaches realized in common GIS.

Technological advances in Artificial Intelligence (AI) are providing new tools with potential application in spatial-data analysis. These technologies not only promise to add accuracy and speed to decision making and geospatial-process prediction, they also to provide a GIS with a more user-friendly and intellectual interface.

Some of the main methods and technologies that can be used in GIS intellectualization include the following:

- intellectual data processing,
- intellectual data mining,
- expert systems, and
- machine learning.

Utilizing methods and technologies of AI greatly improves the efficiency of GIS and allows us talk about so-called intelligent GIS (IGIS). In this chapter, IGIS is considered as a type of GIS that includes integrated tools and/or systems of Artificial Intelligence [5].

Usually IGIS includes the following basic components [5, 9]:

- a knowledge-based system (ontology),
- an expert system and inference machine,
- a visual environment for developing classes and objects of a subjects domain,

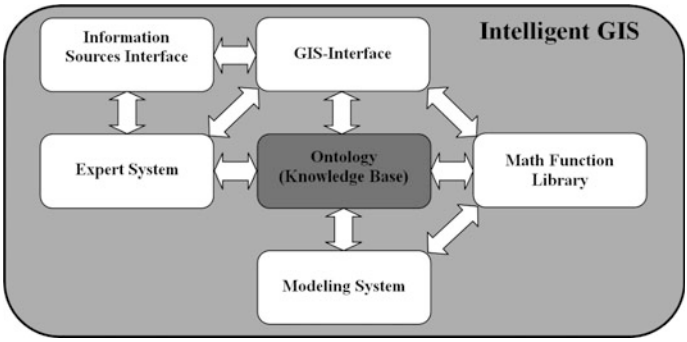


Fig. 5 Intelligent GIS architecture

- a visual environment for developing models of objects, and
- a system for scenario implementation in real-time using a special visual-development environment.

According to Fig. 5, IGIS includes the following components: knowledge base (ontology), information-sources subsystem (interface), expert system, modeling system, math function library, and GIS interface.

The central part of IGIS is a knowledge base (ontology). Ontology provides the "framework" for the representation of concepts and relations between them in the subject domain. Another part of the knowledge base is the storage of subject domain object instances, which can later be used for real-time situation assessment.

The information-sources interface performs the communication between sensors (radars, sonars, etc.) and IGIS. In addition, we consider the following information system as information sources:

- vessel traffic–monitoring systems,
- remote-sensing systems,
- hydro-meteorological and ice-conditions data base (including operative data-bases), and
- hydro-acoustic monitoring system of ice conditions.

The GIS interface provides a set of interfaces for interaction with other IGIS components. GIS Interface is used for the following purposes:

- displaying IGIS incoming data in various variants;
- visualizing results of spatial processes analysis, prediction and modeling; and
- choosing combinations of algorithms for modeling.

The Math-Function Library is an important part of IGIS. The library includes the set of special mathematical functions and models that can be used by any IGIS subsystem. The set of functions supports extension and variability. For example, for modeling spatial processes associated with a dangerous situation the following functions and models from the library can be used:

- mathematical models of different dangerous situations (e.g., oil spill–area assessment, weather conditions–area assessment, seizure of vessel by pirates, terrorist, etc.);
- vessel navigation along the predefined route modeling;
- searching algorithms (for maritime rescue operation) and etc.

Expert system is a rule-based system that uses human expert knowledge to assist decision-makers in developing responses to dangerous-situation risks. Expert knowledge is described in the form of producing rules and saved in the IGIS knowledge base (ontology). It is necessary to mention that any of the mathematical functions from the Math-Function Library can be used in rules.

The IGIS-modeling system provides a powerful tool for computer modeling of various geospatial processes. The important part of modeling system is an interactive visual environment for scenario development. Mathematical models and functions from the Math-Function Library are also the basis for the IGIS modeling system.

5 Situation Planning and Correction of a Pre-planned Vessel Route

Situation planning of a vessel route is based on the deep understanding of situations and their possible development, thus allowing to make more reasonable decisions as well as predict dangerous maritime situations and to take timely action to prevent them.

The authors suppose that in order to handle the problems, new techniques must be used. One of the most perspective approaches to solving the problem is to use GIS-technologies with Artificial Intelligence and situational-planning methods.

In this chapter, the basis for situation vessel-route planning is the situation-awareness model proposed by Endsley [10]. Situation awareness is understood as the "perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" [10].

The proposed situation-awareness model includes three levels [10]:

- Level 1 (perception of the elements in the environment): perception of the current situation, its properties, and dynamic development of elements related to the observed situation in the environment.
- Level 2 (comprehension of the current situation): synthesis of the disconnected elements of the situation perceived in level 1, comprehension and processing of information, integration of various heterogeneous data, and determination of its significance for particular situation.
- Level 3 (projection of future): available prognosis of future actions and future situation development, based on knowledge of the situation status and the development of its elements, for taking decisions on future actions beforehand.

For effective use of situation awareness, an integrated approach based on ontologies and IGIS-technologies is applied.

As in paper [7], we try to define possible dangerous situations in terms of infons and apply rule-based inference for situation assessment. Objects for the situation planning of vessel routes are ships (cargo vessels, tankers, etc.), ice-covered regions, weather conditions, and such areas as oil-spill area, areas that are closed to navigation area, and others. Vessels make a voyage from a starting point to a final point. On all waterways, there can be various dangerous situations influencing vessel safety. In this chapter, we consider only three dangerous situations: influence of the ice situation, weather conditions on the vessel route, and a closed area due to an oil spill.

Let us consider typical situations that are common for navigation in the Arctic seas. The first dangerous situation we call "IceDanger." Status of the ice cover affects vessels' velocity and increases travel time on a planned route. In addition, the vessel's velocity is reduced due to the danger of damage to the hull inflicted by ice. Mathematically, this information can be presented by the following relation tuples: near(Ice, Vessel), clash(Ice, Vessel), and threat(Ice, Vessel). By means of infons, these relations can be written as follows:

$$
\begin{aligned}
&IceDanger| = \ll location, Ice, L\_Ice, 1 \gg,\\
&IceDanger| = \ll velocity, Ice, V\_Ice, 1 \gg,\\
&IceDanger| = \ll location, Vessel, L\_Vessel, 1 \gg,\\
&IceDanger| = \ll velocity, Vessel, V\_Vessel, 1 \gg,\\
&IceDanger| = \ll near, Ice, Vessel, 1 \gg,\\
&IceDanger| = \ll clash, Ice, Vessel, 1 \gg,\\
&IceDanger| = \ll isSafety, Ice, Vessel, 0 \gg,\\
&IceDanger| = \ll threat, Ice, Vessel, 1 \gg .
\end{aligned}
$$

Here *L_Ice* is ice location; *L_Vessel* is vessel location; *V_Ice* is ice velocity, and *V_Vessel* is vessel velocity.

Another situation that occurs during vessel-route planning in the Arctic seas is the bad weather conditions (meteorological information) affecting navigation (WeatherDanger). This situation is similar to the IceDanger situation. Infons and polarities for this situation can be written as follows:

$$
\begin{aligned}
&WeatherDanger| = \ll isWind, Weather, Vessel, 1 \gg,\\
&WeatherDanger| = \ll isHeavy, Weather, Vessel, 1 \gg,\\
&WeatherDanger| = \ll isFog, Weather, Vessel, 1 \gg,\\
&WeatherDanger| = \ll isSafety, Weather, Vessel, 0 \gg,\\
&WeatherDanger| = \ll threat, Weather, Vessel, 1 \gg
\end{aligned}
$$

Figure 6 shows two infons for the WeatherDanger situation.

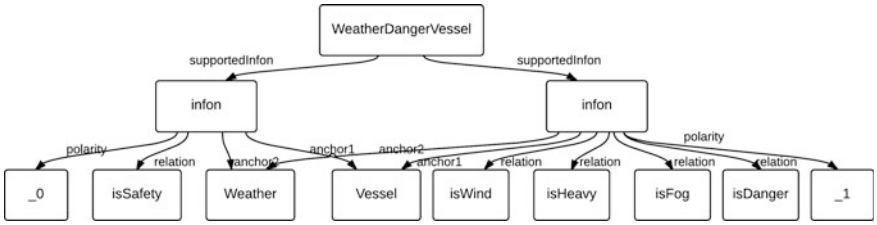


Fig. 6 Two infons for WeatherDanger situation

Finally, closing of an area for navigation due to an oil spill along the planned route is possible (OilSpill). In this case, it is necessary to make changes in the planned route with consideration of by-passing the closed area as well as meteorological and ice conditions. The additional infons for this situation are the following:

$$OilSpill| = \ll near, Vessel, OilSpillArea, 1 \gg,$$
$$OilSpill| = \ll changeRoute, Vessel, OilSpillArea, 1 \gg,$$
$$OilSpill| = \ll isSafety, Vessel, OilSpillArea, 0 \gg .$$

For each of above-mentioned infons, we can define rules that allow to describe the set of IGIS-based actions in the case of changes in the current situation. For this, it is necessary to define classes for each of the three parts of the rule. Also, we suppose that the following classes are already defined in the ontology: IceThreaten and WeatherThreaten.

6 Case Study: Intelligent GIS–Based Vessel-Route Planning

In this section, the results of the application of the proposed approach are given. IGIS has been used to define the areas at risk and vessel route corrections.

IGIS provides powerful tools for vessel-traffic monitoring, route planning, weather conditions, and ice-situation analysis in a given sea region. The IGIS Expert system can be used for control of navigation safety and decision-making support in case of dangerous situations on the Northern Sea Route.

Main scenario: A cargo vessel V travels along a pre-planned route on part of the Northern Sea Route (Kara Sea). The east part of the Kara Sea is ice-covered; an icebreaker is opening a safe passage through the ice field of Malygina Strait (Fig. 7). In Fig. 7, circles with numbers show the number of vessels at sea.

It is also known that some parts of the Kara Sea along the pre-planned route are dangerous due to weather conditions (storm alert, WMO sea state code: from 5 to 9) (Fig. 8).

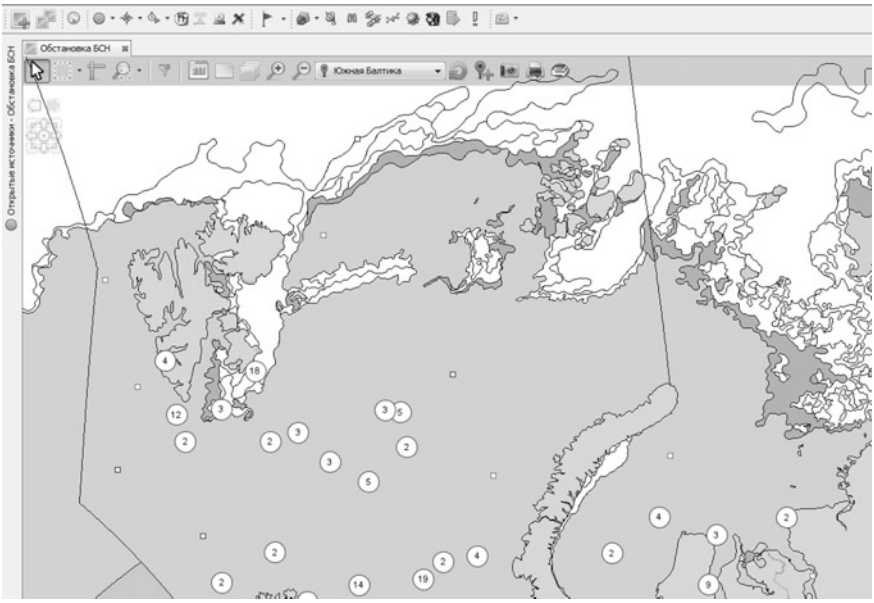


Fig. 7 The ice situation in the Barents and Kara Seas on September 15 (including prognostic information)

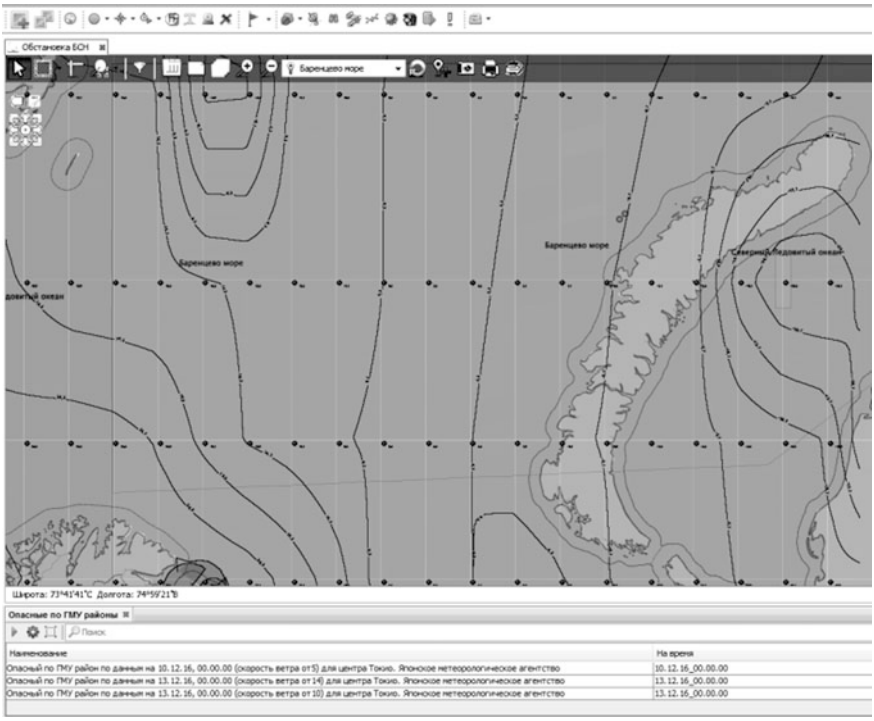


Fig. 8 Wind conditions along the Northern Sea Route

Ice and weather-conditions maps are used to construct object, which involves further calculations. Such objects may include weather-dangerous regions, ice-dangerous regions, closed regions, and others. Then dangerous regions are transformed into multi-point polygons. These polygons are instances of an ontology class [6].

Oil-spill scenario: At approximately 00:00 am, the tanker T collides with a cargo ship, the C, and spills oil from a damaged tank. The tanker T radios to report that the vessel has leaked approximately 4000 tons of crude oil. The location of the collision and oil spill is known and marked on digital map. At the same time, the vessel V is already traveling along the planned route on the part of Northern Sea Route where the dangerous ice-covered areas are present.

The IGIS has been used for quantifying the oil-spill size and trajectory [7] and for estimating the oil-spill area. The model takes into consideration the following factors:

- the coastline,
- the ice situation in the area of interest,
- the water flows, and
- the weather conditions.

In addition, the IGIS knowledge base can store information from previous incidents and present it through additional expert system rules, which can be used in future.

The current dangerous situations along the pre-planned route of the cargo vessel V are shown in the panel "Situation awareness" and on the digital map of IGIS MSRS (IGIS Maritime Surveillance and Recognition Systems) (Fig. 9).

The rules for inferring this situation and supporting route-change decision-making are as follows:

If
belongsTo(X, Vessel)
and belongTo(Y, WeatherDangerArea)
and near (Y, X)
Then
threat(Y, X)
If
belongsTo(X, Vessel)
and belongTo(Y, IceDangerArea)
and near (Y, X)
Then
threat(Y, X).

Then we define the situation type for any given situation. At any time, the vessel that by-passes an area closed to navigation can meet ice, and the ice will be a threat to the vessel if it is near the ice and moving in the direction of the ice. This condition can be represented in the form of mathematical expression as sub-sumption rule as follows:

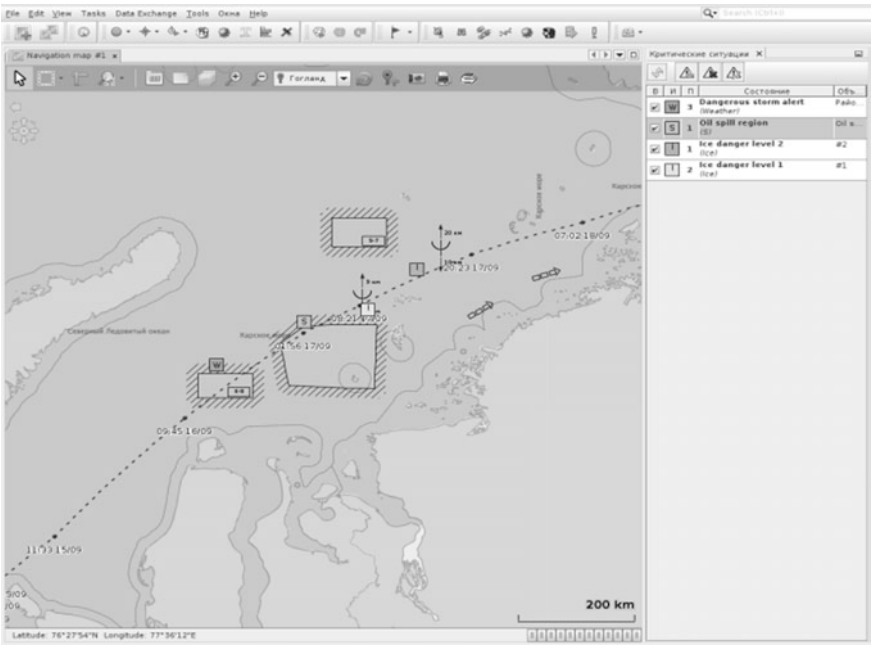


Fig. 9 Initial vessel route built with IGIS MSRS

$$S_1 \cap S_2 \Rightarrow S_3,$$

where,

$$S_1 = \{s|s| = \ll near, Ice, Vessel, 1 \gg\}$$
$$S_2 = \{s|s| = \ll inDirectionOf, Ice, Vessel, 1 \gg\},$$
$$S_3 = \{s|s| = \ll clash, Ice, Vessel, 1 \gg\}.$$

The subsumption rule is the basis for description logic [3], which is the underlying logic for an IGIS MSRS modeling subsystem.

The above-mentioned rules are stored in the IGIS MSRS knowledge base and used for situation awareness about a pre-planned vessel route. The results of the rules inference are shown on the digital map of IGIS with special marks along the vessel route. Throughout the descriptions of detected dangerous situations are presented in the panel "Situation awareness."

An alternative vessel route was for vessels instead of the regularly used one to avoid storm-dangerous waters and ice-covered regions of the Kara Sea (Fig. 10). The alternative vessel route was automatically suggested and shown on the digital map of IGIS MSRS. Thus, IGIS supports intelligent decision making for maritime safety monitoring in the Arctic seas.

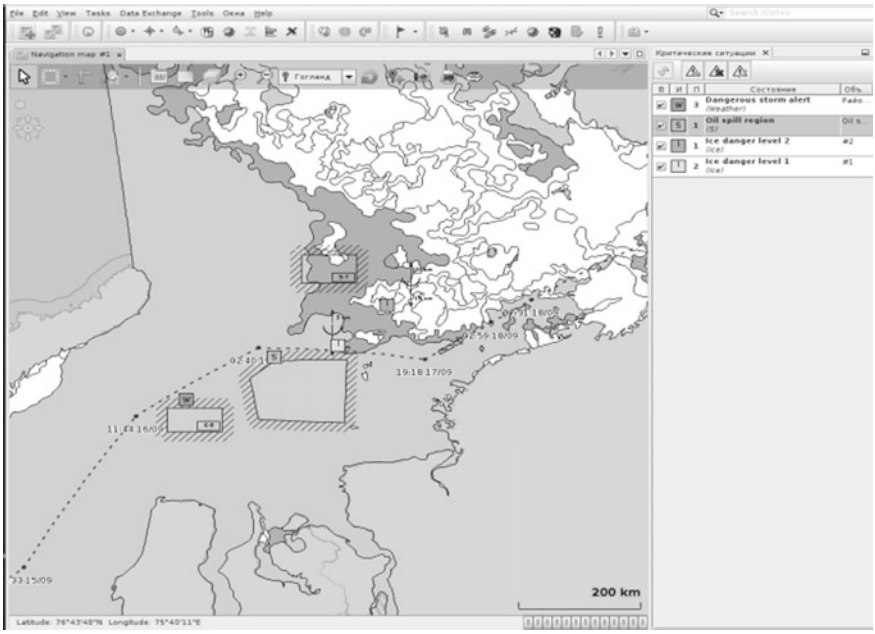


Fig. 10 Alternative vessel route built with IGIS MSRS

7 Conclusions

This chapter has presented an approach to vessel-route planning in the Arctic seas along the Northern Sea Route. The approach has also demonstrated a fusion of different science and technology knowledge: GIS, AI, and mathematical methods.

A constant need in monitoring different dangerous situations in the Arctic seas is related to the constantly changing navigation conditions typical for the region. There is a need to perform real-time monitoring of current situations (weather conditions, ice conditions) and their operational forecasting to provide safe navigation along parts of the Northern Sea Route.

Furthermore, we considered the development and implementation of the IGIS concept and technology for maritime-safety monitoring in the Arctic seas with the use of a situation-awareness model. The model provides situational planning of a vessel route based on a system comprising ontology and expert knowledge. The approach taken so far has been implemented as described in IGIS MSRS, and the initial evaluation has been found to be positive.

Future work by the authors is to extending the approach by adding to the vessel-route planning process new dangerous maritime situations, particularly addressing aspects of navigation in the Arctic seas.

References

1. Humpert M, Raspotnik A (2012) The future of shipping along the transpolar sea route. The Arctic Yearbook 1(1):281–307
2. Smith CL, Scott RS (2013) New Trans-Arctic shipping routes navigable by midcentury. In: Proceedings of the National Academy of Sciences of the United States of America 110.13, E1191–E1195, PMC Web, 26 July 2015
3. Goralski RI, Gold CM (2007) The development of a dynamic GIS for maritime navigation safety. In: Proceedings ISPRS'07 Workshop, Urumchi, China, pp 47–50, 28–29 August 2007
4. Baylon AM, Santos EMR (2013) Introducing GIS to TransNav and its extensive maritime application: an innovative tool for intelligent decision making? Int J Marine Navig Saf Sea Transp 7(4):557–566
5. Popovich V (2013) Intelligent GIS conceptualization, information fusion and geographic information systems. In Proceedings IF&GIS 2013 Conference, St. Petersburg, Russia, pp 17–44, May 2013
6. Smirnova O, Tsvetkov M, Sorokin R (2014) Intelligent GIS for monitoring and prediction of potentially dangerous situations. In: Proceedings SGEM 2014 Conference, vol 1, pp 659–666, 17–26 June 2014
7. Kokar M, Matheus C, Baclawski K (2009) Ontology-Based Situation Awareness. Inf Fusion 10(1):83–98
8. Sorokin R (2011) Advantages of intelligent geographic information system research prototype, information fusion and geographic information systems: towards the digital ocean. Springer, Heidelberg
9. Sorokin RP, Shaida SS, Smirnov AV (2003) Intelligent geo-information systems for modeling and simulation. In: Proceedings Harbor, Maritime and Multimodal Logistics Modeling & Simulation Conference, Riga, Latvia, 2003, pp 395–398
10. Endsley MR SAGAT (1987) A methodology for the measurement of situation awareness. Northrop Corp, Hawthorne, CA

Ranking of Information Sources Based on a Randomised Aggregated Indices Method for Delay-Tolerant Networks

Oksana Smirnova and Tatiana Popovich

Abstract INFORMATION for various monitoring systems usually comes from heterogeneous sources such as AIS, radars, Internet, etc. However, not all sources can be considered equally reliable; furthermore, the quality of transmitted data also differs. It has been deemed necessary to solve the problem of ranking sources according to their credibility based on expert information and a randomised aggregated indices method (RIAM).

Keywords Data sources · Expert information · Randomised aggregated indices · Delay-tolerant network

1 Introduction

Geographical information systems (GIS) have rapidly become a popular widespread technology for the monitoring and observation of geo-special processes of various nature.

Monitoring systems, as a rule, obtain information from different heterogeneous sources. Among such sources we can name: AIS, radar, Internet sources, and others. However, not all information sources can be considered equally reliable. Therefore, it is necessary to solve the problem of ranking such information sources according to their reliability based on pre-collected expert estimations. We shall use a randomised aggregated indices method (RIAM) for solving this problem.

To determine the relevance of each source, they are assigned with an aggregated weight coefficient that denotes source's reliability and trustworthiness.

O. Smirnova (✉) · T. Popovich
SPIIRAS Hi Tech Research and Development Office Ltd., 199178 St. Petersburg, Russia
e-mail: sov@oogis.ru

T. Popovich
Bonch-Bruevich St. Petersburg State University of Telecommunications,
St. Petersburg, Russia
e-mail: t.popovich@oogis.ru

V. Popovich et al. (eds.), *Information Fusion and Intelligent Geographic Information Systems (IF&IGIS'17)*, Lecture Notes in Geoinformation and Cartography, DOI 10.1007/978-3-319-59539-9_7

Such aggregated weight coefficients are calculated according to a RIAM. This method uses expert knowledge about the relevance of each source of information. In this chapter, the problem of ranking sources according to their credibility based on expert information and a RIAM is applied for a delay-tolerant network (DTN) environment.

2 Related Works

Most modern researches [1–4] have been dedicated to determining the quality of information in general that, by association, transfers to information sources. In [4], it was proposed to classify information quality criteria into sets: content-related, technical, intellectual, and instantiation-related. To determine the quality of data, researchers usually employ statistical data along with expert estimations [4]. There are several works on information sources ranking [4, 5]. They maximise criterion derived from a linear convolution of a list of weighted criteria such as availability, price, consistency, etc., with a given threshold score for each source. However, this method seems rather cumbersome and a bit overdone, especially for distributed-monitoring systems where the data supply is limited. In this chapter, we propose an application of a more elegant RIAM that can deal with expert information in its rawest form and is capable to operate with purely numerical criteria as well.

3 Application of DTN Networks

For monitoring systems, one of major problems is the continuous transmission of data about an observed situation. One of the innovative ways of solving this problem is the implementation of DTN technology.

The currently used TCP/IP protocol, which is applied throughout all of the modern Internet, assumes that a data package has a remotely small size, can be transmitted quickly, and does not need to be stored for lengthy periods of time. If the next node of the route is unavailable, the according response is sent, and the package is deleted [6].

Existing Internet protocols are not well adapted for currently developing environments, such as distributed GIS, due to some fundamental assumptions built into the Internet architecture [7]:

- that an end-to-end path between source and destination exists for the duration of a communication session;
- (for reliable communication) that retransmissions based on timely and stable feedback from data receivers is an effective mean for repairing errors;
- that the end-to-end loss is relatively small;

- that all routers and end stations support the TCP/IP protocols;
- that applications need not worry about communication performance;
- that endpoint–based security mechanisms are sufficient for meeting most security concerns;
- that packet-switching is the most appropriate abstraction for interoperability and performance;
- that selecting a single route between sender and receiver is sufficient for achieving acceptable communication performance.

DTN, in contrast, is a network composed of a set of networks. The DTN architecture embraces the concept of occasionally-connected networks that suffer from frequent transmission breaks and can consist of more than one mixed set of protocols or protocol families. DTN supports compatibility with other networks and adjustment to prolonged breaks and delays in connectivity and protocol recoding. This functions allow DTN to match mobility and limited power of developing wireless communication tools.

DTN overcomes the problems associated with intermittent connectivity, long or variable delay, and high error rates by using store-and-forward message switching. Entire messages—or pieces (fragments) of such messages—are moved (forwarded) from a storage place on one node (switch intersection) to a storage place on another node along a path that eventually reaches the destination [1].

The DTN architecture is conceived to relax most standard TCP/IP assumptions based on a number of design principles [7]:

- use variable-length (possibly long) messages (not streams or limited-sized packets) as the communication abstraction to help enhance the ability of the network to make good scheduling/path selection decisions when possible;
- use a naming syntax that supports a wide range of naming and addressing conventions to enhance interoperability;
- use storage within the network to support store-and-forward operation over multiple paths and over potentially long time-scales (i.e., to support operation in environments where many and/or no end-to-end paths may ever exist); does not require end-to-end reliability;
- provide security mechanisms that protect the infrastructure from unauthorized use by discarding traffic as quickly as possible;
- provide coarse-grained classes of service, delivery options, and a way to express the useful lifetime of data to allow the network to better deliver data to serve the needs of applications.

Initially, DTN were developed for operation under interstellar conditions [8, 9] where even the speed of light may seem bit slow and delay tolerance is essential. However, DTN can have a wide range of applications on Earth. DTN can use different types of wireless communication such as radio frequencies (RF), ultra-wide bandwidth (UWB), free-space optics (FSO), and acoustic technologies (sonar or supersonic communications) [8]. Fields of DTN application include the following:

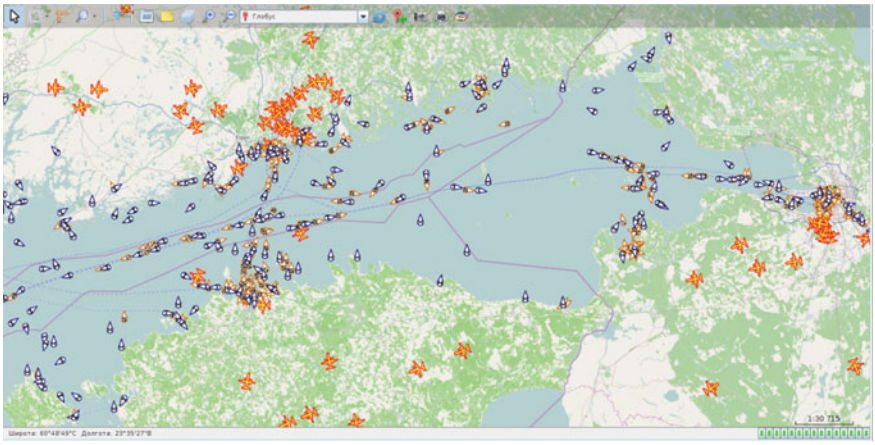


Fig. 1 Representation of object data from various sources in GIS

- space agencies: maintaining communications with space stations, interstellar transmissions, and space-debris monitoring;
- military and intelligence: mobile specialised networks for wireless communications, monitoring, surveillance, search-and-rescue communication, and drone-handling;
- commercial: cargo and transport monitoring (road, railway, maritime, or air transport), tracking belongings, transaction data, etc.;
- social services and security: emergency communication, support of emergency response services in "smart cities," transport networks support, etc.;
- monitoring of the environment;
- personal uses.

Application of such technology in GIS can aid in the achievement of representation of only contemporary information without time lapses and with insignificant data corruption. Constant data flow is also crucial for distributed monitoring systems that are territorially distributed and that aggregate and process data from a large number of heterogeneous sources (Fig. 1). Such systems are also characterized by the necessity of making operational decisions in a limited span of time [10]. Therefore, at the point of sending data through limited power networks in cases of transmission breaks and delays in communication, we should determine which bundles of information possess most value and thus should be transmitted in primary order.

4 Randomised Aggregated Indices Method

For our research, we shall use expert information as a basis for ranking information sources according their relevance. Experts rarely can provide precise numeric estimations and usually give information about alternatives' probabilities in the form of

comparative statements like "alternative A_i is more probable than alternative A_j" or "alternative A_i has the same probability as alternative A_j" [8]. While we may have prior knowledge about weights of chosen experts: "the estimations, given by k-th expert, are more weighty then analogous estimations, given by lth expert," but even this joined information may be incomplete, i.e., an amount of this information is not enough for the unambiguous (unique) determination of probabilities and weights under consideration. Therefore, it will be realistic to suppose that we posess only non-numeric (ordinal), non-exact (interval), and non-complete (incomplete) expert knowledge (NNN-knowledge, NNN-information) about alternatives' probabilities and weights of different information sources [11]. Hence, it has been proposed [12] to model uncertain choice of admissible (from the point of view of appropriate NNN-information) probabilities and weights by a random choice from corresponding sets of probabilities and weights. By such Bayesian randomization of uncertainty, random alternatives' probabilities and random experts' weights are obtained, and mathematical expectations of these numerical random variables are interpreted as the needed numerical image of corresponding NNN-information [12] (Fig. 2).

Consider at a present time-point t_1 a complex system that can proceed to one of the finite number of alternatives $A_1, \ldots, A_r$ at a future time-point t_2. Suppose that a decision maker has m different sources of information (experts) about probabilities $p_i = P(A_i), \quad i = 1, \ldots, r, p_i \geq 0, \; p_1 + \cdots + p_r = 1$ of the system transition into alternative states $A_1, \ldots, A_r$ at the time-point t_2. The decision maker obtains NNN-knowledge I_j from "experts," which can be presented as a vector $I = (I_1, \ldots, I_m)$ that consists of systems of equalities and inequalities for probabilities $p_1, \ldots, p_r$. The decision maker has NNN-knowledge J about the comparative "weights" of the sources of information. Thus, the entire set of NNN-knowledge constitutes vector $(I; J) = ((I_1, \ldots, I_m); J)$ where I_j, is information obtained from j-th source, and system J is NNN-knowledge that expresses the preferences of the decision maker about the comparative significance of sources of the information [12]. If we take into account information I_j, we can form a set $P(r; I_j)$ of all admissible

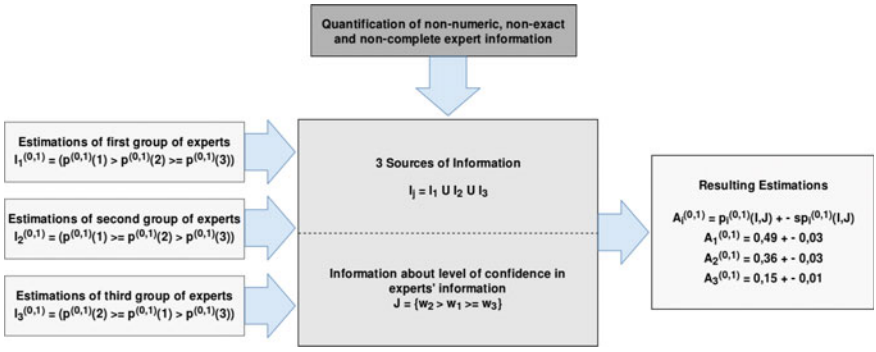


Fig. 2 Quantification of expert information algorithm

(according to information I_j) probability vectors $p = (p_1, \ldots, p_r)$. Modelling uncertain choice of vector $p = (p_1, \ldots, p_r)$ from the set $P(r; I_j)$ by a random choice, we obtain a random probability vector $\tilde{p}_1(I_j) + \cdots + \tilde{p}_r(I_j) = 1$ $\tilde{p}(I_j) = (\tilde{p}_1(I_j), \ldots, \tilde{p}_r(I_j)), \tilde{p}_i(I_j) \geq 0$, which has a uniform distribution on set $P(r; I_j)$ [8]. A component $\tilde{p}_i(I_j)$ of random vector $\tilde{p}_i(I_j)$ represents a random estimation of the probability of an alternative A_i according to information I_j obtained from j-th source. Vector of mathematical expectations $\bar{p}(I_j) = (\bar{p}_1(I_j), \ldots, \bar{p}_r(I_j))$ we will interpret as numerical image NNN-information I_j . In other words, vector $\tilde{p}_i(I_j)$ is a result of information I_j quantification. In these terms, a random vector $\tilde{p}_i(I_j)$ is a stochastic quantification of NNN-information I_j [12]. High qualification of an experts and careful interpretation of his or her estimations allow to significantly improve estimations of the regarded value. However, low qualification of experts in a chosen field of study or inaccurate interpretation of his or her knowledge can have a negative impact on overall estimations.

5 Case Study

To determine the relevance of every information source, each of them is to be denoted with a weighting coefficient, which is calculated according to RIAM. This method uses expert information about the credibility of each source of information. Expert information is quantified into corresponding weighting coefficients for the sources (Fig. 3).

Knowledge and information concerning the credibility and relevance of information sources, collected from experts, are rarely presented in numerical form and usually are given in form of comparative statements such as "source 1 is more reliable than source 3 and source 3 is more reliable than source 2" and so on (NNN-information). To transform such statements into numerical coefficients, we apply RIAM. The level of confidence in experts' estimations can also be calculated using the same method.

For each source of information from the given set, we denote an index: source 1 corresponds to index p_1; source 2 corresponds to index p_2; and, consequently,

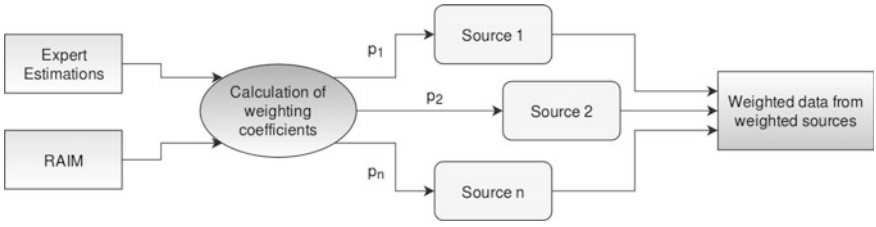


Fig. 3 Algorithm of sources' ranking

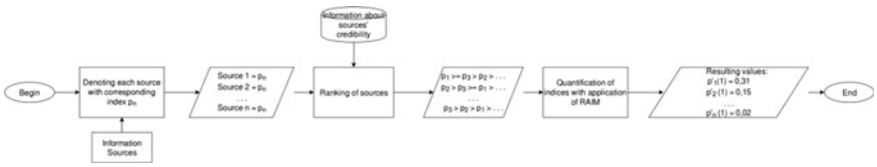


Fig. 4 Algorithm of quantification of weighting coefficients

Table 1 Weighting coefficients for information sources

	Expert 1	Expert 2	Expert 3	Expert 4	w_{1j}	w_{2j}	Resulting weighting coefficients
Relevance	0.32	0.4	0.08	0.2	0.67	0.33	
Source 1	0.5	0.53	0.33	0.5	0.5	0.5	0.5
Source 2	0.35	0.31	0.53	0.33	0.35	0.34	0.35
Source 3	0.15	0.16	0.14	0.17	0.15	0.16	0.15

source n corresponds to index p_n. Next, according to the NNN-information, which is delivered by selected experts, sources are prioritised according to their relevance. The comparative statements of one expert are represented in the form of inequality: $p_1 \geq p_3 > p_2 > \cdots > p_n$. The RIAM is applied to such inequalities, after which they take shape of numerical weighting coefficients (Fig. 4).

The level of experts' qualification also has an impact on the resulting weighting coefficients. Each expert, whose statements were used in the quantification of coefficients, is denoted with an index as well: expert 1 corresponds to w_1; expert 2 corresponds to w_2; and expert m corresponds to w_m. Experts are ranked according to their qualifications and trustworthiness of their statements: $w_3 > w_2 > \cdots > w_m$. Such inequality is quantified according to the same principle as the sources' coefficients.

As an example, let us choose three sources of information about vessels in GIS: source 1 (radar A_1), source 2 (AIS A_2), source 3 (Internet resources, e.g., ShipFinder A_3). Let us rank them according to the relevance of the transmitted information. We will use the knowledge of four qualified experts as the basis for ranking our chosen sources. Experts' estimations are given in the form of inequalities: the estimation of the first expert : $A_1 \geq A_2 > A_3$; the estimation of the second expert : $A_1 > A_2 \geq A_3$; the estimation of the third expert : $A_2 > A_1 > A_3$; and the estimation of the fourth expert : $A_1 \geq A_2 \geq A_3$. The level of credibility of each experts' estimation is also denoted by inequality: $w_2 \geq w_1 > w_4 > w_3$. In spoken words, it means that "the second expert is more or equally trustworthy as the first when the first is more trustworthy than the fourth," etc. The resulting weighting coefficients for information sources are listed in Table 1.

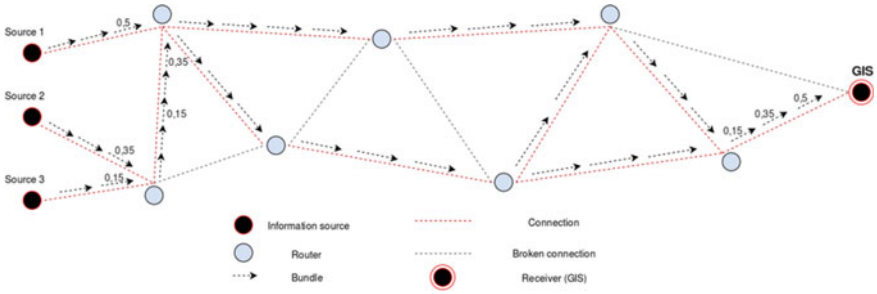


Fig. 5 DTN with weighted bundles

In the end, each source is denoted by an aggregated numerical weighting coefficient that represents the relevance and adequacy of the information transmitted through it. Each bundle of information, transmitted through DTN, is also denoted by the according coefficient that defines its priority in information flow (Fig. 5).

6 Conclusion

DTN can be considered a very perspective technology for application with GIS and especially with distributed monitoring systems. DTN allows to maintain connectivity between the source of information and the receiver in conditions of limited power networks and in cases of transmission breaks and delays in communication. However, the problem of ranking information according to its relevance remained open. In this chapter, the authors described the possibility of successful ranking of information sources with application of a RIAM in DTNs.

References

1. Angeles M, MacKinnon L (2010) Assessing data quality of integrated data by quality aggregation of its ancestors. Comput Sist 13:331–344
2. Canini K et al (2011) Finding credible information sources in social networks based on content and social structure. In: Proceedings of the IEEE international conference on privacy, security, risk, and trust
3. Chen Y et al (1998) Query processing with quality control in the World Wide Web. World Wide Web 1:241–255
4. Naumann F (2002) Quality-driven query answering for integrated information systems. Lecture notes in computer science, vol 2261, Springer, Berlin. p 168
5. Naumann F et al (1999) Quality-driven integration of heterogeneous information systems. In: Proceedings of the 25th VLDB conference, Scotland
6. Hovanov NV (1986) Stochastic models in qualimetric scales theory. Leningrad State University Press, Leningrad, USSR (in Russian)

7. Warthman F, Delay- and disruption-tolerant networks (DTNs): a tutorial/electronic resource: http://ipnsig.org/wp-content/uploads/2012/07/DTN_Tutorial_v2.05.pdf
8. Cerf V et al (2007) Delay-tolerant networking architecture. In: The IETF Trust. Electronic resource: https://tools.ietf.org/html/rfc4838
9. Jenkins A et al (2010) Delay/disruption-tolerant networking: flight test results from the international space station. In: Proceedings of the IEEE aerospace conference
10. Popovich V (2012) Hybrid networks for heterogeneous distributed maritime observation systems. In: Proceedings of the IEEE 1st AESS European conference on satellite telecommunications, pp 30–36
11. Hovanov NV, Analysis and synthesis of indices in information shortage conditions. St. Petersburg State University Press, Russia (in Russian)
12. Hovanov N, (2009) Multicriteria estimation of probabilities on basis of expert non-numeric, non-exact and non-complete knowledge. Yudaeva M, Hovanov N, (eds) Eur J Oper Res 195:857–863
13. Hovanov NV (2010) Prognosis of characteristics of financial and economics factors based on expert and statistical information synthesis. In: Proceedings of 10-th International Science School "Modelling and analysis of safety and risk in complex systems", pp. 114–122 (in Russian)
14. Hovanov N (2012) Stochastic quantification of non-numeric measurement scales: theory and application. In: Hovanov N (ed) International seminar "mathematics, statistics, and computation to support measurement quality", pp 28–31
15. Vladyko V et al (2013) Application of aggregated indices randomization method for prognosing the consumer demand on features of mobile navigation applications. In: Information technologies and telecommunications, no 4, The Bonch-Bruevich St. Petersburg State University of Telecommunications (in Russian)
16. Zhang W (2004) Handover decision using fuzzy MADM in heterogeneous networks. In: Proceedings of the IEEE Communications Society, WCNC, pp 653–658

Route Planning for Vessels Based on the Dynamic Complexity Map

Zhe Du, Liang Huang, Yuanqiao Wen, Changshi Xiao and Chunhui Zhou

Abstract Regarding aiming at multiple mobile objects in a complex water traffic–navigation environment, the chapter puts forward a route-planning method based on the complexity map. First, a Complexity Map was established according to the theory of complexity measurement. Then, combined with the A* algorithm, the actual cost was modified by making use of the distribution of the complexity values. Meanwhile, to reduce the distance of the whole voyage and avoid the local minimum in the process of path-finding, the Euclidean distance from the current point to the target was set to estimate heuristic cost. In addition, by using the standardization method, the value of complexity and distance became dimensionless. Finally, considering the ship dimensions, a channel-boundary constraint was added. The experimental results show that in the case of satisfying the ship dimensions, the best path is close to the shortest route and avoids all the high complex areas.

Keywords Complexity map · A* algorithm · Multiple mobile objects · Route-planning

1 Introduction

The rapid growth of the economy in China makes the shipping industry develop rapidly. In addition, the increase of cargo types and freight volume makes the water-traffic situation more complicated: The trend is that the type of ship has become more diverse; the size of the vessel has become larger; and the navigable waters have become more intense [1], thus increasing the dangers of water traffic.

Z. Du (✉) · L. Huang · Y. Wen · C. Xiao · C. Zhou
School of Navigation, Wuhan University of Technology, Wuhan, China
e-mail: 1149212283@qq.com

L. Huang · Y. Wen · C. Xiao · C. Zhou
Hubei Key Laboratory of Inland Shipping Technology, Wuhan, China

L. Huang · Y. Wen · C. Xiao · C. Zhou
National Engineering Research Center for Water Transport Safety, Wuhan, China

V. Popovich et al. (eds.), *Information Fusion and Intelligent Geographic Information Systems (IF&IGIS'17)*, Lecture Notes in Geoinformation and Cartography, DOI 10.1007/978-3-319-59539-9_8

That's why the time and cost saving concerned by the planning route gradually transformed into security. Therefore, how to quickly and efficiently plan a safe and reliable route in a complex navigation environment has become one of the focuses of current research.

The core of the problem is path planning. The related research has accomplished much in terms of gaining information, and this information has been widely used in various fields such as a robot's autonomous motion with a collision-free, unmanned aerial vehicle (UAV) and obstacle-avoidance penetration flight [2], GPS navigation, urban-road planning, vehicle and management resource–allocation challenges [3], etc. However, there has been little research on water transportation. Only a few studies have proposed different planning algorithms according to the types of aircraft and the planning objectives.

For large ocean vessels, economics are often concerned. That makes the minimum distance and time important goals. With the goal of minimum total time, Ji [4] established a mixed-integer programming regional container-transport model and used a genetic algorithm to solve the model. Li [5] focused on the shortest total distance and proposed a route-optimization scheme based on the Tabu Search (TS) algorithm. Zhang [6] considered the shortest time and minimum cost to complete re-supply missions as the objective function and put forward a path-planning scheme for a large ship material–handling system. Small vessels, such as the unmanned surface vessel (USV), usually have a task and are greatly influenced by the environment; thus, the final goal is often to complete the tasks independently. Chen [7] takes advantage of the Ant-Colony algorithm to optimize the Dijkstra algorithm that lets the unmanned rescue boat get a better collision-free path and then complete the search and rescue mission. In order to solve the problem of encountering underwater obstacles when USVs are gathering information, Phanthong [8] reconstructed the original path using multi-beam front sonar combining with A* algorithm to realize the real-time path planning. Xie [9] aimed at solving USV cruise problems under conditions of a complex environment and proposed a method that combines global (Fast Dijkstra algorithm) and local (Artificial Potential Field method) planning to implement safe navigation in unknown and dynamic environments.

From the above, regarding water traffic path–planning research, the safety of the path has been simplified whether the efficiency of large ships or the mission completeness of small vessels is considered. However, with the navigation environment becoming more complex, it is more difficult to ensure safe path-planning. Even if the above-mentioned documents can solve part of the problems, the situation is safe only considering a single or a few dynamic objects. If the environmental assumptions are too simple, the value of study for path-planning in a real situation will lack reference. Aiming at solving the problem of encountering multiple mobile objects in the water-traffic navigation of complex environments, this chapter makes use of the theory of complexity measure to build a complexity map; the path has been searched to meet the requirements of sailing safety, route distance, and shipping-dimension constraints based on the A* algorithm. The experimental results show that the method can plan the shortest safe path that satisfies all of the constraint conditions.

2 Complexity-Map Modeling

2.1 Concept of Complexity

The concept of complexity is often referred to in the aviation field. This concept has put forward the idea of an "air traffic–management center" (ATM) [10]. This would involve choosing a suitable method to measure the work load and the complexity of traffic behavior for the Air Traffic Services Center. After much research, some foreign scholars believe that there is a close relationship between the complexity of traffic conflicts and traffic flow. Hong et al. [11] took an airspace as research object to study the effect of traffic situations when aircraft invade the airspace. By analyzing the relationship among traffic complexity, aircraft azimuth, and course angle, a conceptual model of the complexity map was proposed. As for water-traffic complexity, however, there is less research on home and abroad. So far, only Huang [12] and Wen [13] have established a model for water traffic flow–complexity measurement, which is based on the research of the water traffic–flow characteristics. Therefore, we took advantage of this model to conduct the study reported in this chapter.

In research waters, every two ships constitute a basic traffic unit. First, according to the traffic status and traffic-trend factors, we measure the complexity between every two ships; then, we compute the complexity among more ships through a nonlinear superposition; last, we divide the waters into several identical grid sizes and calculate each grid-complexity value to obtain the distribution of the complexity in the waters: That is the Complexity Map, and the greater the value, the higher the risk.

2.2 Complexity Calculation Model

Traffic-Status Factor

Traffic status can be measured by the relative distance between two ships, which is the most obvious information available. When there is a trend toward convergence, closer relative distance can often create a greater threat. Even if the closest approach point has been passed, we cannot ignore the influence of distance between the two ships. Therefore, according to relative distance, we can build traffic-status model C_S [12] as follows:

$$C_s = \lambda e^{-k\frac{D_{ij}}{R_s}} \tag{1}$$

where k and λ are weight coefficients; R_S is safety distance; and D_{ij} is relative distance.

Traffic-Trend Factor

The measurement of traffic trend can be divided into the following processes: convergence judged, consequences estimated. and urgency estimated.

1. Convergence judged

The convergence situation between two ships can be seen as a change of the relative distance between them as follows:

$$\frac{d\overrightarrow{D}_{ij}}{dt} = v_{ij}\cos\left(\overrightarrow{D}_{ij}, \vec{v}_{ij}\right) \tag{2}$$

where $\left(\overrightarrow{D}_{ij}, \vec{v}_{ij}\right)$ is the angle between the relative distance and relative velocity,- denoted as θ and $\theta \in [0, \pi]$. According to the relative position, we can determine that when $0 < \theta < \pi/2$, $d\overrightarrow{D}_{ij}/dt > 0$, there is a converging tend between two ships; when $\pi/2 < \theta < \pi$, $d\overrightarrow{D}_{ij}/dt < 0$, there is a separating tend between two ships. Therefore, the formula $d\overrightarrow{D}_{ij}/dt$ can be used to judge if there is a converging tend between two ships as follows:

$$C_{T1} = \begin{cases} 1, & \frac{d\overrightarrow{D}_{ij}}{dt} \leq 0 \\ 0, & \frac{d\overrightarrow{D}_{ij}}{dt} > 0 \end{cases} \tag{3}$$

Namely, when $C_{T1} = 1$, the two ships are converging; when $C_{T1} = 0$, the two ships are separating.

2. Consequences estimate

In the traffic unit, if two ships do not change their course and continue to sail until their distance is closest, the increment of complexity is called "convergence consequences." At the time, the relative distance between two ships is the distance to the closest point of approach, namely, DCPA. DCPA can be used to measure the degree of danger between two ships: The smaller the DCPA, the higher the conflict complexity. Because DCPA describes future position relations between the two ships, the potential consequences in traffic unit can be denoted as C_{T2} [12] as follows:

$$C_{T2} = \lambda \cdot \left(e^{-\alpha\frac{D_{\mathrm{CPA}}}{R_S}} - e^{-\alpha\frac{D_{ij}}{R_S}}\right) \tag{4}$$

where α are the weight coefficients; and the other parameters are the same as in Eq. (1).

3. Urgency estimate

According to Eq. (2), C_{T2} is the increment of complexity, but it cannot measure the potential degree of urgency. The degree of urgency is the time from D_{ij} to DCPA, usually denoted as T_{CPA} (time to closest point of approach). By using T_{CPA}, the time-collision risk C_{T3} can measure the degree of urgency [12] as follows:

$$C_{T3} = \begin{cases} 1 & T_{\mathrm{CPA}} \le t_a \\ \left(\frac{t_b - T_{\mathrm{CPA}}}{t_b - t_a}\right)^a & t_a < T_{\mathrm{CPA}} \le t_b \\ 0 & T_{\mathrm{CPA}} \ge t_b \end{cases} \tag{5}$$

where t_a and t_b are time nodes relating to the security area radius of ships; and a is a coefficients >1.

Therefore, the traffic-trend quantity can be measured by the above three factors: C_{T1} is used to judge whether there is a traffic conflict in the traffic unit; C_{T2} is used to evaluate the complexity increment when the two ships are close but without taking necessary action; and C_{T3} is used to evaluate the degree of urgency between two ships. The traffic-trend model can be expressed as [12] follows:

$$C_T = C_{T1} \cdot C_{T2} \cdot C_{T3} \tag{6}$$

Complexity Model

In a traffic unit, the traffic complexity comprises the traffic-status factor and the traffic-trend factor; it can be defined by adding the following format [12]:

$$C = C_S + C_T = \lambda e^{-k\frac{D_{ij}}{R_S}} + C_{T1} \cdot C_{T2} \cdot C_{T3} \tag{7}$$

When the two ships are separating, $C_{T1} = 0$, and there is no conflict trend; when they are converging, $C_{T1} = 1$, and there might be conflict trend. The consequences are determined by DCPA, and the urgency is determined by TCPA.

2.3 Building the Complexity Map

In order to reflect the complexity distribution of the waters, first divide them into several identical grid sizes, and calculate each grid complexity value to obtain the distribution of the complexity in the waters. Assume ship i and j in the waters: $T_{i1}(k)$ and $T_{j1}(k)$ are the moment when i and j enter the grid k, respectively, and

$T_{i2}(k)$ and $T_{j2}(k)$ are the moment when i and j leave the grid k, respectively. Then, the complexity of i and j in grid k is as follows:

$$C_{i,j}(\mathrm{k}) = \sum_{t=t_1}^{t_2} C(\mathrm{t}) \tag{8}$$

where $C(t)$ is the complexity between i and j, $t_1 = \max\left(T_{i1}(k), T_{j1}(k)\right)$, $t_2 = \max\left(T_{i2}(k), T_{j2}(k)\right)$.

Meanwhile, taking into account all the other ships related to the grid whose complexity and the influence of the overall traffic scales, define the complexity value of the grid k as:

$$C(k) = \sqrt{\left(\sum_{i=1}^{n-1}\left(\sum_{j=i+1}^{n} C_{i,j}(k)\right)^2\right) \cdot n} \tag{9}$$

3 Route-Planning Based on the Complexity Map

3.1 The Main Idea

The main idea of searching for the shortest route is based on the A* algorithm using the complexity map to improve the cost function. The A* algorithm is a kind of typical heuristic algorithm. Its main idea is to choose the appropriate cost function according to the final purpose. It calculates the cost value with the current node reaching all the notes it can; then it determines the minimum from the cost value of the collection; and it makes the minimum value as the next of the current node. According to the process search, which covers the whole area of waters constantly, the algorithm continues until the final point is reached. The general expression of the algorithm cost function is as follows:

$$f(x) = g(x) + h(x) \tag{10}$$

Among them, $g(x)$ is the direct cost function from the starting point to the current point; and $h(x)$ is the heuristic cost function from the current point to the final point [14]. The shortest distance is the ultimate goal of the traditional A* algorithm, and so $g(x)$ is the Euclidean distance, and $h(x)$ is the Manhattan distance.

3.2 Route-Planning Methods

Cost Function

In the chapter, route planning occurs in the environment of dynamic multi-objects, so the ultimate purpose is to avoid all the dynamic objects in order to safely to arrive at the final point.

According to part 2, the complexity map is a grid map that has weight. That is, based on the situation of the dynamic objects and the calculation, each grid has only the corresponding numerical values. The value shows that the complexity (risk) of different areas on the map: The greater the value, the greater the risk; when the risk is lower, i.e., when there is no danger, the value is 0. Regarding riverbanks and static objects, they are not accessible for ships. So the grid is set to minus infinity (−INF) or non-numeric (NAN). To reach the target safely, compute the total value of all of the grids that the route has passed at minimum. Let $g(x)$ be the complexity-cost function and compare the complexity of the surrounding eight-grid (eight figure puzzles) from the current node. Select one of the smallest nodes as the next node, and the current node became the parent node with a pointer to it. Cycle search like this, so the final complexity value of the route must be a minimum.

However, as a matter of fact, the following two questions must be considered.

(1) Ship-size restraint

Because a ship has a certain size, once the generated route is close to the channel boundary, the risk of grounding could happen, which decreases the safety of navigation. Thus, in the process of the path-planning, we cannot regard the ship as a particle directly; the size should also be considered. To simplify the problem, here we approximate ship as a rectangle.

When the ship is sailing, the course is approximately outspread along the channel direction. Therefore, we only care about the relationship between the beam and the channel boundary. Assume the beam of ship as B and a single size of grid as a; then the number of grids the ship has occupied is determined as follows:

$$n = \lceil B/a \rceil \tag{11}$$

Symbol "⌈⌉" is the ceiling. After n is determined, such constraint conditions have been added in the process of the search: Calculate the minimum number of grids m, which extend from the current node to the channel boundary nodes. If $m > n$, continue to search; if $m \leq n$, change the direction of the next node to make sure that n is no longer reduced.

(2) Distance restraint

Usually when a ship is sailing in a waterway, most of the waters are safe. Therefore, in the complexity map, the number of grids that have complexity value is relatively small, and most of the values are zero. It is not accurate if we only consider the factor

of complexity. Even in some cases, the route cannot be planned. Therefore, other factors should also be considered. Here, we consider the distance factor.

Let the Euclidean distance be the heuristic cost function $h(x)$. In the process of calculating the complexity, compute the distance from the current point to the target point. The benefits of doing this are as follows:

1. In safe waters, route-planning is mainly determined by the distance factor. This not only avoids the ship being trapped in the local minimum when all the surrounding grids' complexity values are zero, it also decreases the distance of the planned route.
2. The distance heuristic cost makes the search route trend toward the target. Thus, it eliminates an unnecessary search point and improves the algorithm's speed.

Standardize Complexity and Distance

Considering that the complexity and the distance value cannot operate each other, standardization must proceed. By the Min-Max standardization processing method [15], both the complexity and distance are transformed into a dimensionless value between 0 and 1. The conversion formula is as follows:

$$x_j = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{12}$$

Among them, x_i is the data after standardization; x is the original data; and $x_{\min}$ and $x_{\max}$ is the minimum and maximum values of x, respectively.

As for complexity, suppose the original complexity value as C; the maximum is $C_{\max}$; and the minimum is 0. Then, the standardized complexity C_i is as follows:

$$C_i = \frac{C}{C_{\max} - 0} = \frac{C}{C_{\max}} \tag{13}$$

Because the complexity value is generated along with the complexity map, the standardization process can be completed before the beginning of route searching. As for distance, because the result of the calculation is from the current point to the target point, standardization must be carried out during the process of route searching. Suppose the distance from the current to target is D; the maximum is $D_{\max}$; and the minimum is 0 (when the current is the target). Then, the standardized distance D_i is calculated as follows:

$$D_i = \frac{D}{D_{\max} - 0} = \frac{D}{D_{\max}} \tag{14}$$

After this step, the two values can be calculated directly. A flow chart (Fig. 1) shows the entire algorithm.

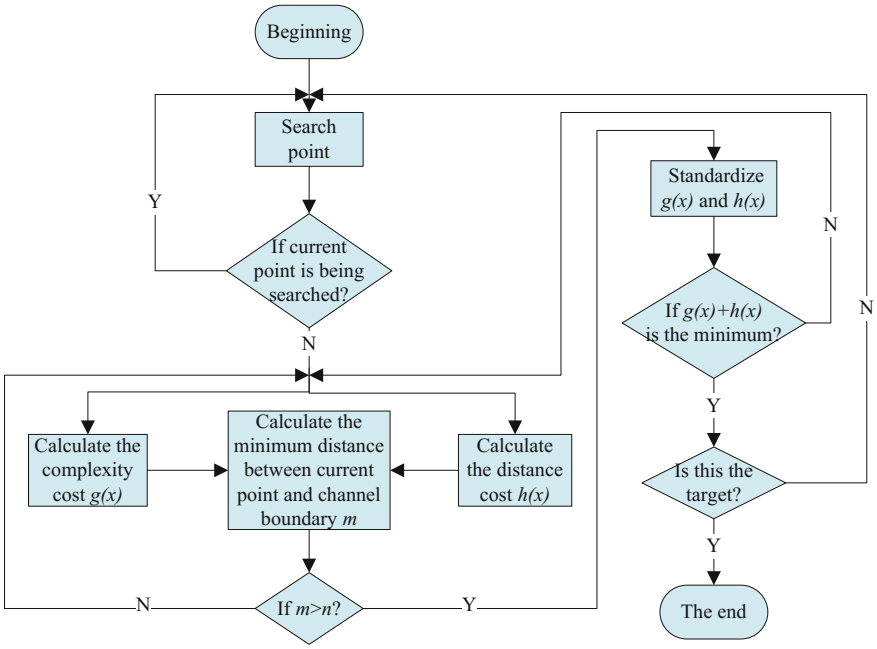


Fig. 1 Algorithm flow chart

4 Simulation and Results

The simulation environment is Matlab R2015a, and three sets of the experiment have been performed. The first two groups are under virtual conditions: Group 1 compared the algorithm considering only complexity and distance (traditional); group 2 compared the arithmetic with complexity–distance as well as complexity alone; and group 3 group simulated dynamic multi-objects under real channel conditions.

4.1 Comparison of Complexity Only Versus Distance Only

Set up a grid map of 30 * 30, and assign each grid random numbers from 0 to 1 (i.e., the complexity value); in the upper and lower parts of the map, set different values [0, 0.1] (small) and [0.9, 1] (large) to distinguish the complexity. Compare the two results of the algorithm.

From the results, no matter the distribution of complexity, the searching route of the distance-aimed algorithm is constant (see Figs. 2a and 3a). On the premise of avoiding all of the static obstacles, the route is the shortest. For algorithm aimed at complexity, the searching route passes by the area of the low-complexity grid (the light-colored area). As shown in Fig. 2b, the upper half of the map is low

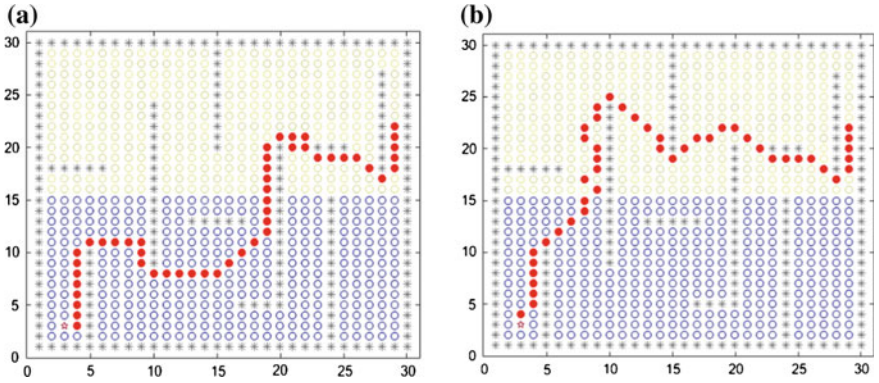


Fig. 2 Comparison of the algorithm with complexity alone versus distance alone (the up-complexity is high). *Panel a* shows distance only, and *panel b* shows complexity only. The *asterisks* are static obstacles

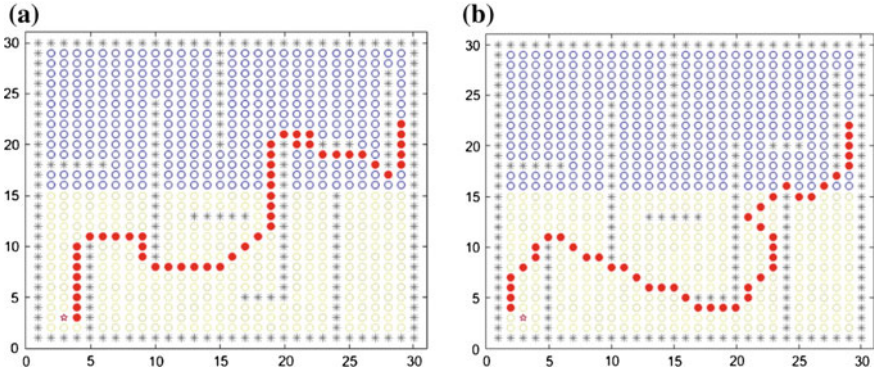


Fig. 3 Comparison of the algorithm with complexity alone versus distance alone (the up-complexity is low). *Panel a* shows distance only, and *panel b* shows complexity only. The *asterisks* are static obstacles

complexity, but the lower half's complexity is high, so the planning route is distributed mostly in upper half; on the contrary, the route in Fig. 3b is distributed mostly in the lower half.

4.2 Comparison of Complexity–Distance and Complexity Only

Set up a grid map of 30 * 30, and assign each grid random numbers as the complexity values. This experiment compared final distance and complexity values under two kinds of algorithm.

Table 1 Comparison of the cost of two algorithms

	Complexity only	Complexity–distance
Distance h (x)	18.0986	16.0346
Complexity g (x)	13.3558	13.6084

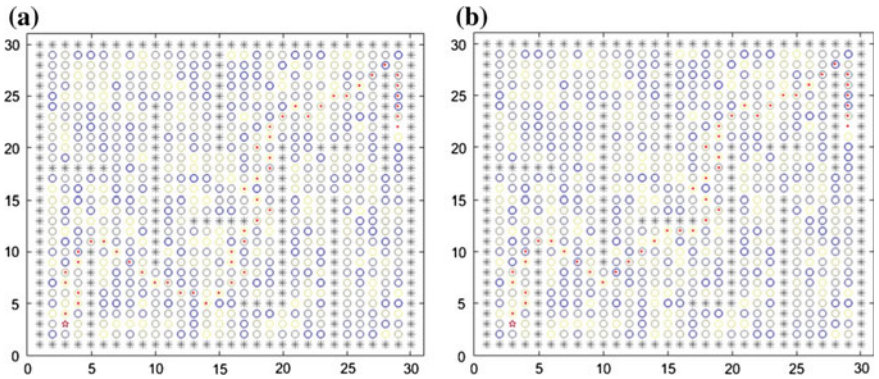


Fig. 4 Comparison of the algorithm with complexity only versus complexity–distance: *Panel a* is complexity only, and *panel b* is complexity–distance. The *asterisks* are static obstacles

Figure 4 shows the results of two kinds of algorithm. From the results, we can conclude the following:

(1) Although the two routes were different, the waypoints are distributed in the area of low complexity (the light-colored area), and this confirmed the conclusion of experiment 4.1.
(2) To compare the differences further, Table 1 provides the cost value of the two algorithms. As seen from the table, the algorithm aiming at complexity only has lower value of complexity, but the distance value is greater; the two gaps are obviously. In contrast, the algorithm aimed at complexity–distance has a slightly higher complexity value (i.e., slightly large), but the distance value is much lower than previous one, and the two gaps are also less (i.e., more reasonable).

4.3 Route Planning Under Real Conditions

The experiment was conducted in part of a segment from the Shenzhen West Coast. In order to make the test condition more close to reality, we set up 16 underway ships (moving objects). In the figure, the colored part of the channel indicates areas with complexity, and the deeper the color, the greater the complexity.

Figure 5 shows the results for route-planning (red line is the route). We can draw the following conclusions from the results:

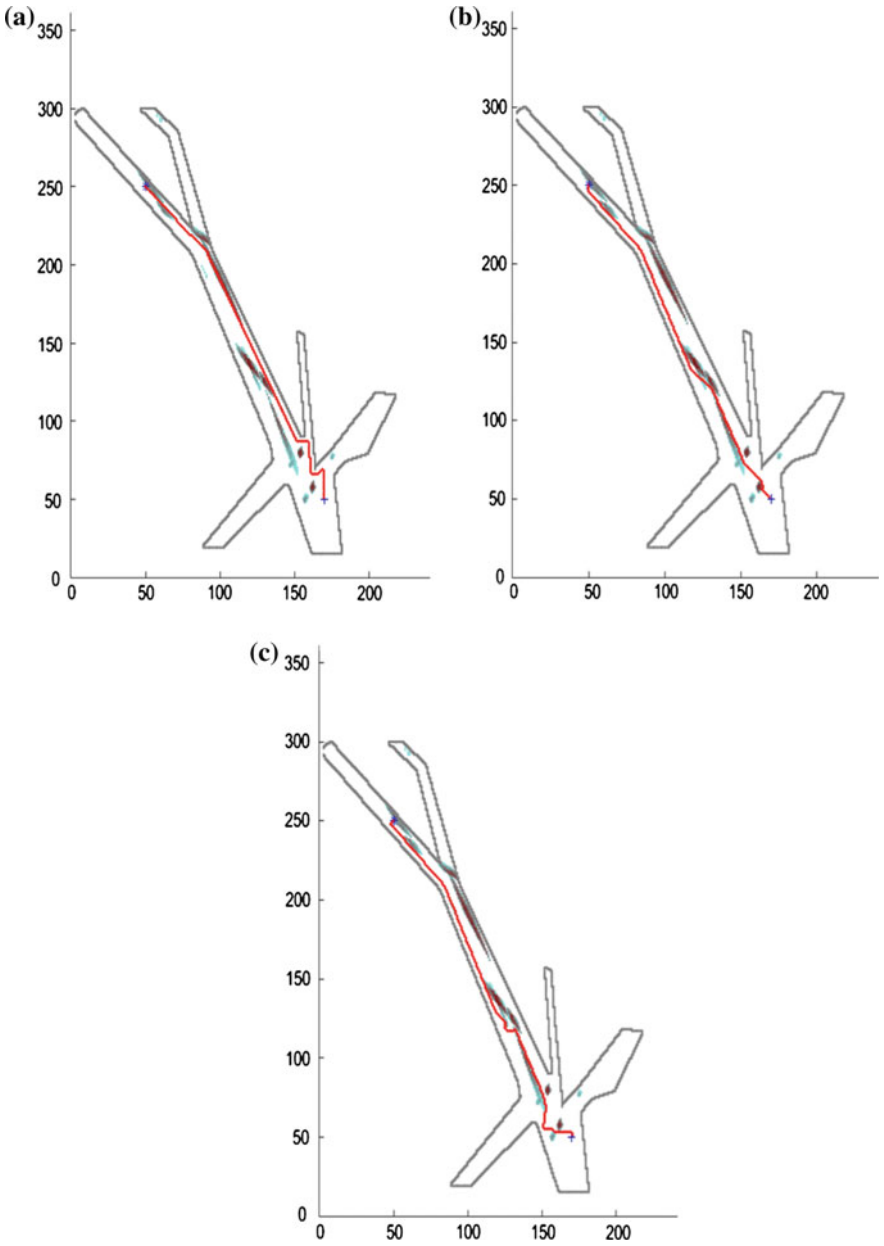


Fig. 5 The route-planning result along the Shenzhen West Coast: *Panel a* is distance only; *panel b* is complexity–distance and *panel c* is complexity only

1. Overall, the three planning routes are not too close to the channel line, i.e., the algorithm meets the constraint of ship size.
2. Fig. 5a shows the results when distance is the goal. The route trend is consistent with the direction of the channel extension, and there are few inflection points. However, the waypoints go through the complex areas, thus increasing the risk level of the entire route. Hence, the route is not safe.
3. Fig. 5c shows the results when complexity is the goal. The route avoided all the complex area, so this route is the safest. However, there are plenty of inflection points, and the route has more twists and turns; there are even detours. This, this route risks economy as well as the ecosystem.
4. Fig. 5b shows the results when complexity–distance is the goal. The route combined both of the above-mentioned two advantages: It is close to the shortest route in the direction of the target point, and it avoids most of the complex areas. In concrete terms, compared with Fig. 5a, the safety is increased greatly, and the distance was almost the same; compared with Fig. 5c, the distance was decreased significantly, but the security was not reduced by much.

5 Further Discussion

The complexity map is built by calculating the complexity value within each grid to get the distribution of complexity in the waters to be traveled. The grid-size setting can affect the accuracy of the complexity map and thus affect the result of route-planning. To explore this question, the authors conducted the following a set of comparative experiments (the length unit of the experiment is nautical miles [nm]).

In Sect. 1, part of the complexity-map modeling, the authors specified the size of the moving objects (i.e., the other ships): The length is approximately 0.12 nm, and the beam is approximately 0.025 nm. Considering the size of the ship domain, the ship was as abstracted as an ellipse with long axis 4 times the length and a short axis 1.6 times the beam. That is, the ellipse's long axis is 0.48 nm, and its short axis is 0.04 nm. Thus, the size of grid was set to 0.05 nm, which is slightly larger than the short axis of the ellipses. On the basis of this standard, the authors enlarged and shrank the grid, respectively, by twice its size to compare the planning route. The results of the experiment are shown in Fig. 6.

From the results, the three planning routes are approximately the same. There are differences in the middle segment where the complexity was populated. The complexity map in experiment x is too rough; compared with the other two experiments complex areas are missing from the channel. The complexity map in experiment z is the most accurate, and the continuity of the complexity distribution is the best; meanwhile, it avoided almost all the complex areas. To understand the experiment process further, we compared the cost and running time as listed in Table 2.

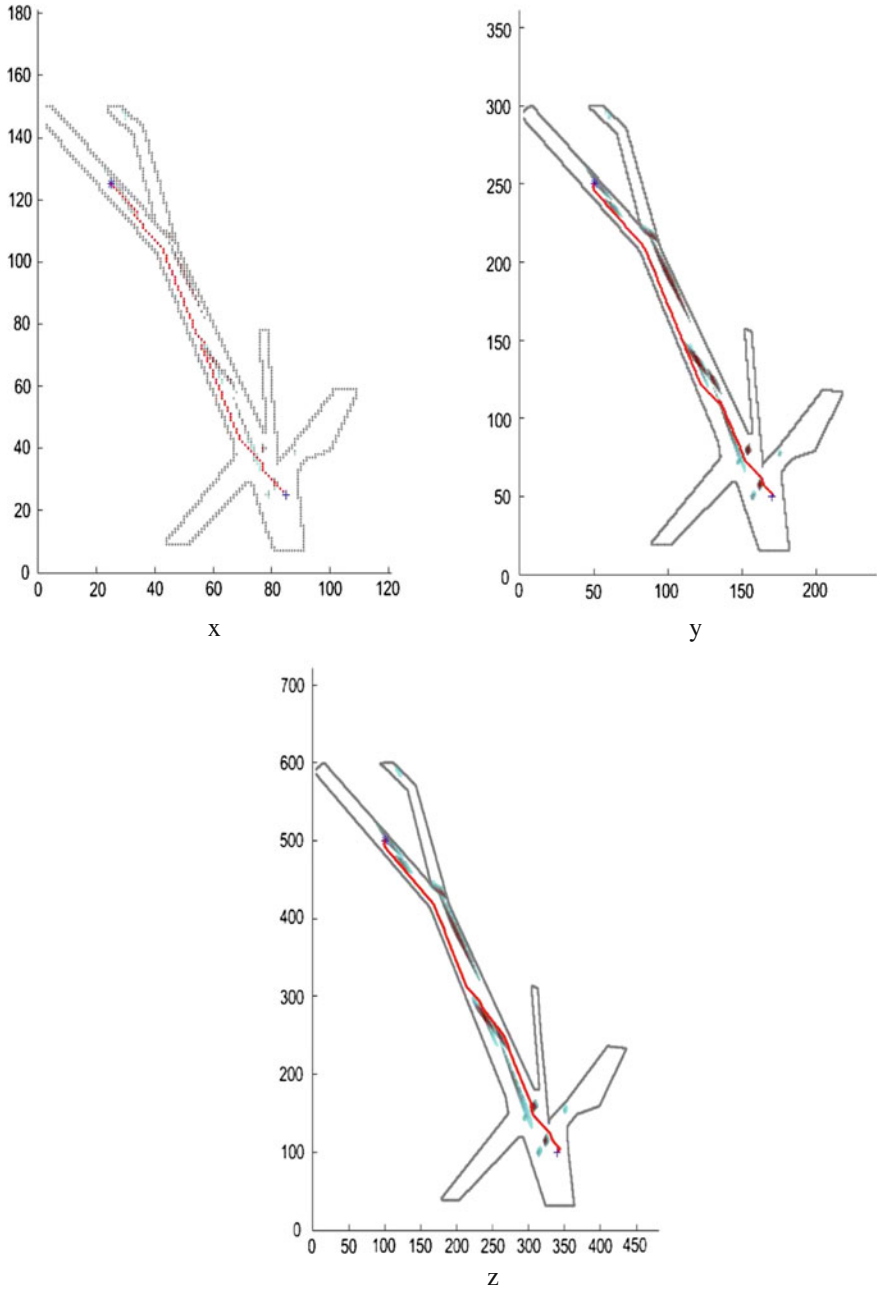


Fig. 6 Comparison of the route under different grid sizes: x is 0.1 nm; y is 0.05 nm; and z is 0.025 nm

Table 2 Comparison of the cost and running time

	0.1 nm (x)	0.05 nm (y)	0.025 nm (z)
Complexity (g)	2.2714	0.2049	0.0027
Distance (h)	53.4997	54.7675	54.6874
Running time (t)	41.4472	176.3031	1000.9454

From Table 2, we can draw the following conclusions

1. Regarding running time, experiment spanned the maximum number of grids and thus had the longest running time, longer than in the other two experiments. Spanning the fewest number of grids, experiment x had the shortest running time.
2. Regarding distance, there is no obvious difference among three experiments. The cost of experiment x is slightly less than that of the other two experiments. In addition, the distance of experiments y and z are almost the same.
3. Regarding complexity, the cost of experiment x is the greatest. This is because of the rough complexity map, for which the distribution of value is not complete. Thus, during the process of route-searching, it went through some complex areas. Moreover, the cost of experiment z is the lowest, almost 0, i.e., route z avoids all of the complex areas.

In addition, there is a need to explain why the cost value of the complexity and the distance differ so much in Table 2. In the process of the searching, the algorithm chose the grid, insofar as possible, with minimal complexity, namely, the value of grid is 0. Therefore, the value of the grid passed by the final planning route is mostly 0 because only a few grids have a complexity value. After accumulating, the final numerical is often very small. Although the distance cost is calculated from current point to end point, there is a distance cost from the starting point. After accumulating, the final numerical increases gradually. Therefore, from the numerical point of view, the final cost of distance is much larger than the complexity.

Synthesizing the previous discussion, the experiment proved that the smaller the grid, the better the planning results. However, enlarging the grid lead to considerably increased time, and considering the requirement of control time in the process of practical applications, the results argue that the appropriate grid size is slightly larger than the oval short axis of the ship domain.

6 Conclusion

The chapter puts forward a route-planning algorithm for navigating multiple mobile obstacles. The Complex Map was established based on the theory of complexity measurement. With the constraints of route security, route distance, and ship dimensions, a route was planned in a complex water environment using the idea of the A* algorithm and the Complex Map. The experimental results show that in the

case of satisfying the constraint of ship dimensions, the best path is close to the shortest route in the direction of the target point and avoids all of the high complexity areas.

However, in dealing with the ship, the algorithm considered it as a simple rectangle. In reality, it is much more complicated. In addition, the characteristics of ship-motion dynamics also influence the outcomes of the final route-planning. These problems will be studied in future work.

Acknowledgements This work was supported by the National Science Foundation of China (Grant No. 51679180) and Double First-Rate Project of WUT; the National Science Foundation of China (Grant No. 51579204) and Double First-Rate Project of WUT; and the China Postdoctoral Science Foundation (Grant No. 2016M602382).

References

1. Zong G, Hu B, Han J (2012) Research on complexity spatial structure for chinese coastal port network. China Soft Sci 12:171–178
2. Hua Z (2011) Key technology research on route planning for UAV based on intelligent optimization algorithm. Nanjing University of Aeronautics and Astronautics, Nanjing
3. Hao Y, Liu Y (2015) Review of genetic operators applied in VRP. J Southeast Univ (Philosophy and Social Science) S2:85–87
4. Ji M, Chen Z, Wang Q (2011) Optimization algorithm for regional transportation route of container ship. J Transp Eng 04:68–75
5. Li X, Xiao J, Wang X (2011) Shipping route optimization based on tabu-search algorithm. Intelligent information technology application association. 2011 International Conference on Machine Intelligence (ICMI 2011 V4). Intelligent Information Technology Application Association, 2011:6
6. Zhang Y (2013) Path planning research on large ship material handling system. Huazhong University of Science and Technology
7. Chen J (2013) Study of path planning algorithm on unmanned rescue ship. Wuhan University of Technology
8. Phanthong T, Maki T, Ura T, Sakamaki T, Aiyarak P (2014) Application of A* algorithm for real-time path re-planning of an unmanned surface vehicle avoiding underwater obstacles. J Mar Sci Appl 01:105–116
9. Xie S, Wu P, Liu H, Yan P, Li X, Luo J, Li Q (2016) A novel method of unmanned surfacevehicle autonomous cruise. Ind Robot Int J 43(1):121–130
10. Zhang J, Minghua Hu, Zhang C (2009) Research on complexity on air traffic management. Acta Aeronautica et Astronautica Sinica 30(11):2132–2142
11. Hong Y, Kim Y, Lee K (2013) Application of complexity assessment for conflict resolution in Air Trac Management system. AIAA Guidance, Navigation, and Control (GNC) Conference, Boston, US
12. Huang Y (2014) Research on water traffic complexity measure. Wuhan University of Technology
13. Wen Y, Huang Y, Yang J, Xiao C, Zhou C, Wu X (2014) Modeling for complexity of water traffic flow structure. Navig China 02:62–68
14. Hart PE, Nilsson NJ, Raphael B (1968) A formal basis for the heuristic determination of minimum cost paths. IEEE Trans Syst Sci Cybern SSC-4(2):100–107
15. Guo Y, Yi P (2008) Properties analysis for method of linear dimensionless. Stat Res 02:93–100

Part III
Intelligent GIS Integration with Acoustic, Remote-Sensing, and Radar Systems

Calibration and Verification of Models Defining Radar-Visibility Zones in Marine Geoinformation Systems

Sergey Vavilov and Mikhail Lytaev

Abstract In the present research, we are concerned with the problem of forecasting marine radar-visibility zones under various meteorological conditions. The existing widespread models describing the electromagnetic wave propagation around the Earth, such as AREPS and PETOOL, demonstrate the presence of some weak points in their conceptions. The method of calibration and verification of the models, which contain artificial unknown parameters, was elaborated to obtain the desired accuracy in calculations of the electromagnetic field.

Keywords Radiowave propagation · Radar-visibility zone · Inhomogeneous medium

1 Introduction

Determination of radio-visibility zones under various meteorological conditions and terrain structures constitutes an important part in the design and construction of radars, provision of wireless communication, and ensuring navigation safety [1]. Spatial variations of the refractive index and characteristics of terrain may have a significant impact on the propagation of electromagnetic waves [2]. The effects of tropospheric propagation, which strongly depend on particular meteorological conditions, can significantly increase the zone of radio-visibility, but also provoke the existence of invisible areas for radar. The above-mentioned areas are also known in the literature as "skipping zones" and "radar holes." Special ducts arising

S. Vavilov
Saint Petersburg State University, St. Petersburg, Russia
e-mail: savavilov@inbox.ru

M. Lytaev (✉)
The Bonch-Bruevich Saint Petersburg State University of Telecommunications, St. Petersburg, Russia
e-mail: mikelytaev@gmail.com

V. Popovich et al. (eds.), *Information Fusion and Intelligent Geographic Information Systems (IF&IGIS'17)*, Lecture Notes in Geoinformation and Cartography, DOI 10.1007/978-3-319-59539-9_9

along the wave-propagation path capture of almost all electromagnetic energy within the ducts and simultaneously strongly diminish the radar coverage outside of them. Thus, the object positioned outside the mentioned waveguides may remain undetected even if their presence was expected in the coverage area of the radar.

At the present time, the use of geographical information systems (GIS) for the determination and evaluation of areas of radio-visibility, for instance AREPS and TEMPER [3], have gained increasing popularity. Such systems are capable of seeking meteorological real-time information from various sources along with further processing by making use of special numerical algorithms, which also provide the desired visualization in the most convenient form. An important component of such software is the subroutine, which provides calculations of electromagnetic-wave propagation in the inhomogeneous medium. Analysis of the existing widespread models, describing the electromagnetic-wave propagation in the vicinity of the Earth, demonstrates the presence of some weak points in their conceptions. In particular, it is relevant to mention the well-known fact that the split-step Fourier method, which is used in the above-mentioned systems to provide the numerical solution to the problem within the method of parabolic equation, yields false reflections of the waves [4]. These artificial reflections arise because of the compelled truncation of the originally infinite integration area. In order to get rid of the above-mentioned false reflections, a special absorption layer is to be placed in the vicinity of the upper boundary of the truncated domain. The parameters of such a layer strongly depend on the characteristics of the transmitting antenna, the terrain parameters, and the troposphere refractive index. Moreover, the parameters of the absorption layer are to be determined as a result of some procedure that also includes empirical considerations, which makes calibration and verification of the procedure necessary. Despite these drawbacks, we should keep in mind that the vast popularity and wide application of the split-step Fourier method may be easily explained by its high numerical efficiency. Thus, the main goal of our study was to elaborate the method of calculating electromagnetic wave–propagation in the vicinity of the Earth by the transmitting antenna, without introducing an artificial absorption layer, in the original elliptic formulation of the problem. It conferred an opportunity to provide the calibration procedure and verification of the exactness of the existing methods and also to determine the range of their limitations from the application point of view. One should notice that in contrast to split-step Fourier method, the high numerical efficiency of such software is not obligatory.

This chapter is organized as follows. In the next section, we give the mathematical formulation of the problem. Reformulation of the original problem on the basis of the “method of the equivalent sources” is given in Sect. 3. Simultaneously we derive the Fredholm second-kind integral equation, which provides a solution to the reformulated problem. Additionally, we point the numerical scheme for construction of an approximate solution. Numerical comparison of the “method of parabolic equation” and the “method of equivalent sources” is given in the process of the radiowave-propagation study in Sect. 4.

2 Study Problem

In this research, we present a preferred systematic approach to solving a class of problems modelling wave propagation called the "paraxial direction." The problem of wave propagation is associated with the appropriate complex field component $\psi(x, z)$ defined on the set: $0 \le x < \infty$, $0 \le z < \infty$, where x and z are longitudinal and transversal variables, respectively. The function $\psi(x, z)$ follows the Helmholtz equation:

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial z^2} + k^2(1 + i\alpha + n(x, z))\psi = 0 \tag{1}$$

where $k = 2\pi/\mu$ is the wave number in a vacuum; μ is the wave length; $\alpha > 0$ is a small parameter; and $n(x, z)$ is the inhomogeneous component of the refractive index. As shown below, the introduction of a small amount of loss through parameter α provides the uniqueness of a solution to the problem sought owing to the principle of limiting absorption. Furthermore, function $\psi(x, z)$ is subject to the impedance boundary condition:

$$\left(\frac{\partial \psi}{\partial z} + q\psi\right)\Big|_{z=0} = 0 \tag{2}$$

where $q = q_1 + iq_2$ is a complex number with $q_2 > 0$. The most distinctive feature of the considered problem consists in the fact that the wave field is generated by the additional condition

$$\psi(0, z) = \psi_0(z) \tag{3}$$

with the given function $\psi_0(z)$. Moreover, we are seeking a solution to problems (1)–(3) that is stable to the function $\psi_0(z)$ perturbations in one or the other reasonable metrics. In our further study, we will proceed from the assumption that $\psi_0(z)$ belongs to the $L_2[0, \infty)$ functional space. The reason to set condition (3) is stipulated by the fact that the aperture field is not directly known, but instead the source is defined through its far-field beam pattern. For instance, the Gaussian beam patterns are often used: Besides having excellent numerical properties, they provide a good representation for paraboloid dish antennas. The aperture field corresponding to a source at height z_0 for a Gaussian beam of half-power beam width β and elevation angle θ_0 generates function $\psi_0(z)$, which is defined as follows:

$$\psi_0(z) = A\frac{k\beta}{2\sqrt{\pi}\log 2}\exp(-ik\theta_0 z)\exp\left(-\frac{\beta^2}{8\log 2}k^2(z - z_0)^2\right) \tag{4}$$

where A is a normalization constant. From here on in, for the sake of brevity we will write Eq. (1) in the dimensionless form by putting $k = 1$. The latter implies the introduction of dimensionless variables x and z by setting $x := xk$, $z := zk$. In our

reasoning, we proceed from the fact that the generated waves propagate mostly in a preferred direction, called the "paraxial direction." In the present study, the positive x-dimension was chosen as the paraxial one. Thus, we discard waves propagating in the opposite direction. Nevertheless, it is worth noting that in contrast to the parabolic-equation method, including its modifications based on a Pade approximation [2], we do not suppose that the energy propagates mostly in a narrow-angle or a wide-angle cone centered on the preferred direction. In our study original problem (1)–(3) is reduced to the Helmholtz equation with the "equivalent source" $Q_0(z)\delta(x-x_0)$ on its right-hand side, where δ is the Dirac function; and $x_0<0$ is taken in a vicinity of zero. Function $Q_0(z)$ is defined through $\psi_0(z)$ as a solution to the first-kind Fredholm integral equation. One should emphasise that in the present research, we are not limited to a specific width of the wave-propagation angle.

3 Problem Reformulation

Suppose that function $V(x,z)$, defined on the set $\Omega:\{-\infty<x<\infty, 0\le z<\infty\}$, follows the Helmholtz equation:

$$\frac{\partial^2 V}{\partial x^2}+\frac{\partial^2 V}{\partial z^2}+(1+i\alpha+m(x,z))V=Q_0(z)\delta(x-x_0) \tag{5}$$

with a priori unknown function $Q_0(z)$ and $m(x,z)=0$ when x belongs to the interval $[2x_0,0]$ and $m(x,z)=n(|x|,z)$ when $x>0$ and $x<2x_0$, whereas $n(x,z)$ is expanded as an even function for all negative values of x. The choice of $x_0<0$ is clarified later in the text. Additionally $V(x,z)$ is subject to the impedance boundary condition as follows:

$$\left(\frac{\partial V}{\partial z}+qV\right)\Big|_{z=0}=0 \tag{6}$$

and the principle of limiting absorption.

In one line with $V(x,z)$, we take into consideration the auxiliary field component $W(x,z)$ defined on the same set and satisfy the "truncated" Helmholtz equation as follows:

$$\frac{\partial^2 W}{\partial x^2}+\frac{\partial^2 W}{\partial z^2}+(1+i\alpha)W=Q_0(z)\delta(x-x_0) \tag{7}$$

subject to the same restrictions

$$\left(\frac{\partial W}{\partial z}+qW\right)\Big|_{z=0}=0 \tag{8}$$

including the principle of limiting absorption.

In our reasoning, we proceed from the fact that the wave propagation in problem (1)–(3) takes place in the preferred x-positive direction. In this connection, we introduce the interrelationships between field components ψ, V and W as it is presented in Fig. 1. Consequently, the field component W generated by the source $Q_0(z)\delta(x - x_0)$ should match function $\psi_0(z)$ at $x = 0$. By making use of the Fourier transformation with respect to the x variable, relationships (7) and (8) may be rewritten in the following form:

$$\frac{d^2\hat{W}}{dz^2} + \left(1 - \lambda^2 + i\alpha\right)\hat{W} = Q_0(z)e^{-i\lambda x_0}$$
$$\left(\frac{d\hat{W}}{dz} + q\hat{W}\right)\Big|_{z=0} = 0$$

where $\hat{W}(\lambda, z) = \int_{-\infty}^{+\infty} W(x, z)e^{-i\lambda x}dx$.

Consider the Green function $\hat{G}(z, z', \lambda)$, which is defined as follows:

$$\frac{d^2\hat{G}}{dz^2} + \left(1 - \lambda^2 + i\alpha\right)\hat{G} = \delta(z - z')$$
$$\hat{G}\big|_{z\to z'+} = \hat{G}\big|_{z\to z'-},$$
$$\frac{d\hat{G}}{dz}\Big|_{z\to z'+} - \frac{d\hat{G}}{dz}\Big|_{z\to z'-} = 1,$$
$$\left(\frac{d\hat{G}}{dz} + q\hat{G}\right)\Big|_{z=0} = 0.$$

Keeping in mind the principle of limiting absorption, one can write it down in the following explicit form:

$$\hat{G}(z, z', \lambda) = -\frac{1}{2\gamma}e^{-\gamma|z-z'|} + \frac{q+\gamma}{2\gamma(q-\gamma)}e^{-\gamma(z+z')},$$

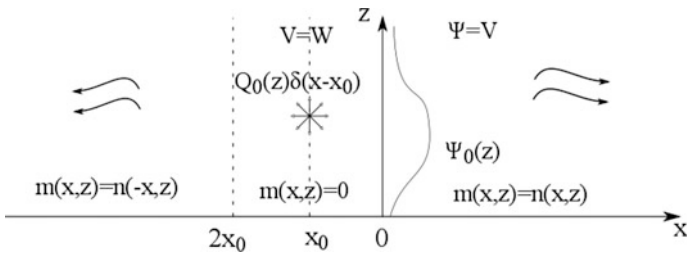


Fig. 1 Interrelationships between field components ψ, V and W

where $\gamma = a + bi$,

$$a = \sqrt{\frac{\sqrt{\left(1-\lambda^2\right)^2+\alpha^2}-\left(1-\lambda^2\right)}{2}},\ b = -\sqrt{\frac{\sqrt{\left(1-\lambda^2\right)^2+\alpha^2}-\left(1-\lambda^2\right)}{2}}$$

Thus, we arrive at the following relationship:

$$W(x,z) = \frac{1}{2\pi}\int\limits_{-\infty}^{+\infty}\int\limits_{0}^{+\infty}\hat{G}(z,z',\lambda)Q_0(z')e^{-i\lambda x_0}e^{i\lambda x}dz'd\lambda.$$

The equality $W(0,z) = \psi_0(z)$ implies that the following:

$$\frac{1}{2\pi}\int\limits_{-\infty}^{+\infty}\int\limits_{0}^{+\infty}\hat{G}(z,z',\lambda)Q_0(z')e^{-i\lambda x_0}dz'd\lambda = \psi_0(z).$$

Changing the order of integration and taking into account the following function:

$$K(z,z',x_0) = \frac{1}{2\pi}\int\limits_{-\infty}^{+\infty}\hat{G}(z,z',\lambda)e^{-i\lambda x_0}d\lambda \tag{9}$$

we arrive at the first-kind Fredholm integral equation with respect to function $Q_0(z)$ as follows:

$$\int\limits_{0}^{+\infty}K(z,z',x_0)Q_0(z')dz' = \psi_0(z). \tag{10}$$

The value of $x_0 < 0$ may be taken arbitrarily but is not equal to zero. The point is that $x_0 \neq 0$ provides convergence of the integral on the right-hand side of (9) when $z = z'$ because of the Dirichlet convergence criterion. We also pay attention to the fact that the field component V becomes symmetric with respect to x_0 when x_0 aspires to zero.

Equation (10) is an ill-posed problem that is tackled by making use of specially elaborated methods [5]. One should notice that exponentially rapid convergence of the kernel in (10) to zero at the infinity gives us the opportunity to realize a truncation procedure. Thus, the infinite integration domain in (10) is substituted for the succession of bounded sets expanded to the original one. In the class of functions with limited variation, the solution to such type of integral equations, i.e., those with continuous kernel, bounded integration domain, and the right-hand side from L_2 space, may be constructed in the framework of the "quasisolution" notion

[6]. This procedure is easily realized numerically and provides the piecewise uniform convergence of approximations. Moreover, the obtained quasisolution to Eq. 10 is stable to the perturbations of its right-hand side $\psi_0(z)$ in $L_2[0, +\infty)$ metrics.

In the new statement of, problem (5), (6), (10) the problem benefits greatly because it is amenable to a solution by special methods [7]. In particular, when the refractive index does not depend on x, i.e., $n(x, z) = n(z)$, the problem is reduced to the second-kind Fredholm integral equation with respect to function $\hat{V}(\lambda, z)$ as follows:

$$\hat{V}(\lambda, z) + \int_0^{+\infty} \hat{G}(z, z', \lambda) n(z') \hat{V}(\lambda, z') dz' = \int_0^{+\infty} \hat{G}(z, z', \lambda) e^{-i\lambda x_0} Q_0(z') dz', \quad (11)$$

where $\hat{V}(\lambda, z) = \int_{-\infty}^{+\infty} V(x, z) e^{-i\lambda x} dx$. One should notice that in Eq. 11, function m is substituted for n as x_0 may be taken arbitrary close to zero. To compute the inverse Fourier transformation, an effective adaptive integration method based on work [8] was elaborated. It is remarkable that in contrast to split-step Fourier method for the parabolic equation in the case $n(z)$ when it is a finite (compactly supported) function with respect to z, the original problem (1)–(3) may be solved directly within the limits of the proposed approach through Eq. (11) without introduction of the artificial absorption layer [2]. The latter means that under the pointed-out circumstances, problem (5), (6), (10) is reduced to a second-kind integral equation with the Fredholm kernel. In addition, the existence of the Fourier inverse transformation may be also easily proven in this case.

4 Numerical Examples

The validation, verification, and calibration (VV&C) process with respect to radiowave-propagation problems was applied to modelling software PETOOL in [9], but only in the case when the refractive index is linearly decreasing with the height growth. The point is that in this case the problem admits analytical solution. The proposed method makes it possible to generate calibration tests with respect to any finite refractive index, to any arbitrary antenna pattern, and to each boundary condition containing impedance.

We will conduct simulations for the refractive index depicted in Fig. 2 and in the case of a Gaussian antenna (4) with the beam width $\beta = 35^0$. This antenna is located at the altitude of 10 m above the sea surface and radiates a signal equal to 3 GHz. The height of the computational domain and the absorbing layer in the simulation, as calculated by the method of parabolic equation, was taken as 400 and 200 m, respectively. Furthermore, we introduce the notion of the propagation factor PF, which is connected with the field component V by the dependence as follows:

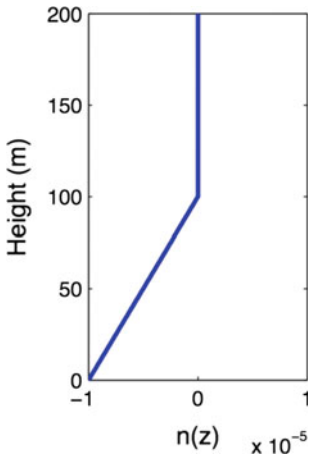


Fig. 2 Dependence of the refractive index with respect to height

$$PF = 20\log|V| + 10\log x + 10\log\ \mu.$$

One can see in Fig. 3 that the PF calculated using the above-mentioned different methods is not distinguishable at distances ≤ 50 km from the source of radiation. However, the method of the parabolic equation gives a significant overestimation of the electromagnetic field in comparison with the method of the equivalent sources at ranges >50 km. Apparently, such a discrepancy is caused by the bad choice of the artificial parameters in the split-step Fourier method. It is remarkable that usually it is possible to achieve good matching of the above-mentioned methods for a wide range of distances by increasing the height of the computational domain and selecting the appropriate weight function for the absorption layer. This fact indicates that the weak point of the split-step Fourier method lies in the problem of choosing the suitable initial parameters for its realization. Simultaneously, the numerical realization of the method of equivalent sources does not require manual selection of the input parameters including the meshes sizes for the longitudinal and vertical variables. Their choice is based exclusively on the requirements of visualization and does not affect the accuracy of calculations.

Furthermore, we demonstrate the results of simulation for a medium with the refractive index depicted in Fig. 4. The antenna is located at an altitude of 20 m and emits a signal with a frequency of 10 GHz. In this case, the height of the computational domain and the absorbing layer were chosen equal to 3000 m. One can see in Fig. 5 that for the chosen parameters, both methods demonstrate an acceptable coincidence. The spatial two-dimensional distribution of the propagation factor is depicted in Fig. 6.

The maximal dynamical range of the calculations in a standard arithmetic equation of double precision constitutes the same value for both of the methods, which is equal to approximately 110 dB. Here we are to keep in mind that in the method of equivalent sources, the accuracy of calculations is always achieved

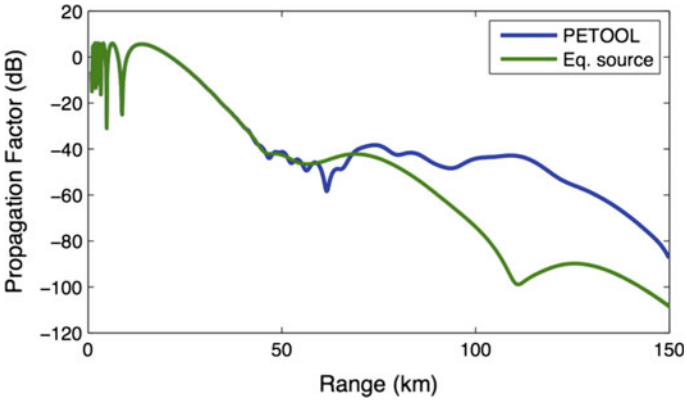


Fig. 3 Propagation factor at 50-m altitude

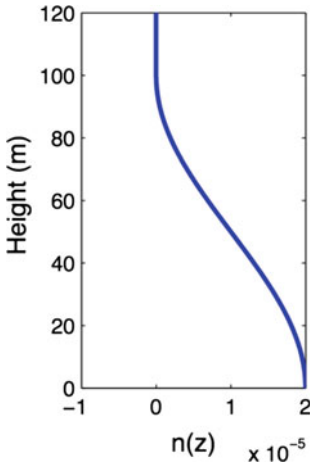


Fig. 4 Dependence of the refractive index with respect to height

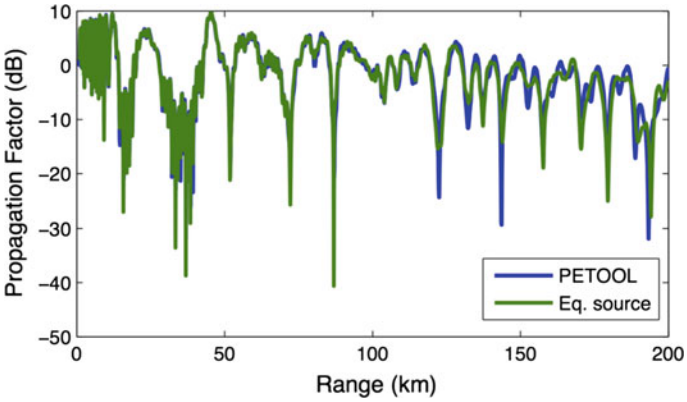


Fig. 5 Propagation factor at 70-m altitude

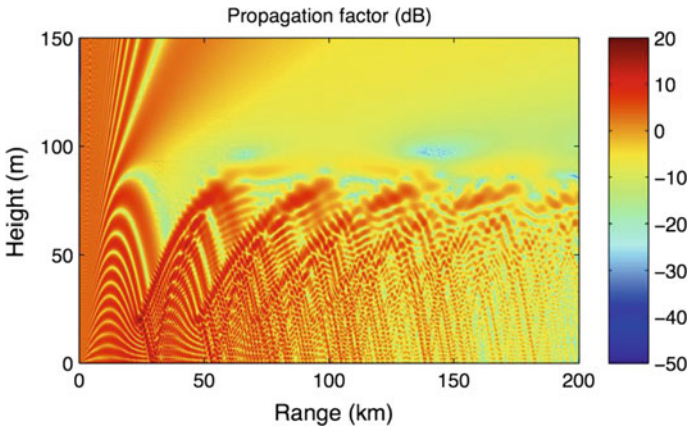


Fig. 6 Propagation factor in a duct

automatically owing to use of the adaptive algorithm of numerical integration. In contrast to the method of equivalent sources, while solving the corresponding parabolic equation by split-step Fourier method it is usually impossible to distinguish errors of numerical integration from false reflections from the upper boundary of the integration domain. In should be noted that despite the advantages of the method of equivalent sources, the numerical efficiency of the split-step Fourier method, even with properly selected parameters of the computational meshes and the absorbing layer, turns out to be several times higher. Therefore, only algorithms based on split-step Fourier method for parabolic equations remain suitable for prediction of the effects of the tropospheric influence on electromagnetic wave–propagation in real time. Additionally, one should pay attention to the fact that the method of equivalent sources proves useful not only for the purposes of calibration and verification of the existing GIS forecasting the radar-visibility zones under various meteorological conditions: This method may be also used in applications that require a high guaranteed accuracy of calculations, for example, in the planning and analysis of experimental results.

5 Conclusion

In the chapter, we proposed a novel approach to the calculation of an electromagnetic field in an inhomogeneous troposphere. In contrast to the method of parabolic equation and algorithms of its numerical solution, the proposed approach gives an opportunity to solve the original problem without introducing an artificial absorbing layer and other empirical schemes. Considering the problem within the paraxial approximation, we do not suppose that the energy propagates mostly in a narrow-angle or a wide-angle cone centered on the preferred direction. The

proposed method does not impose restrictions on the antenna-beam width and the maximum value of the propagation angle. In addition, the proposed algorithms of constructing the numerical solution do not require manual selection of artificial parameters depending on the initial data, and the required precision is achieved automatically in the process of solving the problem. In comparison with the split-step Fourier method, the above-mentioned algorithm proved its high effectiveness, in particular, for calibration and verification of numerical methods providing the functioning of the corresponding forecasting systems of radar visibility within the method of parabolic equation.

References

1. Smirnova OV, Svetlichny VA (2015) Geoinformation systems for maritime radar visibility zone modelling. Information Fusion and Geographic Information Systems, Springer, pp 161–172
2. Levy MF (2000) Parabolic equation methods for electromagnetic wave propagation. The Institution of Electrical Engineers, London
3. Brookner E, Cornely PR, Lok YF (2007) AREPS and TEMPER-getting familiar with these powerful propagation software Tools. IEEE Radar Conference, pp 1034–1043
4. Zhang P, Bai L, Wu Z, Guo L (2016) Applying the parabolic equation to tropospheric groundwave propagation: a review of recent achievements and significant milestones. IEEE Trans Antennas Propag 58:31–44
5. Ivanov VK, Vasin VV, Tanana VP (2002) Theory of Linear Ill-posed problems and its applications. VSP, Utrecht
6. Tikhonov AN, Arsenin VY (1977) Solution of Ill-posed problems. Halsted Press, New York
7. Atkinson KE (1997) The numerical solution of integral equations of the second kind. Cambridge University Press, New York
8. Domínguez V, Graham IG, Smyshlyaev VP (2011) Stability and error estimates for filon-clenshaw-curtis rules for highly oscillatory integrals. IMA J Numer Anal 31:1253–1280
9. Ozgun O, Apaydin G, Kuzuoglu M, Sevgi L (2011) PETOOL: MATLAB-based one-way and two-way split-step parabolic equation tool for radiowave propagation over variable terrain. Comput Phys Commun 182:2638–2654

Algorithmic Component of an Earth Remote-Sensing Data-Analysis System

Vasily Popovich and Filipp Galiano

Abstract The article describes the algorithm component of universal an Earth remote-sensing data (RDS) system. A comparative analysis of algorithms that are used to solve problems related to all stages of analysis—preliminary processing, selection of informative elements, description of informative elements, and classification of descriptions as well as their interfaces—is proposed. Data structures to store raw data and intermediate calculations are discussed. Theoretical results are confirmed with computer experiments and practical realization in an RSD-analysis system. The structure of the developed universal RSD-analysis system and an example of scenarios for the analysis of real data is considered.

Keywords Remote-sensing data · Image processing · Segmentation · Classification · Scenario approach

1 Introduction

The use of remote-sensing satellites, as well as scope of their application, is developing rapidly. Today it is difficult to detect those branches of knowledge, industry, and business where space technology has not yet been applied. However, widespread use of remote-sensing data forms a whole set of new problems to be solved. An example of such problems is the operational analysis of remote-sensing data. Manual analysis methods cannot be effective because they require many human resources, which leads to the high price of such analysis as well as unacceptable delays.

Systems for the analysis of remote-sensing data are based on the integration of different analysis methods from different knowledge domains such as image pro-

V. Popovich · F. Galiano (✉)
SPIIRAS Hi Tech Research and Development Office Ltd, 14 Line, St. Petersburg, Russia
e-mail: galiano@oogis.ru

V. Popovich
e-mail: popovich@oogis.ru

V. Popovich et al. (eds.), *Information Fusion and Intelligent Geographic Information Systems (IF&IGIS'17)*, Lecture Notes in Geoinformation and Cartography, DOI 10.1007/978-3-319-59539-9_10

cessing and pattern recognition. Main stages of analysis of remote sensing include the following:

- preliminary processing of image(s);
- selection of informative elements such as segments, corners, edges, etc. (segmentation algorithms are one of most used on this stage because they allow to solve the wide range of analysis tasks);
- description of informative elements; and
- classification of descriptions.

On the last stage, classical methods of classification and clustering (artificial neural networks, cluster analysis, etc.) can be used. Important peculiarities of remote-sensing data analysis include the following:

- considerable size of the image at hand; and
- related analysis of the images corresponded to several spectral bands.

These peculiarities impose limits on the computational complexity of algorithms used for analysis. The common peculiarities of commercial systems used for the analysis of remote-sensing data include the following:

- There is a need for significant work of highly qualified experts for the selection, specification, and setting of algorithms to be used for data analysis. Direct application of classical methods is complicated due to high computational complexity of the task.
- It is difficult to reuse algorithms, which often leads to having to develop a unique system for each class of tasks.

In general, is impossible to solve practical problems using a single algorithm sequence; a combination of separate algorithms is required for constructing the analysis process. Thus, to automate the analysis of different types of remote-sensing data for different purposes, it is important to develop a set of algorithms, each one of which has a standardized interface. This approach allows to adapt the system to the peculiarities of the subject domain without restricting the set of algorithm used. Let us consider the components of the system here.

2 Basic Classes of Analysis Algorithms

Algorithms of RSD analysis can be divided into classes according to the analysis stage. The basic classes of analysis algorithms are listed in Table 1. Typical input parameters are not specified because they vary widely from algorithm to algorithm.

Structures for representation of the original and processed data can vary significantly (Table 2). Methods of data formal representation in feature space depend on the data itself and the classification algorithms used. The size of the training set used to train the classification algorithm is usually significantly smaller than the number of pixels in the image, which reduces the requirements of the data structure

Table 1 Basic classes of RSD-analysis algorithms

No.	Class	Important subclasses	Input data	Output data	Example of algorithms	Note
1.	Preliminary processing	Histogram-based algorithms, local-filtration algorithms	Image, in most cases multi-spectral	Image, in most cases multi-spectral	Median filtration, equalization	Used for visualization and subsequent processing
2.		Segmentation, selection of corners or edges	Image, in most cases multi-spectral	List of recognized objects and coordinates	Flood fill segmentation, Harris corner detector, Canny edge detector	For segmentation; computationally effective calculations can be achieved using only dynamic trees Kharinov [3]
3.		Calculation of segments, properties, or edge parameters	Dynamic tree, list of points, etc.	Vector of properties (real numbers from a given range) that describe informative element	Segment perimeter or average bright	Methods vary significantly for different subject domains
4.		ANN, Bayesian classification, etc.	Vector of properties (real numbers from a given range) that describe informative elements	Class tag assigned to corresponding element	Nearest neighborhood classification, «naive» Bayes	For this stage a wide set of algorithms has been developed Gonzalez and Woods [5]
5.	Post-processing	–	Input image, coordinates of recognized objects	Properties of recognized objects	Course of detected ship	Task-specified
6.	Use of geospatial information	–	Input image (georeferenced)	Geodata calculated for particular part of Earth surface	Calculation of Earth surface mask	Can be used in different stages of analysis

Table 2 Data structures for different levels of analysis

Object class	Subclass	Data structure	Advantages	Disadvantages
Single-band (gray-scale) image	Input image	Array of integer numbers	Absence of error accumulation in sequential operations	Need to check for integer overflow in certain operations
		Array of real numbers	Image can be used for representation of intermediate operations	Direct visualization is impossible
	Segmented image (one level of segmentation)	An associative array (key is the pixel coordinates, value is the pixel brightness)	Native support in many programming languages	High computational costs
		Dynamic tree	High speed of computations	Complex software realization
	Segmented image (hierarchy of segmentation levels)	Multi-dimensional associative array	Native support in many programming languages	High memory consumption
		Indexed tree	High speed of operation on property values	Complex software realization
Feature space	Explicit representation	Two-dimensional array of real numbers	Ability to improve classification accuracy by analyzing the relationship between property values	High computational costs
	Implicit representation	Indexed tree	High speed of operation on property values	Classification accuracy is slightly lower

used to store them. In many cases, matrices or vectors of features are used. Its representation is one- or two-dimensional arrays of real numbers. In the simplest case, the dimension of the feature vector is equal to the spectral resolution of the remote-sensing data multiplied by the number of features calculated for each spectral band, and their number does not exceed the number of pixels of the image. For segmented images, the optimal data structure is a dynamic tree (so-called "system of disjoint sets") [1], Sleator and Tarjan [2]. If image segmentation at various levels of approximation forms a hierarchy, an indexed (dynamic) tree can

be used. This uses a fixed amount of memory for storing a hierarchy of image partitioning into disjoint segments [3]. The results of contour detection can also be stored in an indexed tree and processed as segments.

All of the classification algorithms can be divided into two groups: (1) ones that require presence of all the training samples at a time or (2) ones that are capable of incremental learning [4]. The algorithms of the latter class can significantly reduce the memory requirements for the classification, but they sometimes demonstrate lower classification accuracy.

3 Preliminary-Processing Algorithms

A significant impact on the accuracy and stability of the analysis is provided by image preprocessing algorithms. The distinguishing feature of this type of algorithms is that their input and output data is an image. However, not every image is obtained because of preliminary processing can be visualized directly because the range of pixel brightness can exceed the dynamic range of the display device used.

Among preliminary-processing algorithms, the two largest classes are brightness transformation and local filtering. First-class algorithms are used for visualization, and they provide stability of the image histogram parameters for subsequent analysis. The latter are mainly used for noise suppression and isolation of informative features on the image (contours, homogeneous areas, etc.). A considerable part of the filters are based on fast Fourier transform or simultaneous analysis of the frequency and spatial information (using wavelet transform) [5].

Example of a brightness-conversion algorithm is an invariant representation of the image [6, 7]. The invariant representation of the image at full-brightness resolution is obtained by the arithmetic transformation of the image pixels' brightness. Representation is considered invariant because it does not depend on transformation of the image-brightness scale if such transformation preserves the order and number of bins (non-zero elements) of the histogram. For example, increasing the brightness of each element value by the same integer number can be considered. Furthermore, invariant representation is stable with respect to some other distortions such as change in a small percentage of image pixels. It is important that the invariant representation of any image is idempotent. Its use makes possible (in some cases) to increase the stability of the analysis while reducing the requirements for computational resources.

4 Segmentation Algorithms

In a broad sense, segmentation is any division of images into homogeneous regions in a predetermined way [5]. Automatic segmentation is one of the most complex image-processing tasks. It should be noted that the use of traditional segmentation

Table 3 Some approaches to the task of RSD segmentation

Method	Advantages	Disadvantages
Histogram methods	Simple software realization; High-speed calculations	Does not take into account connectivity and square of the segments
Methods based on segments merging	High segmentation quality; accounting connectivity and square of the segments.	Tendency to accumulate errors during iterations of segmentation
Methods based on Markov models	Effective for textural segmentation; eliminate the need for global parameters Kussul' et al. [18]	High computational costs

algorithms for processing of RSD is limited due to the large volume of data to be processed [8]. The most important segmentation algorithms for the analysis of remote-sensing data are listed in Table 3.

SRM [9] is an example of a segmentation algorithm based on the segments merging or one of a segmentation algorithm based on the Mumford–Shah model [10].

5 Algorithms of Classification and Clusterization

Classification is defined as "assignment of the input signal to one of some predefined category (class)" [11]. Combinations of methods for describing informative elements can be used for the construction of feature space. Many different methods have been used for classification of the resulting feature space elements [12]: fuzzy logic [13]; probability theory and mathematical statistics [14]; structural methods and methods associated with the modeling of biological systems (e.g., artificial neural networks [15], genetic algorithms, etc.) A separate area is the algebraic approach for solving the problems of recognition and classification [16]. Very important are also the cluster-analysis methods [17].

5.1 Cluster Analysis

The classification results can be improved by preliminary clustering of the data being analyzed. Cluster analysis is a group of techniques that are used to cluster the data feature space, i.e., to divide the data into groups of similar elements, called "clusters," without the use of the training set. A brief description and comparative analysis of cluster-analysis methods is listed in Table 4 [17].

Cluster-analysis methods also differ in the distance measure used, which determines the degree of closeness between clusters. Two well-known representatives of

Table 4 Comparative analysis of some -analysis methods

Group of methods	Subgroup	Description	Advantages	Disadvantages
Hierarchical	According to the direction of the process of dendrogram construction: agglomerative and divisive methods	Construction of the graph (so-called "dendrogram") that reflects hierarchy obtained as a result of the analysis	Clearness of the obtained results	Typically the methods are effective only for a small amount of data
Not-hierarchical	Maximises the degree of difference between clusters; allocates clusters in areas of feature space, which are characterized by the highest density of points	Exhaustive search of partitioning variants as long as the stopping condition is not met	Less sensitive to incorrect choice of metric and presence of noise in the input data	Need to set the number of clusters and/or criterion to stop iterations of the clustering process

the non-hierarchical cluster-analysis methods, which are often used in the analysis of remote-sensing data, include the following:

- k-means algorithm: This is based on maximization of the difference between the average clusters values.
- ISODATA algorithm: This is similar to the method of k-means but has different procedures of element redistribution between clusters. It also requires significantly larger amounts of settings to set, for instance, a number of output clusters.

6 Results of Computer Experiments

Analysis algorithms are part of the developed RSD-analysis system. Software implementation of a distributed system for analysis of remote-sensing data includes the following components:

- Analysis server: This performs tasks of remote-sensing data processing at all stages of analysis.
- Storage server: This provides storage of incoming and receiving RSD and returns RSD on request.

- Subsystem for receiving remote sensing data: This provides an interface to external sources of remote-sensing data.
- GIS interface: This visualises input data and the results of analysis and also allows the user to manage the processing of remote-sensing data by compiling a sequence of blocks from the palette (so-called scenario), each of which represents a corresponding algorithm. An example of an image-preview (visualization) scenario is presented in Figs. 1 and 2.

Let us consider in detail an image-preview scenario that can be used as a stand alone or as a sub-process for another scenario (using an "activity" block). The scenario consists of the following stages:

- Detecting the presence of meta-information in the input image (wavelength set for each band). If wavelengths are set, use it to construct a color-output image. If not, simply average all image bands.
- Equalize and linear normalize all image bands.
- Store output image in a storage server so it can be addressed (by its unique ID) in every other process that needs it.

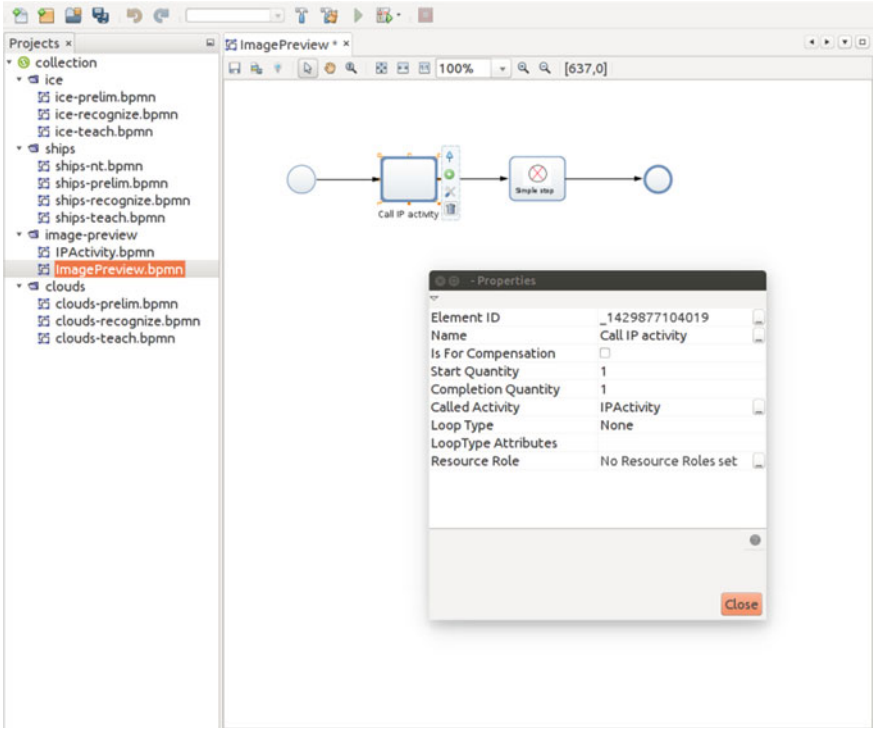


Fig. 1 Example of image-visualization scenario opened in the interface. The content of the "IP activity" cube is presented in Fig. 2

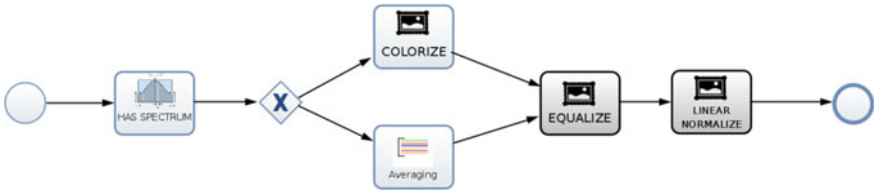


Fig. 2 Main part of image preview. process

Table 5 Parameters of data-analysis scenarios

Scenario name	Input data for test	Data source	Output data	Calculation time	Analysis speed (pixels/s)
Cloud selection	Low-resolution RSD, three bands	GeoEye	Vectorized cloud mask with georeference	1.5 min for 7.5-megapixel image	83,000
Ship recognition	Average-resolution RSD, five bands	GeoEye	Ship coordinates (course and detection probability)	2 min for a 10-megapixel image	83,000
Image visualization	Arbitrary-resolution RSD with or without georeference	GeoEye	Color or gray-scale image	1 min for a 10-megapixel image	166,000
Ice detection	Low-resolution RSD, three bands	MODIS	Vectorized ice mask with georefernce	5 min for a 50-megapixel image	166,000

The developed system also includes a number of auxiliary algorithms (vectorization of natural objects, calculation of parameters of found objects, etc.) totalling >70 algorithms. Examples of scenarios for different tasks solved by the system are listed in Table 5. As a server, Intel Corei7-3770 CPU @ 3.40 GHz, 32 Gb RAM was used.

As an example of the task of artificial-object selection, let us consider the ship-recognition task using multispectral RSD (five spectral bands including the near-infrared and panchromatic band of GeoEye satellite) at 3552*2929 pixels each. Analysis time is approximately 5 min. The system also solves important auxiliary problem such as calculation and visualization of a ship's course based on wake analysis (Fig. 3).

The above-described methods of analysis allow to solve a wide range of different tasks and take into account various object features. However, expert work for creating the combination of algorithms useful for the current task is required.

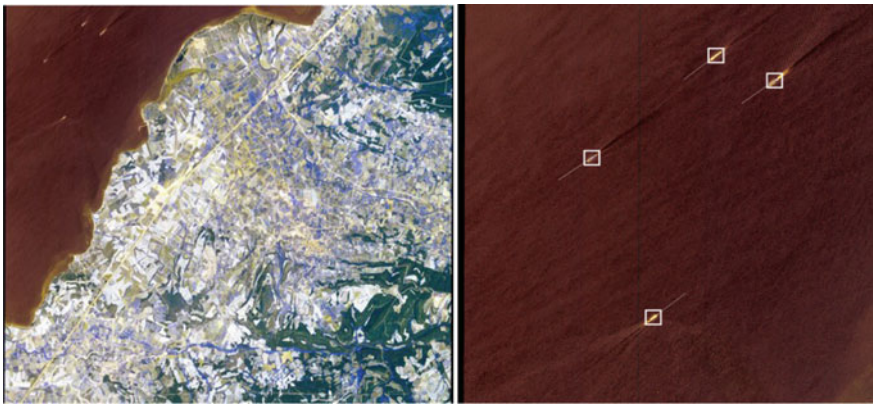


Fig. 3 Example of RSD analysis for ship selection. (*Left panel*) Visualization of input data. (*Right panel*) Results of the analysis (ships and courses). Initial data provided by SCANEX

7 Conclusion

The highlight of the presented technology is not only its self-retention, but also openness, which is realized on both the software and algorithmic levels.

At the software level, any algorithm can be implemented in Java and used as module for the system, thus increasing its capabilities without need of modifying other modules. In addition, the system can (potentially) interact with any third-party image-analysis or numerical-computing software that supports working in batch mode (e.g., Matlab) by using its capabilities to perform some of the steps of analysis (and vice versa). The cost of such interaction depends only on restrictions imposed by the external software (according to its API and the set of technologies used in it).

At the algorithmic level, the system proposes a set of data structures and common but flexible interfaces and process structures that can be used to describe different algorithms employed for RSD analysis. Methods of solving classification and segmentation tasks are an important part of a set of tools that provide advanced space technologies.

The increase of available computing power allows to extend a set of problems to be solved, but obtaining stability of the classification results based on the conditions under which the images are made still does not solve the problem. The need to take into account the different properties of the desired objects does not allows the use of a limited algorithm set. However, a uniform representation of the object's properties in the form of feature vectors almost always allows to consider the whole set of relevant attributes for the classification. In general, the current level of development allows to create a systems that greatly facilitates and accelerates the work of an RSD-analysis expert, but the creation of an automated system of RSD analysis that combines versatility with performance is beyond our current capabilities.

References

1. Tarjan RE (1983) A data structure for dynamic trees. J Comput Syst Sci 26(3):362–391
2. Sleator DD, Tarjan RE (1985) Self-adjusting binary search trees. J ACM 32:652–686
3. Kharinov M (2006) Zapominanie i adaptivnaya obrabotka informatsii tsifrovykh izobrazhenii (Styoring and adaptive information processing of digital images). Izdatel'stvo Sankt-Peterburgskogo universiteta, Saint-Petersburg, p 138
4. Potapov A (2007) Raspoznavanie obrazov i mashinnoe vospriyatie: obshchii podkhod na osnove printsipa minimal'noi dlinny opisaniya (Pattern recognition and machive perception: a general approach based on minimum description length principle). Politekhnika, Saint-Petersburg, p 548
5. Gonzalez R, Woods R (2006) Tsifrovaya obrabotka izobrazhenii (Digital image processing). Tekhnosfera, Moscow, p 616
6. Kharinov MV (2008) Adaptivnoe vstraivanie vodyanykh znakov po neskol'kim kanalam (Adaptive embedding of watermarks by multiple channels): Patent 2329522 RF, No. 20, 41 p
7. Kharinov MV, Galyano FR (2009) Raspoznavanie izobrazhenii posredstvom predstavlenii v razlichnom chisle gradatsii (Images recognition by means of its representation in different number of gradations). In: Proceedings of 14-th All-Russia Conference, "Matematicheskie metody raspoznavaniya obrazov" (Mathematical Methods of Pattern Recognition). MAKS Press, Moscow, pp 465–468
8. Aksoy S, Chen CH (2006) Spatial techniques for image classification, signal and image processing for remote sensing. Taylor & Francis, Boca Raton, pp 491–513
9. Nock R, Nielsen F (2004) Statistical region merging. IEEE Trans Pattern Anal Machine Intell 11:1–7
10. Robinson DJ, Redding NJ, Crisp DJ (2002) Implementation of a fast algorithm for segmenting SAR imagery. DSTO Electronics and Surveillance Research Laboratory, Edinburgh, p 41
11. Luger, G. (2005) Iskusstvennyi intellekt: Strategii i metody resheniya slozhnykh problem (Artificial Intelligence: Structures and Strategies for Complex Problem Solving). Izdatel'skii dom "Vil'yams", Moscow, 864 p
12. Galjano Ph, Popovich V (2007) Intelligent images analysis in GIS. In: Proceedings of IF&GIS-2007. Springer, Berlin, 325 p
13. Etienne EK, Nachtegael M (2000) Fuzzy techniques in image processing. Springer, New York, p 413
14. Storvik G, Fjortoft R, Solberg AHS (2005) A Bayesian approach to classification of multiresolution remote sensing data. IEEE Trans Geosci Remote Sens 43:539–547
15. Haykin S (2008) Neironnye seti: polnyi kurs (Neural networks: a comprehensive foundation). Vil'yams, Moscow, p 1104
16. Yu Z (1978) I. Ob algebraicheskom podkhode k resheniyu zadach raspoznavaniya i klassifikatsii (About the algebraical approach to solve the recognition and classification tasks), Problemy kibernetiki, No. 33, pp 5–68
17. Chubukova I (2008) Data mining (Data mining). Binom, Moscow, p 384
18. Kussul' NN, Shelestov AY, Skakun SV, Kravchenko AN (2007) Intellektual'nye vychisleniya v zadachakh obrabotki dannykh nablyudeniya Zemli (Intellingent computing in tasks on processing Earth remote sensing data). Naukova Dumka, Kiev, p 196

Modeling of Surveillance Zones for Bi-static and Multi-static Active Sonars with the Use of Geographic Information Systems

Vladimir Malyj

Abstract This chapter regards the main features of an efficiency evaluation of bi-static and multi-static active sonar systems on the basis of modeling and visualization of the expected surveillance zones. Assessment of the impact of information-support quality of geographic information systems on the accuracy of calculation of the expected surveillance zones for bi-static and multi-static active sonars under various hydrological and acoustic conditions for various models of the inhomogeneous marine medium is performed.

Keywords Bi-static and multi-static active sonar systems · Surveillance zone · Hydrological and acoustic conditions · Geographic information systems · Inhomogeneous marine medium

1 Introduction

Geographic information systems (GIS) are the most important component of modern underwater surveillance systems. At the same time, the most fast-growing types of sonar, as a part of information acquisition subsystems for modern underwater surveillance systems, are bi-static and multi-static active sonar systems [1–6].

Efficiency-evaluation techniques of a given type of active sonar system on the basis of modeling and visualization of the expected surveillance zone, as well as the requirements of GIS that provide informative realisation [7–9], have essential features and differences compared with traditional efficiency-evaluation techniques of mono-static active sonar systems.

The purpose of this work is (1) to further develop a technique of efficiency evaluation of active sonar systems (for bi-static and multi-static operation modes) under conditions of homogeneous marine medium on the basis of modeling and visualization of the expected surveillance zone; and (2) to conduct an impact

V. Malyj (✉)
SPIIRAS-HTR&DO Ltd, 14 Line, 39, 199178 St. Petersburg, Russia
e-mail: maliy@oogis.ru

V. Popovich et al. (eds.), *Information Fusion and Intelligent Geographic Information Systems (IF&IGIS'17)*, Lecture Notes in Geoinformation and Cartography, DOI 10.1007/978-3-319-59539-9_11

assessment of quality-information support of GIS based on the accuracy of expected surveillance zone–calculation results under various hydrological and acoustic conditions (HAC).

We provide surveillance zone–calculation results for bi-static and multi-static active sonar systems for theoretical conditions in the infinite homogeneous medium (IHM) as well as the expected surveillance zone—calculation results in case of the main types HAC for various options of medium model: For a traditional idealized layered-inhomogeneous model for the marine medium with a flat bottom and for more difficult two-dimensional inhomogeneous model for the marine medium with arbitrary bottom relief as well as for variable vertical sound propagation velocity (VSPV).

2 Modeling, Calculation, and Construction of Surveillance Zones for Bi-static and Multi-static Active Sonars for Conditions of Infinite Homogeneous Medium

As shown in work [10], external borders of the surveillance zone of a bi-static active sonar system for IHM conditions—at the fixed probability of correct detection (PCD) P_{cd} and at placement of the origin of polar coordinates in the center of an *SR* segment that connects locations of the source and the receiver (Fig. 1)—can be depicted by Eq. (1) and represent curves similar in shape to Cassini ovals [11, 12] as follows:

$$10^{0.1\beta\left(\sqrt{\rho^2+\left(\frac{b}{2}\right)^2+\rho b\cos\gamma}+\sqrt{\rho^2+\left(\frac{b}{2}\right)^2-\rho b\cos\gamma}\right)}\cdot\left(\rho^4+\left(\frac{b}{2}\right)^4-2\rho^2\left(\frac{b}{2}\right)^2\cos 2\gamma\right)=10^{0.2\beta D_0}D_0^4 \tag{1}$$

where b is a distance between the source and the receiver; ρ is a distance from the origin of the coordinates to the target; γ is an angle between the source and the

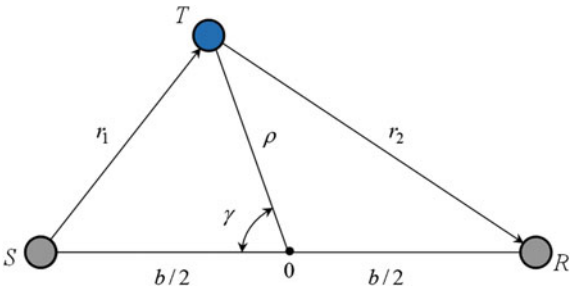


Fig. 1 Target position (*T*) in relation to the source (*S*) and the receiver (*R*) of a bi-static active sonar system (the origin of polar coordinates is in the midpoint of the *SR* that connects the locations of the source and the receiver)

target directions; β is a frequency-dependent coefficient of space attenuation; and D_0 is an active sonar range in IHM during its operation in the monostatic mode (it calculates for a case when all of its technical characteristics coincide with similar characteristics of a bi-static active sonar and as a result it has the same value of noise level and target strength).

The form of these curves, first of all, is defined by the ratio b/D_0 [10]. Figure 2 illustrates examples of the calculation of surveillance-zone external borders for a bi-static active sonar system obtained according to Eq. (1) for the fixed distance b between the source and the receiver and at various D_0 values. Their form can be caused, for example, by a different value of target strength, noise level, radiated sonar power, etc.

The given surveillance-zone borders correspond to different values of ratio b/D_0 = 2.0, 1.6, 1.33, 1.143, 1.0, 0.889, 0.8, 0.727, and 0.667. In Fig. 2, black points designate the fixed locations of the source and the receiver for a bi-static active sonar system. Figure 3 conveys the results of mathematical modeling of the surveillance zone for a bi-static active sonar system obtained according to Eq. (1) at the fixed value of IHM sonar range in the monostatic mode D_0. It corresponds to the same value of target strength, noise level, technical characteristics of active sonar systems source and receiver but for different value of distance b between the source

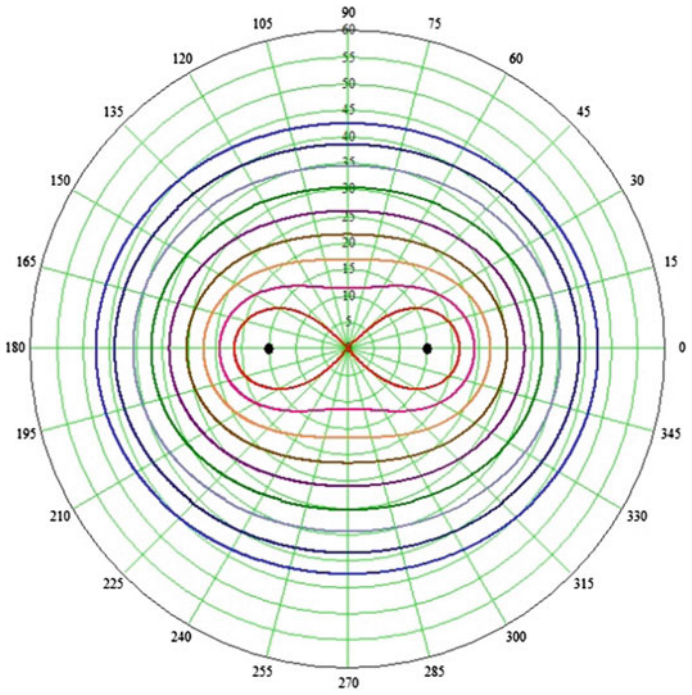


Fig. 2 Example of external borders of surveillance zones for a bi-static active sonar systems with the fixed distance b but for a different value of IHM sonar range in the monostatic mode D_0

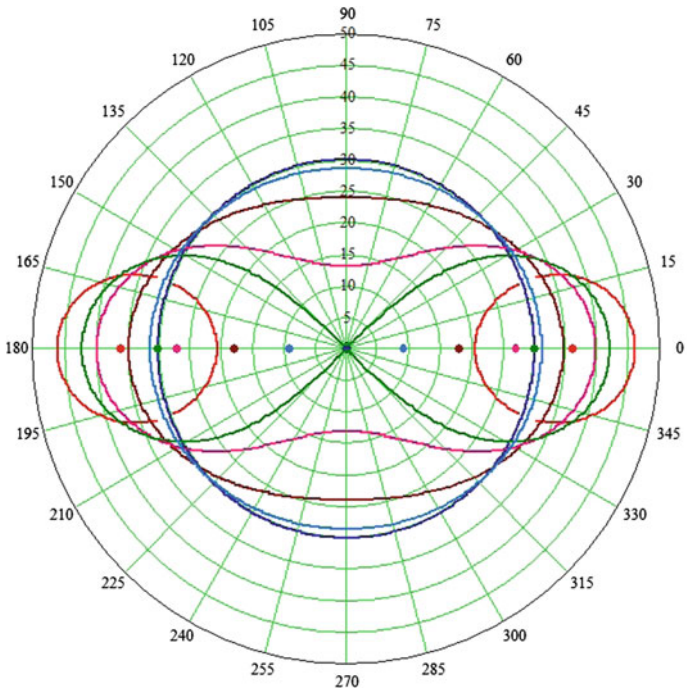


Fig. 3 Example of external borders of the surveillance zone for a bi-static active sonar system with fixed value D_0 in the monostatic mode but for different value of distance b between the source and the receiver

and the receiver. Pairs of symmetrical points denote the locations of source and receiver for bi-static active sonar systems and match with different values of distance b. The locations of the source and the receiver, as well as the corresponding external borders, are painted the same colour. External borders of the formed surveillance zone correspond to different values of ratio b/D_0 = 0, 0.6, 1.2, 1.8, 2.0, and 2.4.

As you can see from Fig. 3, under condition $b = 2D_0\,(b/D_0 = 2)$, the curve describing the surveillance-zone borders transforms into lemniscate of Bernoulli [11, 12]; with a further increase of $b > 2D_0\,(2 < (b/D_0) < \infty)$, it splits into two separate ovals. Each of them extends in the direction of the other.

The internal borders of the surveillance zone for a bi-static active sonar system are defined by the so-called "dead zone," which has the form of ellipses that extend along the line that connects the receiver and the source [1, 10]. The major axis of this ellipse is defined by the following expression:

$$l = b + c\tau \tag{2}$$

The minor axis is defined by the following expression:

$$h = \sqrt{l^2 - b^2} = \sqrt{2bc\tau + (c\tau)^2} \tag{3}$$

where c is the speed of sound in water; and τ is the radiation-pulse duration.

The dead-zone ellipse equation for a bi-static active sonar system in polar coordinates with the origin in the segment center jointing points of the receiver and the source (Fig. 1) is defined by the following expression:

$$\rho(\gamma) = \frac{l/2}{\sqrt{1 - \mu^2 \cos^2(\gamma)}} \tag{4}$$

where $\mu = \frac{\sqrt{(l/2)^2 - (h/2)^2}}{l/2}$ is an eccentricity of the dead-zone ellipse.

An example of ellipses for a "dead zone", with the fixed impulse duration, but for different value of distance b between the source and the receiver, is given in Fig. 4.

However, in modern programs of the efficiency evaluation of active sonar systems, usually the surveillance zone is visualized as "continuous" two-dimensional function of spatial coordinates $P_{cd}(r_1,\alpha) = P_{cd}(q(r_1,\alpha))$ but not as separate borders

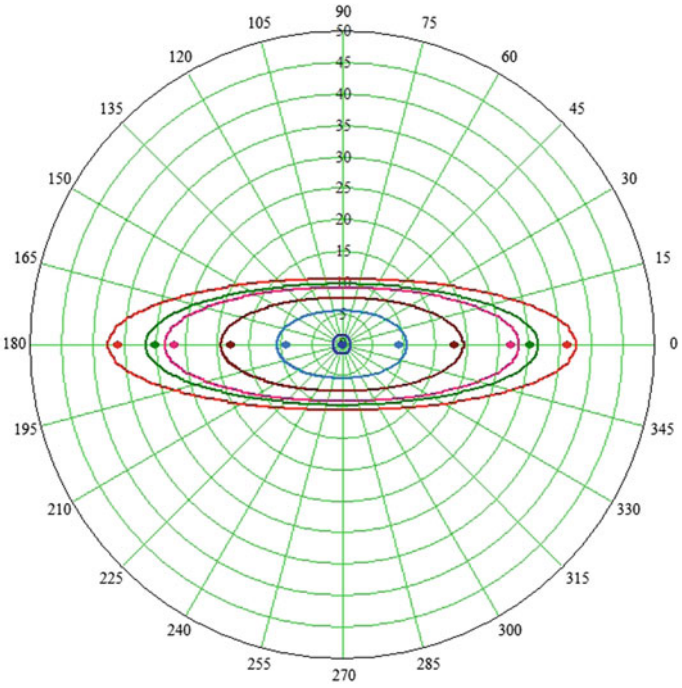


Fig. 4 Example of dead-zone ellipses for a bi-static active sonar system with fixed impulse duration τ for different values of distance b between the source and the receiver

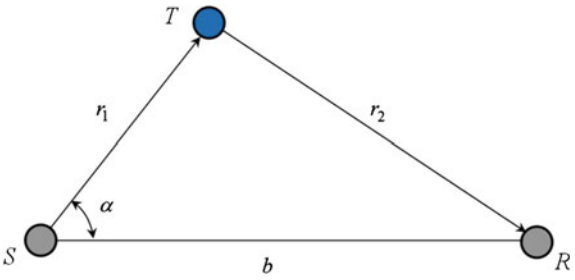


Fig. 5 Target position (*T*) in relation to the source (*S*) and the receiver (*R*) of a bi-static active sonar system (in polar coordinates with the center in the point of the source location)

of the surveillance zone (e.g., for the fixed PCD value of $P_{cd} = 0.9$). The function is calculated from the receiver operating characteristic by substitution of the corresponding dependence of the input signal-to-noise ratio (SNR) from the spatial coordinates of the target $q(r_1,\alpha)$. In this case, it would be convenient to use a polar-coordinates system with a center in the point of the source location (Fig. 5).

For visualization of $P_{cd}(r_1,\alpha)$, we usually use a colour scale with a smooth transition from violet ($P_{cd} = 0$) to red ($P_{cd} = 1$). This scale is convenient for the representation of three-dimensional graphics in the form of flat projections.

In Fig. 6, we can see an example of the surveillance zone–calculation results for similar bi-static active sonar systems under conditions of IHM for different distance b at the fixed D_0, with a given visualization option of the surveillance zone for arbitrary PCD $0 \le P_{cd}(r_1, \alpha) \le 1$.

For efficiency evaluation of a more complicated multi-static active sonar system that includes more than one ($N \ge 2$) receiver in the structure, at the first stage of modeling we construct a separate surveillance zone $P_{cd_n}(r_1, \alpha)$ for each receiver in relation to the general source; then, at the second stage, we calculate the general PCD using a formula for determining the detection probability of the multichannel receiver as follows:

$$P_{cd_\Sigma}(r_1, \alpha) = 1 - \prod_{n=1}^{N} (1 - P_{cd_n}(r_1, \alpha)) \tag{5}$$

Then the total surveillance zone of the multi-static active sonar system is under constructed.

In Fig. 7, examples of the calculation results and the surveillance-zone visualization under IHM conditions for a multi-static active sonar system are given follows:

(a) with two identical receivers ($N = 2$) dispersed in opposite directions from the emitter on equal distances ($b_1 = b_2$ and $b_1/D_0 = b_2/D_0 = 1.2$); and
(b) with three identical receivers ($N = 3$) dispersed in different directions (through 120°) from emitter on equal distances ($b_1 = b_2 = b_3$ and $b_1/D_0 = b_2/D_0 = b_3/D_0 = 1.2$).

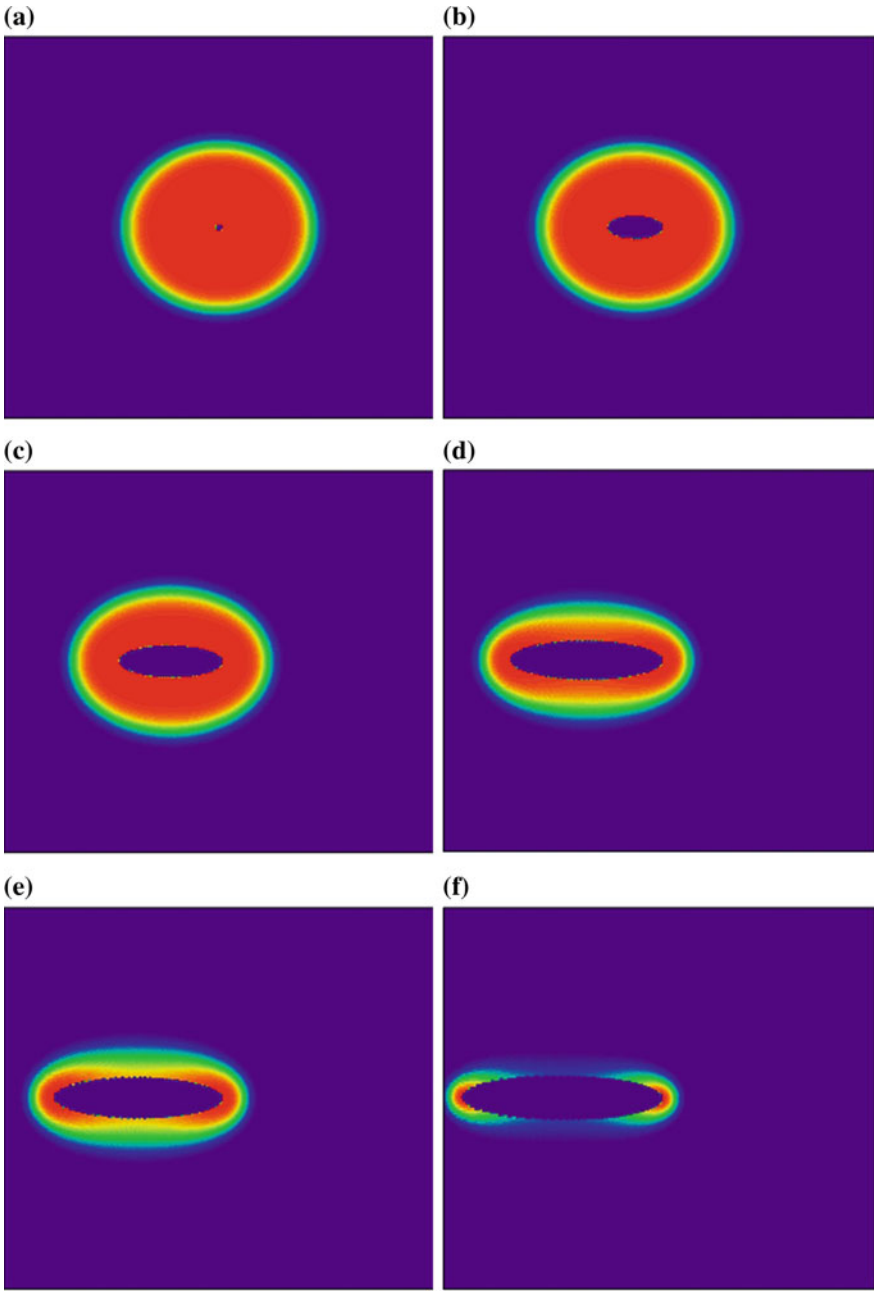


Fig. 6 Examples of surveillance zones for monostatic (**a** [$b/D_0 = 0$]) and bi-static active sonar systems (**b** through **f** [$b/D_0 = 0.6, 1.2, 1.8, 2.0$, and 2.4]) under HM conditions with fixed D_0 in the monostatic mode for different distance b

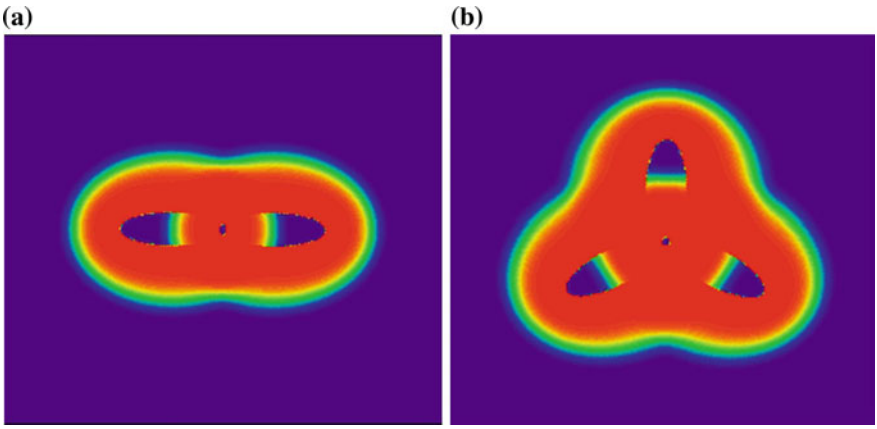


Fig. 7 **a** and **b** Examples of surveillance zones for multi-static active sonar systems with two and three receivers under IHM conditions

In fact, surveillance-zone calculation and analysis for bi-static and multi-static active sonar systems under IHM condition have theoretical nature, and their results can be used only for educational purposes or at the stage of preliminary design of an active sonar system. Techniques and algorithms of expected surveillance-zone calculation and visualization (creation) are of practical interest (for the evaluation of active sonar-systems efficiency under real medium conditions) for more difficult medium models in specific HAC for given depths of the antennas and the target location. For this purpose, at least, it is necessary to know the depth of the sea and the VSPV around the location of the active sonar system's antennas, which can be obtained from the corresponding databases or from the corresponding element of operational oceanology system (OOS).

3 Modeling, Calculation, and Construction of the Expected Surveillance Zones for Bi-static and Multi-static Active Sonar Systems for a Layered-Inhomogeneous Model of the Marine Medium

Figure 9 illustrates results of modeling and calculations for a bi-static active sonar system with similar characteristics and for the same distances b between the source and the receiver (as well as in Fig. 6) but for the concrete HAC corresponding to Barents Sea conditions during winter, i.e., positive refraction from the surface to the bottom (PRSB), sea depth of 200 m, and target and active sonar-system antennas

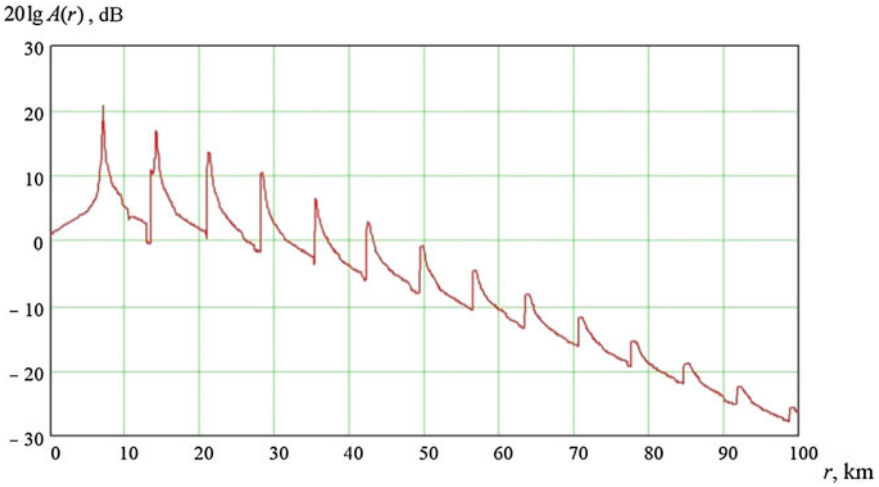


Fig. 8 Sound propagation anomaly level with distance dependence for PRSB conditions and sea depth of 200 m

located at a depth of 100 m. Calculation results of propagation anomaly *versus* distance response are given in Fig. 8.

In the simplified modeling option (using classical model of layered-inhomogeneous marine medium with fixed VSPV and flat bottom), anomaly dependence [with sound propagation from the source to the target of $A_1 = A(r_1)$ and from the target to the receiver of $A_2 = A(r_2)$] is determined only by the distance passed by the signal and does not depend on the concrete directions and trajectories. Consequently, the pressure of the received echo signal–dependence on the target coordinates $P_s(r_1, \alpha)$ can be written as follows:

$$P_s(r_1, \alpha) = \frac{P_1}{r_1 \cdot r_2(r_1, \alpha)} \sqrt{A(r_1)} \sqrt{A(r_2(r_1, \alpha))} \frac{R_e}{2} 10^{-0.05\beta(r_1 + r_2(r_1, \alpha))} \quad (6)$$

where P_1 is a reduced pressure on an axis of the radiating directivity pattern; and R_e is the equivalent sphere radius of the sonar target.

The example of the expected surveillance-zone calculation and visualization results for a multi-static active sonar system with two ($N = 2$, $b_1 = b_2$) and three ($N = 3$, $b_1 = b_2 = b_3$) identical receivers, under PRSB conditions, is given in Fig. 10.

It is pertinent to note that, as can be seen from Figs. 7 and 10, total dead-zone ellipses' areas, always typical of the bi-static mode, can be considerably reduced by rationally locating the receiving antennas in relation to the emitter under a multi-static operating mode of an active sonar system.

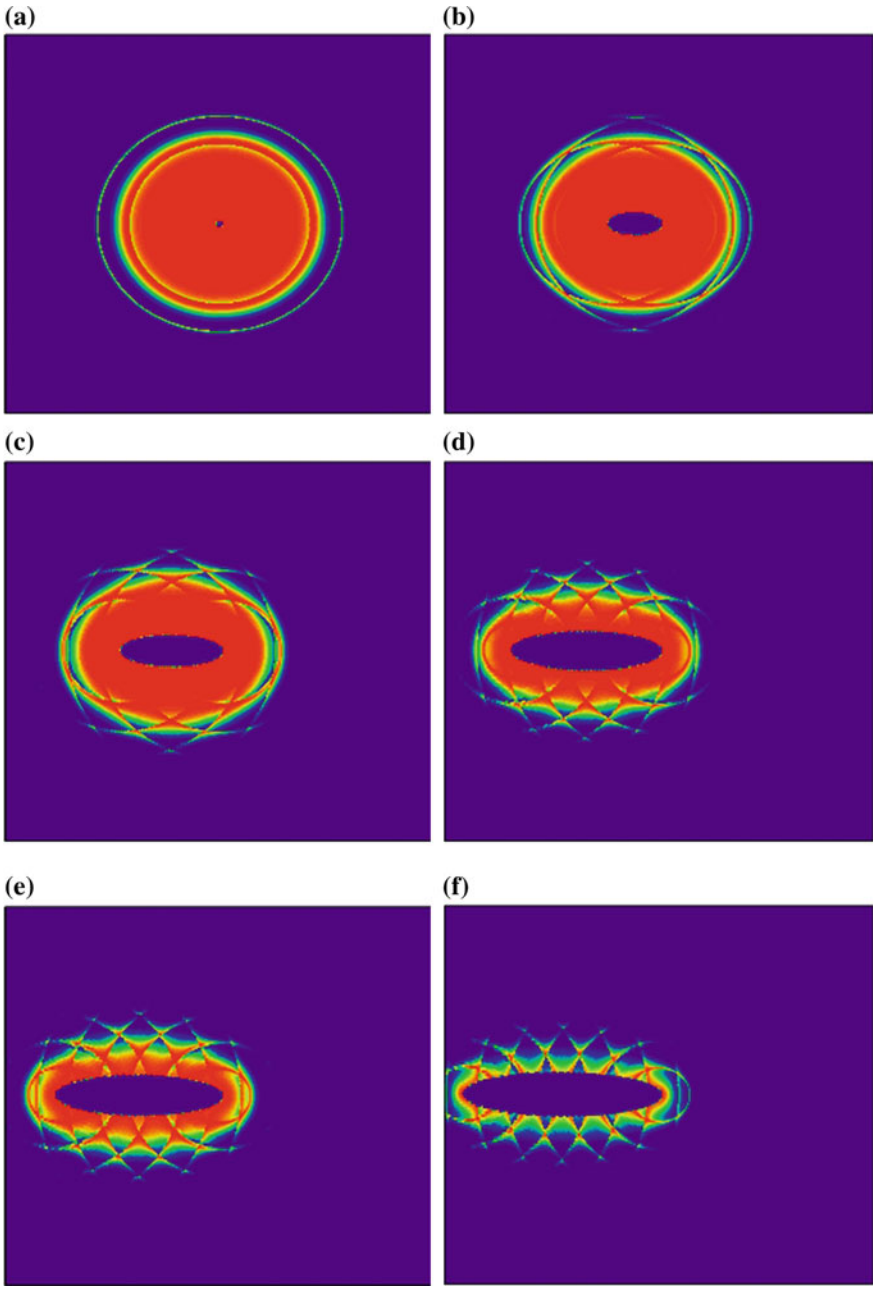


Fig. 9 Expected surveillance zones for a mono-static active sonar system (**a**) and for a bi-static active sonar system (**b** through **f**) under PRSB conditions for different distance b between the source and the receiver

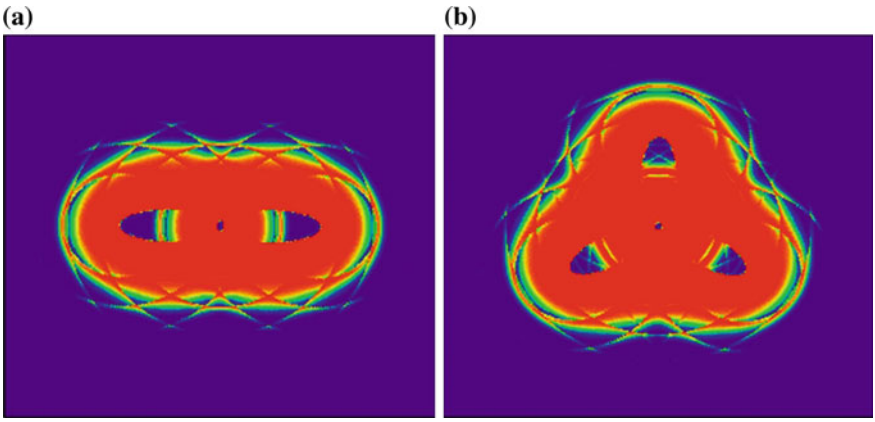


Fig. 10 Example of the expected surveillance zone for a multi-static active sonar system with two (**a**) and three (**b**) receivers under PRSB conditions

4 Modeling and Construction Features of the Expected Surveillance Zones for Bi-static Active Sonar Systems for a Two-Dimensional Inhomogeneous Model of Marine Medium Conditions

A traditional approach with the application of the simplified classical model of layered-inhomogeneous marine medium with fixed VSPV and a flat bottom gives us an idealized symmetric surveillance zone for a bi-static active sonar system: However "under real sea conditions, i.e., with a complicated bottom relief and a changing VSPV profile on the signal-propagation path, this leads to considerable errors. It is especially evident for low-frequency bi-static active sonar systems with a large range of action. Therefore, it is obvious that strict binding of the calculated surveillance zone to a map of the observation region (with a real bottom relief and changing VSPV profile), as well as coordinates of the emitting and receiving antennas" installation, is necessary for real efficiency evaluation of a given active sonar system. This approach can be realized only by means of specialized GIS [7–9] with the appropriate data bases (bottom relief and hydrological conditions) or by means of the hydrological data arriving to the GIS from the OOS.

An example of expected surveillance-zone calculation and creation results for low-frequency bi-static active sonar systems under Norwegian Sea deep-water region conditions during winter (the axis of the underwater sound channel is located at a depth of 600 m)—along with the application of a complicated (and closest to real conditions) model of two-dimensional inhomogeneous marine medium with VSPV variables on the path and a bottom relief—is given in Fig. 11.

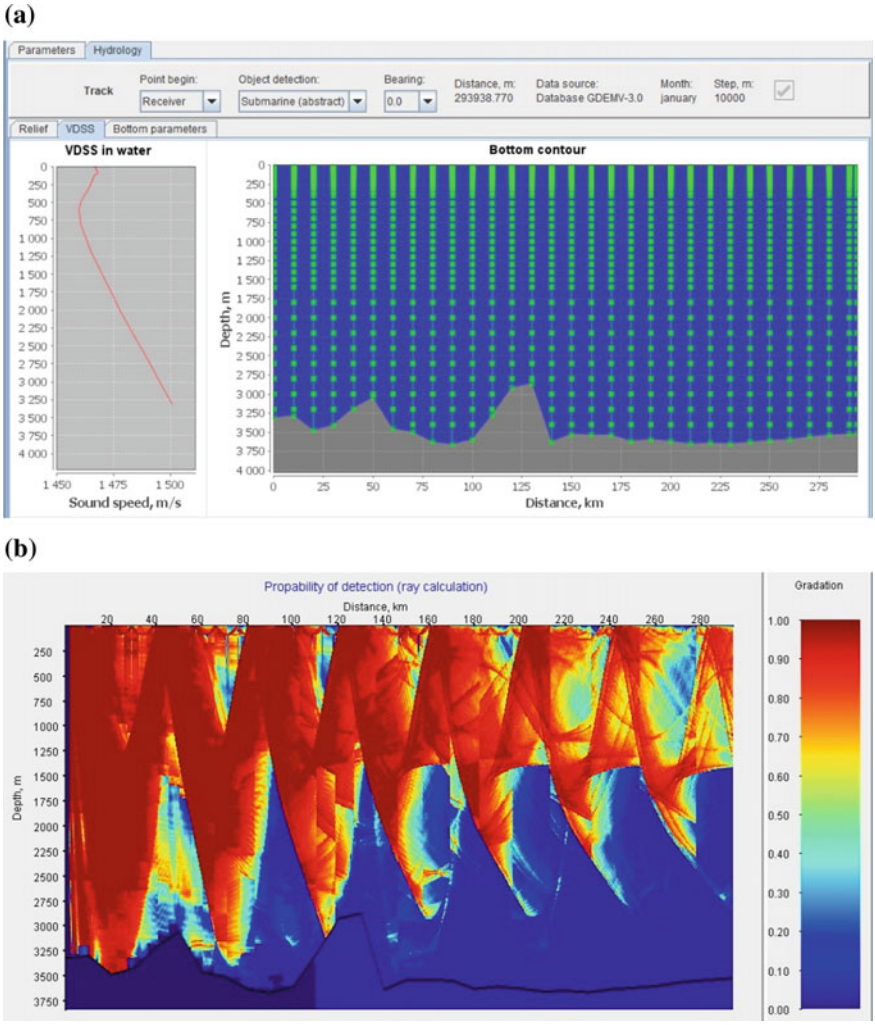


Fig. 11 Expected surveillance zone–calculation results for a bi-static active sonar system with the application of GIS-information support: **a** bottom contour and variable VSPV on the path for bearing 0° from the receiver; **b** a vertical profile of surveillance zone in the direction 0° from the receiver, for arbitrary value PCD $0 \leq P_{cd}(r, \alpha) \leq 1$; **c** example of the expected surveillance zone on a horizon of 80 m for a fixed value of a threshold of PCD $0.9 \leq P_{cd}(r, \alpha) \leq 1$ (with a binding to the map); **d** example of the expected surveillance zone in polar coordinates on a horizon of 80 m for the arbitrary value of PCD $0 \leq P_{cd}(r, \alpha) \leq 1$

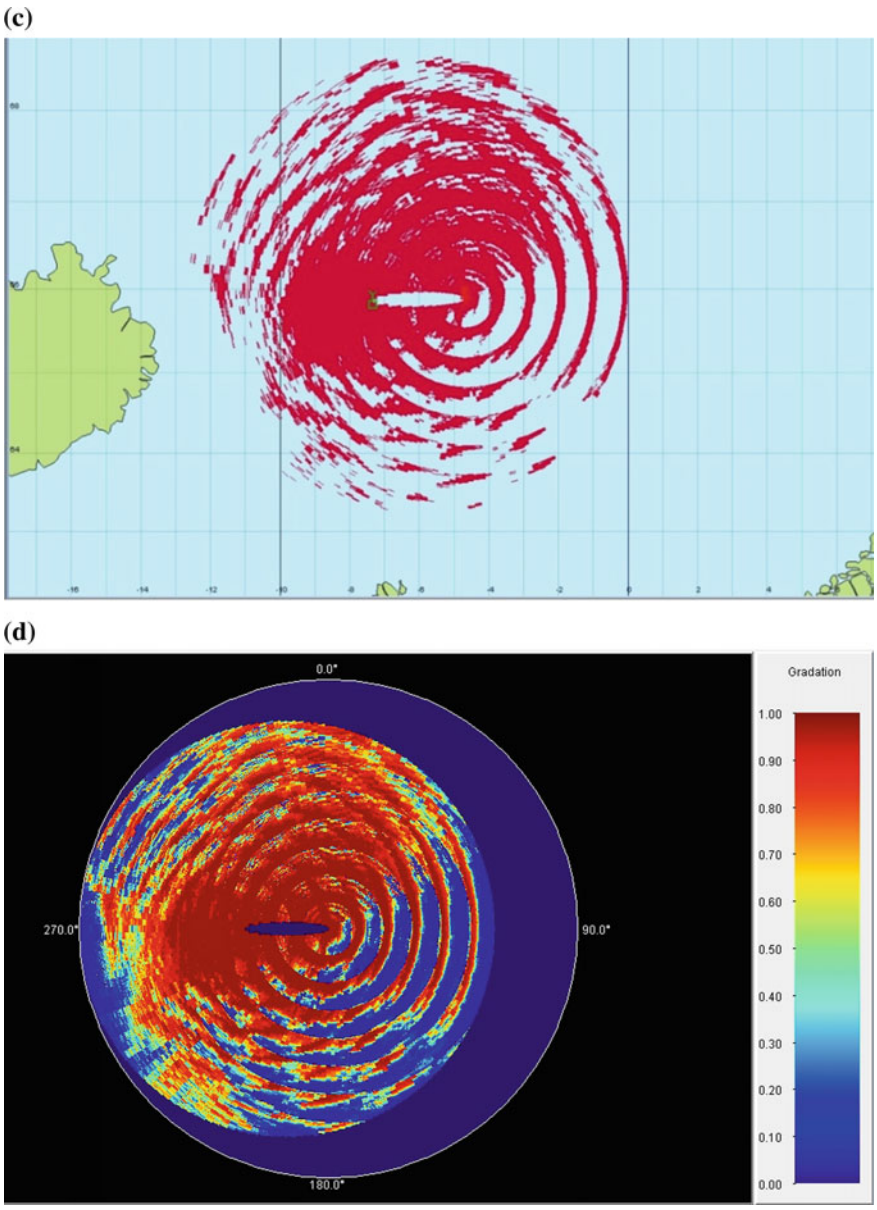


Fig. 11 (continued)

5 Conclusion

Strict binding of the calculated expected surveillance zone to the observation region on the map with a real bottom contour, to coordinates of the emitting and receiving antennas' installation, is necessary for a real efficiency evaluation of bi-static and multi-static active sonar systems, especially for low-frequency active sonar systems with a large range of action. Such an approach can be realized only by means of specialized GIS with the corresponding databases (a bottom contour and hydrology) or by means of the operational hydrological data arriving in the GIS from the OOS [7–9].

References

1. Cox H (1989) Fundamentals of bistatic active sonar. In: Chan YT (eds) Underwater acoustic data processing. Kluwer Academic Publishers. NATO ASI Series, vol 161, pp 3–24
2. Naluai NK, Lauchle G, Gabrielson JJ (2007) Bi-static applications of intensity processing. J Acoust Soc Am 121(4):1909–1915
3. Edwards JR, Schmidt H, LePage K (2001) Bistatic synthetic aperture target detection and imaging with an AUV. IEEE J Ocean Eng 26(4):690–699
4. Bowen JI, Mitnick RW (1999) A multistatic performance prediction methodology. Johns Hopkins APL Tech Digest 2(3):424–431
5. Belous YuV, Kozlovskiy SV, Sergeyev VA (2006) A polystatic method of a location in relation to mobile carriers (in Russian). In: Works VIII of the international conference "Applied technologies of hydroacoustics and hydrophysics." Science, St. Petersburg, pp 81–84
6. Mashoshin AI (2013) Algorithm of optimum positioning of multistatic system of a hydrolocation (in Russian). Sea radio electronics, No. 4 (46), pp 14–18
7. Popovich VV, Yermolaev VI, Leontyev YuB, Smirnova OV (2009) Modeling of hydroacoustic fields on the basis of an intellectual geographic information system (in Russian). Artificial Intelligence and Decision Making 4:37–44
8. Yermolaev VI, Kozlovskiy SV, Makshanov AV (2009) IGIS controlling polystatic detection security marine economic activity (in Russian). In: Popovich VV, Schrenk M, Claramunt C, Korolenko KV (eds) Proceedings of 4th International IF&GIS Workshop. Springer, St. Petersburg, pp 265–276
9. Guchek VI, Yermolaev VI, Popovich VV (2012) Systems of monitoring on the basis of IGIS (in Russian). Defensive Order 2(21):58–61
10. Malyj VV, Leontyev YuB (2016) Features of assessment of a zone of observation of bistatic sonar system for conditions of a boundless homogeneous environment (in Russian). International Scientific Institute "Educatio," vol 1, no 19, part 2. pp 38–44
11. The Soviet Encyclopedia (1982) The mathematical encyclopedia (in Russian) (in 5 volumes). The Soviet Encyclopedia, Moscow, vol. 2
12. Bronshtein IN, Semendyaev KA (1986) The reference book on mathematics for engineers and pupils of technical colleges (in Russian). Science, Moscow

Geoinformational Support of Search-Efforts Distribution Under Changing Environmental Conditions

Victor Ermolaev and Sergey Potapichev

Abstract Modern measuring and control instruments allow to describe the environmental properties that influence the search efforts using monitoring tools. These tools apply for the search of alarmed objects in the interest of navigation security and detecting objects of maritime activities. The scientific literature describes only the problem of the search-effort distribution when the observer' possibilities do not depend on their location and direction of the search object. The interesting case is when changing environmental conditions define unique conditions of object detection at the each point of the search area. In this case, all objects of the search operation are represented as geospatial objects, and the search effort distribution is provided by GIS support. This chapter discusses the following areas of GIS support: attributes of geospatial-object generation; manipulation of geospatial data for problem solving of search-effort distribution; and graphical interpretation of initial, intermediate, and final data of problem solving.

Keywords Search operation · Search effort · Environment · Geospatial data

1 Introduction

A search-operation model includes the following classes of objects: the observer (search), the object of search (target), and the area of search (environment) [1]. The optimum allocation of search-effort problems has been discussed in sufficient detail by several authors. As a rule, they use a homogeneous [1–4], or a layered-heterogeneous [5], environment model in the search area. Let us consider the two-dimensional heterogeneous environment model, which is the most close to real conditions.

V. Ermolaev (✉) · S. Potapichev
St. Petersburg Institute for Informatics and Automation of RAS 39, 14 Liniya,
St. Petersburg 199178, Russia
e-mail: ermolaev@oogis.ru

V. Popovich et al. (eds.), *Information Fusion and Intelligent Geographic Information Systems (IF&IGIS'17)*, Lecture Notes in Geoinformation and Cartography, DOI 10.1007/978-3-319-59539-9_12

The problem is formulated as follows. Suppose the observer performs a search in the confined space $V_0 = L \times B \times H$. The speed of the observer is v_s. The search time is T_p.

The probability-density distribution of the search object is denoted by $w(X)$ where $X = (x, y, z)$. The speed of the search object is v_m. The search capability of the observer is characterized by its search efforts: $\Phi(X)$. The space volume determines the search efforts, which are explored by the observer per unit of time. The search conditions are characterized by two-dimensional heterogeneous environment model $E(X)$.

The problem is to determine such a search strategy $L(X)$ that maximizes the probability of target detection $P_0(t)$ with a minimum of mathematical expectation of the detection time $\bar{T}$.

2 Solution to the Problem of Search Efforts–Distribution

Two-dimensional heterogeneous environment model $E(X)$ determines non-stationary target detection flow by moving the observer. In this case, the probability of detecting a search object during the search time t is determined as follows [1, 4]:

$$P(t) = 1 - \exp\left[-\int_{t_0}^{t_0+t} \lambda(t)dt\right], \; t > 0 \tag{1}$$

where $\lambda(t)$ is the intensity of target detection by observer.

The value $\lambda(t)$ characterizes the expectation of the number of targets detected by the observer per unit time. The detection intensity by the observer depends on their search capabilities at the point $X_t = (x_t, y_t, z_t)$ and the density distribution of the target. Let $\Phi(X_t) = \Phi_t$, then

$$\lambda(t) = \int_{\Phi_t} w(X)d(X) \tag{2}$$

$$\int_{\Phi_t} w(X)d(X) = \iiint_{\Phi_t} w(x, y, z)dxdydz$$

Search efforts of the observer at the point X_t is defined as follows:

$$\Phi_t = S_{ef}(X_t) \cdot \dot{v}_r \tag{3}$$

where $S_{ef}(X_t)$ is the effective area, which is inspected by an observer at the point X_t; and $\dot{v}_r$ is the mean relative speed of search:

$$\dot{v}_r = \frac{2}{\pi}(v_m + v_s)E\left(k, \frac{\pi}{2}\right)$$

Figure 1 illustrates the approach to the calculation of the observer's search effort at point X_t. The target-detection probability calculations in a two-dimensional environment are performed in accordance with the works [6, 7].

$$S_{ef}(X_t) = \int_{Z_{\min}}^{Z_{\max}} L_{X_t}(z)dz \tag{4}$$

The range of integration $Z_{\min}, Z_{\max}$ defines the range of possible target locations along the Z-axis. The length of the segment of effective detection $L_{X_t}(z)$ is the line segment in which we perform effective target detection for fixed coordinate value Z.

$$L_{X_t}(z) = \frac{1}{\Delta K_r} \int_{K_{r_{\min}}}^{K_{r_{\max}}} L'_{X_t}(K_r)dK_r \tag{5}$$

where ΔK_r is variation range of relative courses:

$$\Delta K_r = K_{r_{\max}} - K_{r_{\min}}$$

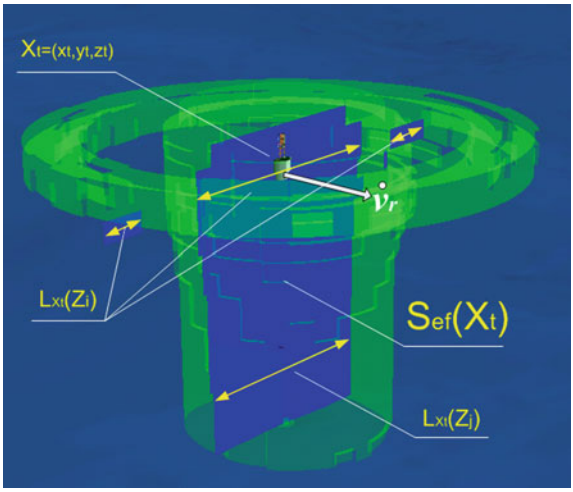


Fig. 1 The approach to the calculation of the observer's search effort

$L'_{X_t}(K_r)$ is the length of the segment of effective detection at the fixed relative course:

$$L'_{X_t}(K_r) = \int_{-\infty}^{+\infty} p_{X_t}(l)dl \tag{6}$$

where $p_{X_t}(l)$ is the probability of the observer detecting the search object at the segment $(l, l+dl)$;

$p_{X_t}(l) = 1 - \exp\left(\int_{-\infty}^{+\infty} \gamma_{X_t}(t)dt\right)$ when an observer performs continuous surveillance of the search space;

$p_{X_t}(l) = 1 - \prod_{i=1}^{M} [1 - p_i(l)]$ when an observer performs discrete surveillance of the search space.

In Fig. 2 shows the clarify calculation of the parameters $p_{X_t}(l)$ and $L'_{X_t}(K_r)$.

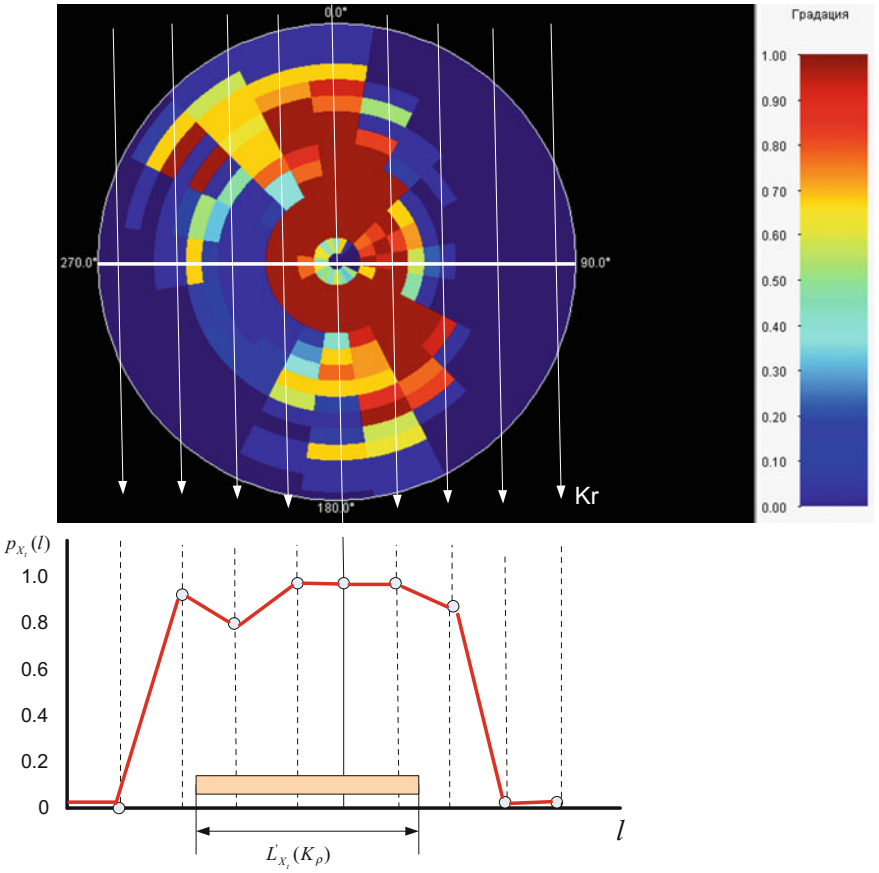


Fig. 2 Calculation of the clarify parameters $p_{X_t}(l)$ and $L'_{X_t}(K_r)$

The optimal distribution of a search effort assumes such search conditions such that that they provide satisfaction of the following criteria:

$$\begin{aligned} &\lambda(t): \int_{\Phi} w(X_t)d(X_t) \rightarrow \max_{\Phi_t \in V_0} \\ &\bar{T}: \int_0^{T_p} e^{-\left(\int_0^t \lambda(t)dt\right)} \rightarrow \min_{t \le T_p} \end{aligned} \tag{7}$$

Considering that we do not know the law of $\Phi(X_t)$, the analytical solution of such an optimization problem is most difficult. In this case, we propose a heuristic algorithm that provides rather good results of a rational search-effort distribution. The algorithm uses the following sequence of actions:

- Formation of three-dimensional regular grid in the search space. There are steps of grid along the x axis is Δh_x, along the y axis—Δh_y, and along the z axis—Δh_z.
- Calculation of search efforts and the target-detection intensity $\lambda\left(X_{ijk}\right); i = 1...I$, $j = 1...J$, $k = 1...K$, $I = L/\Delta h_x$, $J = B/\Delta h_y$, $K = H/\Delta h_z$ in grid points. Mathematical models (2) through (6) are used for the calculation.
- Searching of grid node X_1, which has maximum-intensity target detection:

$$\begin{aligned} &X_1\lambda(X_1) \rightarrow \max_{X_{ijk} \in V_0} \lambda(X_{ijk}), i \in A, j \in B, k \in C \\ &A = \{1, 2, \ldots, I\}, B = \{1, 2, \ldots, J\}, C = \{1, 2, \ldots, K\} \\ &X_1 = \left(x_{p1}, y_{r1}, z_{s1}\right) \end{aligned}$$

- Determination of the second node X_2. Node X_2 is relative to X_1 at a distance equal to the step of the grid. The node is characterized by maximum target-detection intensity:

$$\begin{aligned} &X_2: \lambda(X_2) \rightarrow \max_{X_{ijk} \in V_0} \lambda(X_{ijk}),\ i \in A_1,\ j \in B_1,\ k \in C_1 \\ &A_1 = \{p_1 - 1, p_1 + 1\}, B_1 = \{r_1 - 1, r_1 + 1\},\ C_1 = \{s_1 - 1, s_1 + 1\} \\ &X_2 = \left(x_{p2}, y_{r2}, z_{s2}\right) \end{aligned}$$

Determination of the third and subsequent nodes that belong to the track:

$$\begin{aligned} &X_3: \lambda(X_3) \rightarrow \max_{X_{ijk} \in V_0} \lambda(X_{ijk}),\ i \in A_3,\ j \in B_3,\ k \in C_3 \\ &A_3 = A_2/A_1,\ B_3 = B_2/B_1,\ C_3 = C_2/C_1 \\ &A_2 = \{p_2 - 1, p_2 + 1\},\ B_2 = \{r_2 - 1, r_2 + 1\},\ C_2 = \{s_2 - 1, s_2 + 1\} \\ &X_3 = \left(x_{p3}, y_{r3}, z_{s3}\right) \\ &X_n = \left(x_{pn}, y_{rn}, z_{sn}\right) \end{aligned}$$

Search strategy $L(X)$ assumes the motion for a piecewise-linear trajectory that is to pass through nodes $X_1, X_2, \ldots, X_n$ with the largest values of target-detection intensity. In addition, the algorithm provides the following:

- Uniform inspection over the search space;
- Exception of re-inspection of areas in the search space.

The probability of detection the search object is as follows:

$$P(t) = 1 - \exp\left[-\sum_{i=1}^{N-1} \lambda(X_i)t_i\right],\ t \le T_p,\ t_i = L_i/v_s$$

where L_i is the length of the piecewise-linear part of the trajectory.

In most practical cases, we can isolate areas with relatively similar values of target-detection intensity within the search space. In this case, the following approach to the distribution of the search effort can be applied as follows:

- We separate areas (clusters) $S_1 \ldots, S_n$ within the boards of the search space; these clusters unite the points of space with similar levels of the search object–detection intensity:

$$V_i : \lambda(X_k) - \lambda(X_l) \le |\delta_\lambda|,\ V_i \in V, X_k \in V_0,\ X_l \in V_0$$

- For each cluster, the average value of the detection intensity is calculated:

$$\bar{\lambda}_{S_1}, \bar{\lambda}_{S_2}, \ldots, \bar{\lambda}_{S_i}, \ldots, \bar{\lambda}_{S_n}$$

- We assign ranking of clusters in decreasing order of detection intensity:

$$\lambda_1 > \ldots > \lambda_{j-1} > \lambda_j > \lambda_{j+1} > \ldots > \lambda_n$$

Search strategy $L(X)$ assumes a sequential search in areas (clusters) according to the principle of decrement of target-detection intensity. This provides an advantage in the mathematical expectation of the detection time with the same final target-detection probabilities. It is obvious that the strategy should ensure uniform inspection in the search areas and exclude repeated inspection of the space in each cluster.

Figure 3 illustrates the effect of the sequence of search operations in areas with different intensity on the search effectiveness. The search intensity in different areas is: $\lambda(t_1) = 0.3[1/h]$, $\lambda(t_2) = 0.012[1/h]$, $\lambda(t_3) = 0.6[1/h]$, $\lambda(t_4) = 0.03[1/h]$. The solid red line $[P1(t)]$ depicts the least effective strategy (search with increasing intensity); the dashed green line $[P2(t)]$ depicts the intermediate strategy (search strategy mixed intensity); and the dashed blue line $[P3(t)]$ depicts an optimal plan (search with decreasing intensity).

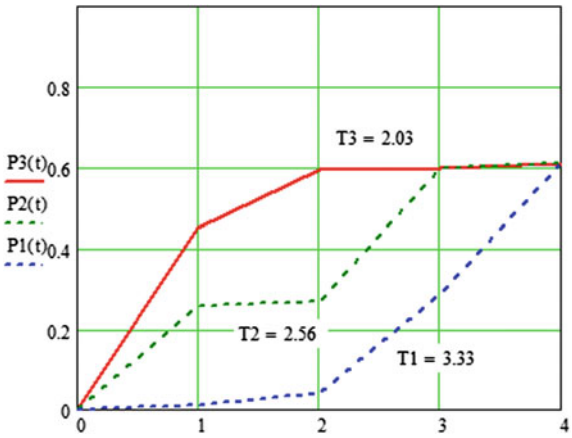


Fig. 3 The least-effective strategy

3 GIS Support of Search-Efforts Distribution

The processing of spatial data is necessary for a practical solution to the problem of search-efforts distribution. GIS support consists of realization of the program and technical measures designed to ensure access to this data and its processing and visualization. GIS-support should provide the following:

(A) Geospatial model search–operation formation:

- Formation, visualization, geo-referencing, and program access to data characterizing the search area;
- Formation, geo-referencing, and program access to data characterizing the probability of density distribution of the target position;
- Provision of access, visualization, and management of data characterizing the properties of the environment in the search space; and
- Formation and geo-referencing of the computational grid;

(B) Formation of the observer properties:

- Calculation of search efforts in the computational grid nodes; and
- Calculation of target-detection intensity in the computational grid nodes;

(C) Optimization of the search-efforts distribution:

- Allocation of areas (clusters) that unite points of space with similar levels of the search object-detection intensity;
- Determination of the starting point of the track;
- Determination of the next track points;
- Creation and visualization of the observer track;
- Calculation parameters of the observer track;

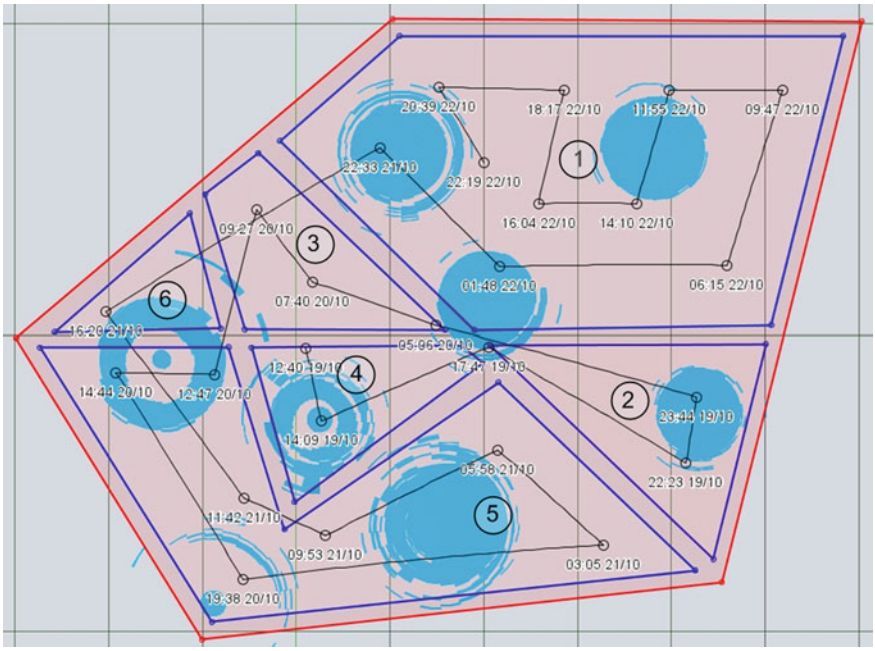


Fig. 4 Example of a rational search-efforts distribution. Clusters of equal target-detection intensities are marked by *blue* polylines

(D) Evaluation of the effectiveness of the chosen strategy.

Figure 4 shows an example of the rational allocation of search effort in A search space area of 1280 m^2. The search space is shown as a limited polyline in red.

Cluster 1 is characterized by the conditions of sound propagations in a shallow sea (Fig. 5).

Clusters 2 through 4 and 6 are characterized by a transition from shallow to deep sea (Fig. 6).

Cluster 5 is characterized by conditions of sound propagations in deep sea (Fig. 7).

Search capabilities of the observer in different nodes of the computational grid are shown in Figs. 4 (two-dimensional representation) and 8 (three-dimensional representation).

Calculations that provide justification for a rational search effort distribution are listed in Table 1.

As illustrated in Fig. 4, the motion track of the observer provides a consistent (from highest to lowest intensity) search in the selected clusters. Estimation of the expected search results according to the chosen strategy of the search-effort distribution are shown in Fig. 9.

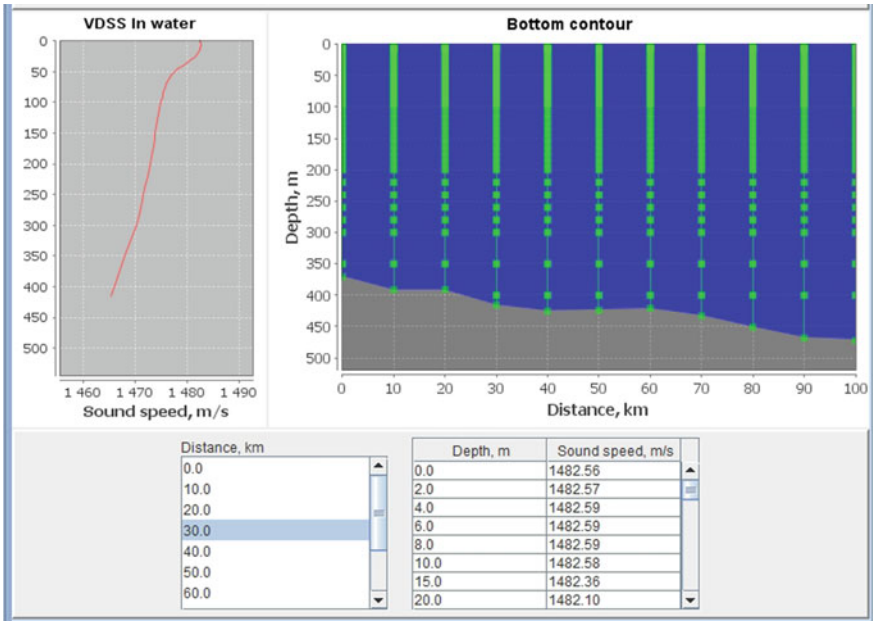


Fig. 5 Typical conditions of sound propagation in cluster 1

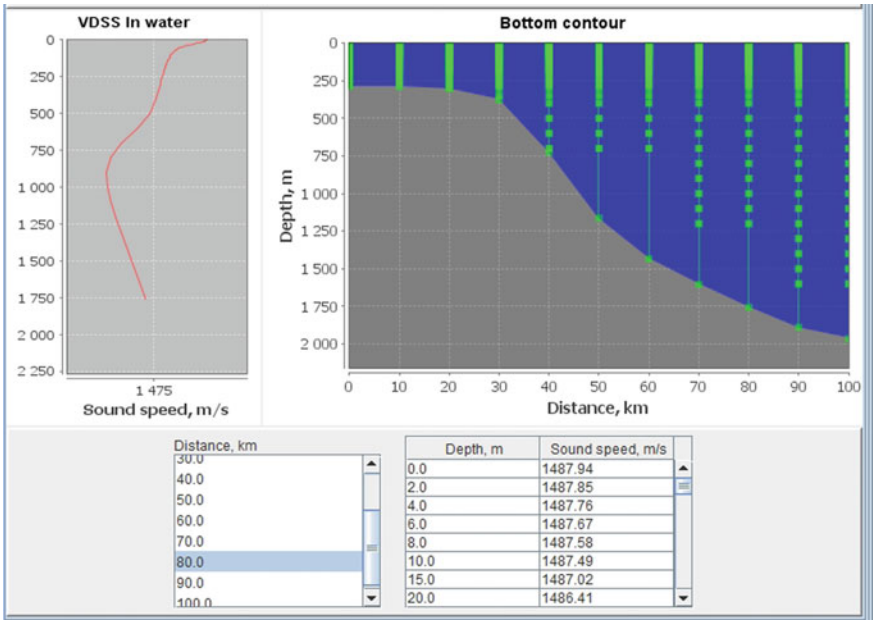


Fig. 6 Typical conditions of sound propagation in clusters 2 through 4 and 6

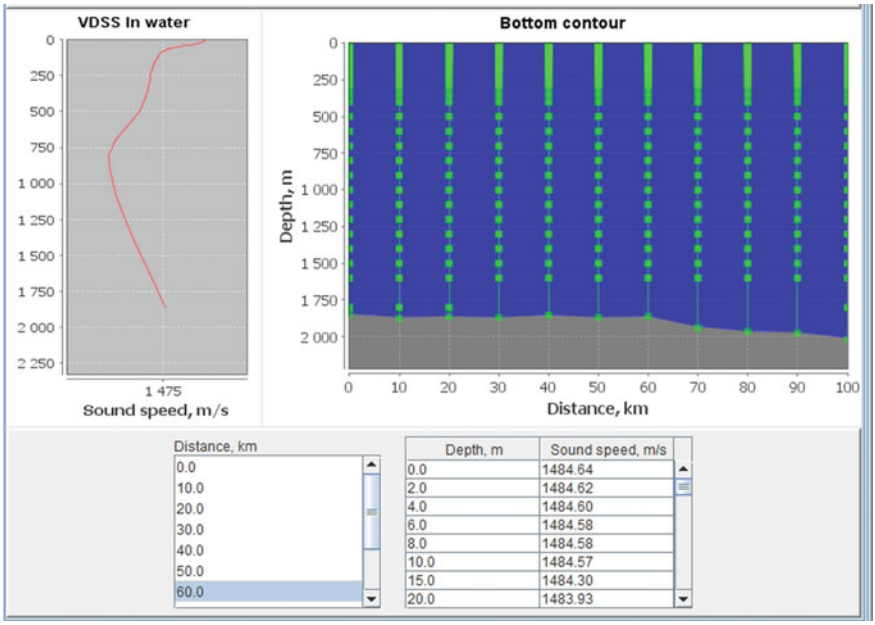


Fig. 7 Typical conditions of sound propagation in cluster 5

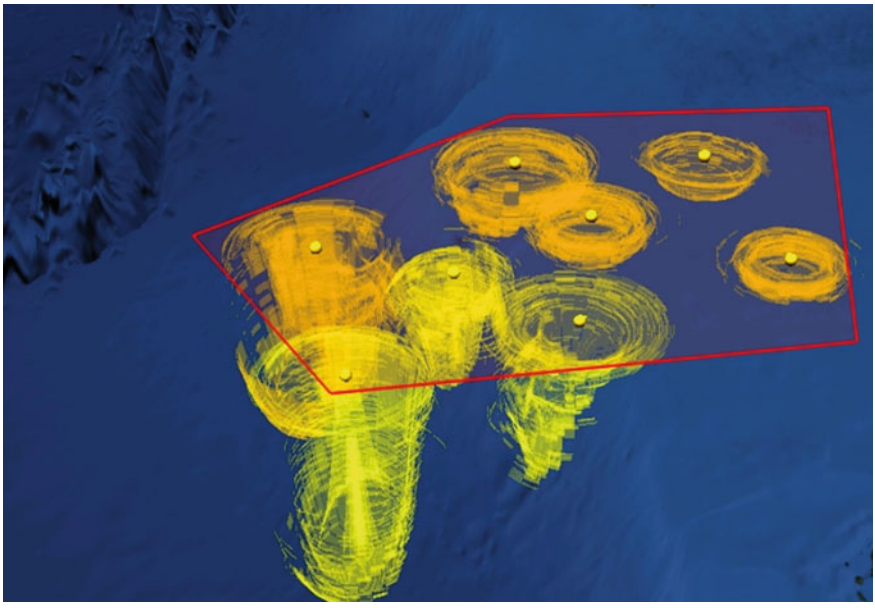


Fig. 8 Search capabilities of the observer in different nodes of the computational grid

Table 1 Calculations that provide justification for a rational search-efforts distribution

Cluster No.	Φ, mile^3/h	w	λ, 1/h	Cluster rank
1.	32.4	29.10^{-5}	0.00939	5
2.	32.4	68.10^{-5}	0.02203	2
3.	40.5	29.10^{-5}	0.01174	3
4.	40.5	68.10^{-5}	0.02754	1
5.	16.2	29.10^{-5}	0.00470	6
6.	16.2	68.10^{-5}	0.01101	4

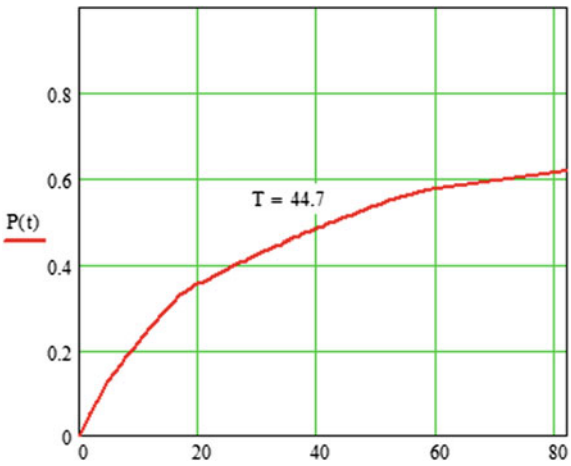


Fig. 9 Estimation of the expected search results

4 Conclusion

1. Solution to the problem of search-efforts distribution in a highly changeable environment is associated with the use of geospatial data characterizing the properties of the observer, the object of observation, and the environment. GIS support for solving such problems is associated with the provision of user services, ensuring the formation of the initial situation simulation, the organization of the computational process, and the provision of access and management of geospatial data, intellectual support for the formation of the search process plan, as well as graphic interpretation of the simulation results.
2. In practice, as a tool of geoinformation support for solutions to this kind of task, it is advisable to use intelligent GIS (IGIS). The composition of IGIS includes components such as a GIS user interface, expert systems, an ontology system, a sonar calculations server, and a computing-tasks complex.
3. The specificity of tasks associated with the distribution of search efforts, as well as other similar tasks, that use the model of two-dimensional medium defines the three-lever client-server architecture of the special software that provides GIS

support. The first level should correspond to the level of data. At the second level, a set of servers and specialized services should be placed. At the third level should be clients (Web clients or thick clients).

References

1. Popovich V (2000) Modeling, evaluation of efficiency and optimization of naval surveillance systems (theory of mobile objects research). St. Petersburg
2. Koopman DJ (1956) Oper Res 5(5):613–626
3. Hellman O (1985) An introduction to the theory of optimal search. Moscow
4. Diner I (1969) Operational research: Naval Academy
5. Prokaev AN (2005) Information fusion in objects search problems with the intellectual geographic information systems. Collected papers of IF&GIS seminar, St. Petersburg
6. Ermolaev V (2013) Modeling of search actions under the conditions of at environment properties' variability. Collected papers of IF&GIS seminar, St. Petersburg
7. Popovich VV, Ermolaev VI, Leontiev YB, Smirnova OV (2009) Hydroacoustic fields modeling based on intellectual geographic information system. Artif Intell Decisions Making, N 4, 2009

Part IV
Ports, Maritime Transportation, and Logistics

Community Structures in Networks of Disaggregated Cargo Flows to Maritime Ports

Paul H. Jung, Mona Kashiha and Jean-Claude Thill

Abstract We investigate community structures in disaggregated cargo flows that are part of the global maritime trade. Individual origin–destination flows of containerized cargo on the hinterland-side of the global maritime trade collectively form a complex and large-scale network. We use community-detection algorithms to identify natural communities that exist in such highly asymmetric networks. The analysis compares the structures that are identified in two geographic regions: China and Europe. We trace shipments by their trajectory patterns considering domestic movement, cross-country movement, and trans-shipments. Then we use the *Infomap* and *Walktrap* algorithms to detect structural equivalence between shipments from hinterland cities to port cities. Data consist of individual flows of containerized trade bound for the United States in October 2006. We draw conclusions on the relative performance of the two algorithms on asymmetric data networks, on the effects of the geographic context and type of flows on their performance, and on the spatial integrity of community structures and the impact of market contestability on the semantics of detected communities.

Keywords Global trade · Maritime network · Community detection · Port hinterlands · Logistics

1 Introduction

In contemporary social data sciences, it is customary to view the world as a set of entities with varied and heterogeneous degrees of connectivity. Entities (people, places, firms, animals, bots, etc.) have certain interactions or functional relationships that can be virtual (e.g., information as in phone calls, email messages, text messages), or physical (e.g., financial transactions, shipping flows, commuter

P.H. Jung · M. Kashiha · J.-C. Thill (✉)
Department of Geography and Earth Sciences, University of North Carolina at Charlotte, 9201 University City Boulevard, Charlotte, NC, USA
e-mail: Jean-Claude.Thill@uncc.edu

V. Popovich et al. (eds.), *Information Fusion and Intelligent Geographic Information Systems (IF&IGIS'17)*, Lecture Notes in Geoinformation and Cartography, DOI 10.1007/978-3-319-59539-9_13

flows). More generally, there is a shared property (e.g., owning the same car manufacturer, attending the same school, having browsed the same Web site) that denotes some form of affinities. Social network analysis (SNA) [14] has emerged as a scientific endeavor aimed at studying social structures in complex and large-scale networks.

The community structure of a network is an important construct in a networked society because it reveals the innate order and organization that regulates relationships between entities, as embedded in big relational data, and points to the possible complexity of such networks. Communities exist if the nodes (entities) of the network can be grouped into natural sets of nodes such that each set of nodes is more densely connected internally than it is connected with nodes belonging to other sets. Given the non-deterministic nature of node interactions and relationships, sets may exhibit a certain degree of overlap.

The detection of community structures is usually performed on "symmetric" networks or "fairly symmetric" networks, whether directed or not, where the number of emitting nodes is approximately the same as the number of receiving nodes. Social media networks would be a good example of such situation as would the network of mobile phone calls between broadcast towers and their surrounding cells. Economic relationships, in contrast, can be much more focused with strong-dependence relationships organized on more hierarchical principles. Thus, the performance of community-detection algorithms known to be effective on symmetric networks may not carry over well to asymmetric social networks. In this chapter, we study the economic territories shown by individual origin–destination flows of containerized cargo on the export-side of global maritime trade at the scale of two large world economic regions, namely, Europe and China. The multitude of export flows to ports is traditionally comprehended by the concept of port hinterland. Because of the tyranny of distance and related transaction costs on economic and business decisions made by agents, the spatial expression of port hinterlands used to take the form of geographic territories that were rather mutually exclusive (not unlike weighted Thiessen polygons) [5]. Although this is hardly the case anymore in certain regions [6] because markets are less protected from competition along the model of spatial monopolies and because competitive forces have increased the degree to which markets are contestable, we believe that network communities can be useful constructs to study the structural properties of economic territories and their spatial expressions.

Geographic regionalization can be implemented in a number of different ways. A straightforward way is by partitioning a continuous land expanse along political boundaries from local jurisdictions to country borders. However, economic relations between regions often extend beyond political boundaries, which makes political boundaries less relevant in studies of regional economics and economic geography. Some localities may have closer economic ties with remote cities than with nearby neighborhoods in terms of physical distance. Moreover, regions divided by political boundaries are mutually exclusive, whereas economic territories may overlap with one another where markets are contestable and the tyranny of distance is under check. Incorporating these notions, we proceed to identify natural

shipping communities in Europe and China and spatialize them by identifying their spatial embeddedness or footprint. Thus, the spatial layout of economic territories can be investigated as a manifestation of the economic or logistical distances (i.e., costs) as experienced by the actors involved in trade across spaces. We conduct this work based on topological relationships within economic networks, not on simple geographical adjacencies.

In this study, we investigate how economic territories are structured within a cross-country region. We focus on the flows of global maritime shipments from their source point to the forwarding ports before leaving the departure continents to the United States. We first examine the formation of clusters of localities that have similar patterns of shipping to forwarding ports within Europe and China, respectively. By using port hinterlands as our proxy for economic territories, our study traces the distribution of economic territories and finds their geographical traits across a cross-country region.

We compare the formation of port hinterlands in Europe and China, which enables us to discuss how geographic constraints affect the formation of economic territories in the context of the economic concept of contestability. Two regions are selected for this study because they can provide good geographical comparison in terms of port shipments. Europe has a similar land size (10,180,000 km^2) as China (9,596,961 km^2), yet the shape of the European coastline is much more complex topologically than that of China, which implies the possibility of widely different and multiple shipping paths from many shipping sources. Europe has an indented coastline with several peninsulas and gulfs and faces seas along its northern, southern, and western sections, whereas China has a simple and topologically straightforward coastline (linear configuration). Thus, the economy of Europe is more exposed to sea-sides than that of China, which enables European sources to have diversified shipping routes and potentially be structured in complex economic territories. We determine the relationship between coastline complexity and the diversification of port hinterlands.

Following a network science approach, we conduct community detection to identify shipping and logistical communities focused on ports within Europe and China. The *Walktrap* and *Infomap* algorithms, which are widely recognized as state-of-the-art community detection methods, are adopted for this data-reduction process. We also compare the results of both algorithms and discuss their relative efficiency and validity. Because these two algorithms use different computation processes, it is necessary to discuss their relative performance to reflect the economic landscape of each of the two world regions.

To check the validity of our results, we perform a diagnostic test on each algorithm. For the *Walktrap* algorithm, we conduct a sensitivity analysis of community structure to obtain the length of random-walk steps. The associations between the length of steps, the number of detected communities, and the modularity score are examined. We also run the *Infomap* algorithm 100 times and check the robustness of the modularity score and community structure. This process informs us to more fully appreciate the reliability of the algorithms and their output results.

The rest of this chapter is organized as follows. In the next section, we provide a general description of the U.S. PIERS Trade Intelligence database and explain how we construct the shipping networks between source localities and forwarding ports. We then present the general framework of community detection and compare *Walktrap* and *Infomap*. In the following section, we provide validity checks of two algorithms to establish the optimality and robustness of community structures. Next, we present the results of community detection by the two algorithms in Europe and China. The last section concludes the analysis.

2 Shipping Networks in Europe and China

Our analysis covers global maritime shipments from Europe and China to the U.S. We track the records of containerized export shipping from Europe and China to the U.S. in October 2006. The Port Import Export Reporting Service (PIERS) Trade Intelligence database provides records of door-to-door freight shipment from localities in Europe and China to the U.S. [7]. This database captures the geographical footprints of shipping containers from addresses of origin through connecting ports and finally to U.S. ports of entry. It also includes commodity information, such as shipment quantity (in 20 foot–equivalent units, TEUs) and commodity type.

As we look through the shipment cases, we find that there are multiple types of shipping routes. For example, some shipments go directly from the forwarding ports to a U.S. port of entry without transshipment, whereas others go through one connecting port before being forwarded to forwarding ports and then leaving the source region for the U.S. Here we split the whole network into two categories: the no-transfer shipment network and the transfer shipment network. We analyze them separately to see how shipping behaviors shape the spatial distribution of economic territories.

To illustrate the logistics of communities focused on maritime ports, the network connections between localities and forwarding ports are examined. We construct the origin–destination matrix by setting source localities as the origin and the last forwarding ports in the same continent as the destination. If a shipment goes through multiple ports along its route, the last port where the shipment leaves before arriving at a U.S. port is considered as the destination in the matrix. Links between nodes are weighted by the amount of shipments or TEUs.

The number of source localities in Europe's shipping network is 3567 including 61 forwarding ports, whereas China has 907 shipping localities and 39 forwarding ports. The associated networks and incidence matrices are correspondingly dimensioned on the basis of dyads formed by source localities–forwarding ports. The total amount of shipments exported from Europe to the U.S. is 202,836.09 TEUs in October 2006 and that of China is 567,965.61 TEUs.

3 Community-Detection Methods

3.1 Algorithmic Principles

We use a community-detection procedure to identify shipping territories exhibiting similar and shared logistical relations to forwarding ports of export in terms of global maritime shipping networks. The shipping routes from the source localities to the forwarding ports are traced to detect the community structures.

To find a community structure within a network, the algorithms search for groups of nodes that are connected densely together. In the shipping networks, bundles of localities in the same community will share the same set of forwarding ports as their shipping-route destinations. This enables us to project communities onto the geographic space, delineate geographic territories of forwarding ports, and investigate their geographic layout such as how they are distributed across regions. Although localities can belong to multiple communities, the territories can be within a single country or span multiple countries.

Various algorithms exist to find community structures [4], but we only adopt two of the most recent ones, namely, *Walktrap* and *Infomap*. These algorithms are both based on the notion that a random walk from a node is more likely to be trapped within a community where it starts [11, 12].

The *Walktrap* algorithm is based on arandom walkrandom-walk connection and on the hierarchical structure of the network [11]. The algorithm groups two nodes in the same community if they have close values of probability such that a random walk from each node reaches a common node with a certain length of random-walk steps. It hierarchically detects community structure by sequential nesting of sub-communities. Thus, at the first level, it nests nodes that have close probabilistic distance and thus compose sub-communities. The probabilistic distance is defined by the following equation:

$$r_{ij} = \sqrt{\sum_{k=1}^{n} \frac{\left(P_{ik}^{t} - P_{jk}^{t}\right)^2}{d(k)}} \tag{1}$$

where r_{ij} is the probability that a random walk from i reaches j within t steps; and $d(k)$ is the degree of node i. At the next level, it nests the communities again based on the probabilistic distance. The probabilistic distance between communities has the same form as the distance between two nodes with communities C_i and C_j substituted for nodes i and j. The algorithm nests the communities when two communities are homogenous in terms of probabilistic distances. When applying this algorithm to real-world cases, the length of random-walk steps is parametrically determined, which can generate different results of community structure.

Infomap has been found to outperform other community-detection algorithms [8, 10]. It has recently been used for geospatial analysis such as the analysis of mobile-communication networks [1], bus-transport networks [13] and taxi-travel

patterns [9]. The *Infomap* algorithm considers the efficiency of a random-walk flow, not the hierarchical structure of communities [12]. Based on information theory, this algorithm starts from an effort to find a method to optimize the length of codes. It examines the flow of a random walk within a community and between communities, and splits the community based on the frequency of movements. Thus, connections between nodes within the same community are significantly frequent, whereas connections between nodes in different communities are rare. The algorithm detects the community structure by the map equation [12] as follows:

$$L(M) = qH(Q) + \sum_{i=1}^{m} p_i H(P_i) \tag{2}$$

where $qH(Q)$ describes movements between communities; and $\sum_{i=1}^{m} p_i H(P_i)$ describes movements within communities. The algorithm minimizes the map equation and finds the best community structure M by optimizing entropy scores of movements within communities and between communities. In this study, community detection with *Walktrap* and *Infomap* is performed in the igraph software package in R [2, 3].

3.2 Validity Check

Neither the *Walktrap* nor the *Infomap* algorithm guarantees to detect consistent outcomes of community structure. The outcomes from *Walktrap* may vary with a parameter value, the length of random-walk steps. Pons and Latapy [11] note that the length should be long enough for the algorithm to catch the topology of the graph, but it should not be too long to keep the probability distribution of random walks from reaching the stationary distribution. *Infomap* also returns outcomes that are stochastic, so they can be variable over multiple runs of the algorithm. It should be noted that the results of *Infomap* may not be accurate when the scale of a network is large. In this case, it requires much time to compute vast amounts of flows. For the efficiency of calculation, it employs a fast stochastic and recursive search algorithm, the result of which may not be consistent, such as the Monte Carlo method. We check the validity and consistency of outcomes by applying the *Walktrap* and *Infomap* algorithms multiple times to the Europe and China data sets, i.e., the whole data set as well as the sub-sets with no-transfer and transfer shipments, respectively. For each batch of runs, we examine how community structure differs across the various runs.

For the *Walktrap* algorithm, we conduct the analysis of sensitivity of the shown community structure to the parameter value. The algorithm is run 30 times by changing the length of random walks to perform from 1 to 30. We then observe how the number of detected communities and the modularity score of each case vary as the parameter value is increased. To appreciate the robustness of the

outcome of *Infomap*, we run the algorithm 100 times and observe how the modularity scores are distributed. The distribution can show the overall level of modularity and suggests how stably the algorithm generates the outcome from the network.

The sensitivity analysis on *Walktrap* illustrates that the detected community structure converges to a certain level as the length of random walks increases. For every region and type of shipment, the number of detected communities shows a reciprocal relationship with the parameter value: It converges to a constant value after a sharp decrease <10 steps (Fig. 1). Moreover, the modularity score has an increasing concave curve, which also converges to a certain level of score with little fluctuation beyond 5–10 steps (Fig. 2). These two patterns indicate that the *Walktrap* algorithm can detect a certain community structure in all cases in Europe and China provided that a sufficient length of random walks can be assumed. It implies that the outcome of community structure would not differ much no matter how long the length of random walks is above and beyond a certain threshold value. Therefore, we can rely on the outcome of the *Walktrap* algorithm. We can choose the outcome and parameter value where the community structure starts to converge and stabilize.

The distribution of modularity scores of the *Infomap* algorithm presents both the overall feature of a detected community structure and a stability of algorithm detection. In every set of 100 runs, we can confirm that there is a strong fluctuation across the runs indicating that the algorithm does not always guarantee consistent results of community detection. Considering all of the results, we find three tendencies in the distribution of modularity scores. First, Europe has a greater modularity than China. The modularity scores based on the entire Europe shipment data set remain approximately 0.4, which is greater than for China (Fig. 3). Even in the

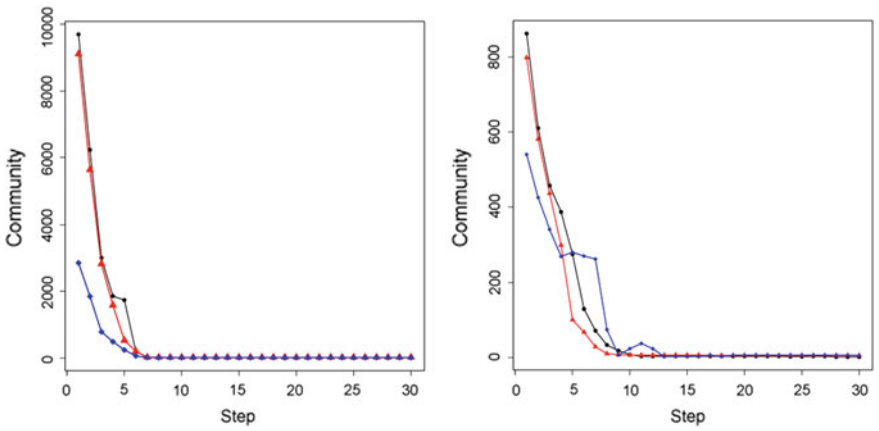


Fig. 1 Sensitivity analysis of the number of detected communities by the *Walktrap* algorithm. (*Left panel*) Europe. (*Right panel*) China. *Black line* the whole network; *red line* no-transfer network; *blue line* transfer network

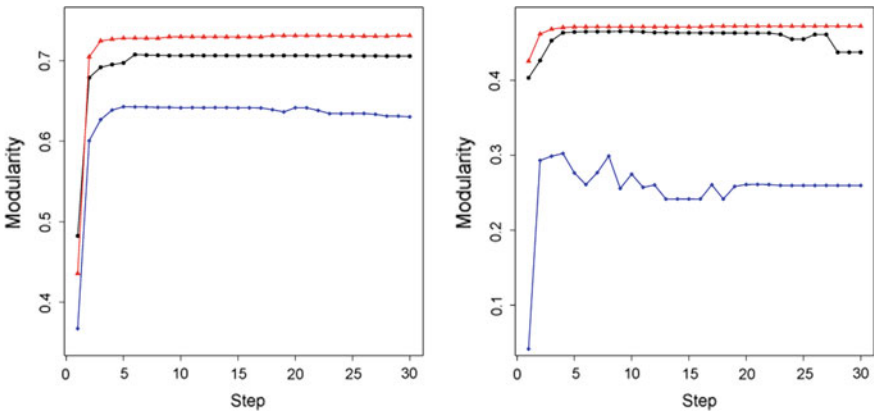


Fig. 2 Sensitivity analysis of modularity by the *Walktrap* algorithm. (Left panel) Europe. (*Right panel*) China. *Black line* the whole network; *red line* no-transfer network; *blue line* transfer network

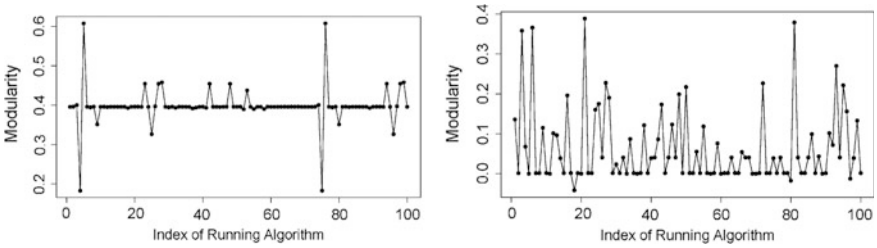


Fig. 3 Robustness test of modularity on the whole network by *Infomap*. (*Left panel*) Europe. (*Right panel*) China

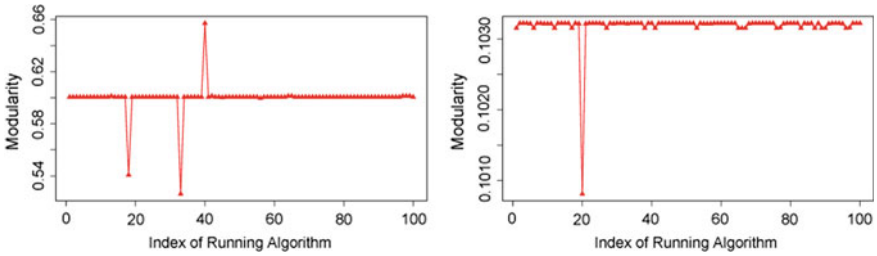


Fig. 4 Robustness test of modularity on the no-transfer network by *Infomap*. (*Left panel*) Europe. (*Right panel*) China

no-transfer network (Fig. 4) and the transfer network (Fig. 5) taken separately, Europe records the values of modularity greater than those of China. This implies that the shipping networks of Europe have more a salient community structure than

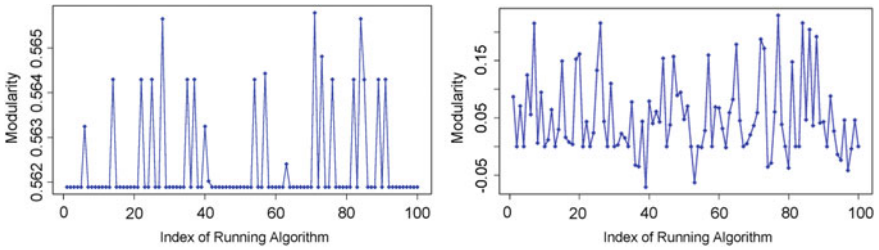


Fig. 5 Robustness test of modularity on the transfer network by *Infomap*. (*Left panel*) Europe. (*Right panel*) China

China, so it is easier to see distinctive divisions of logistical relationships and port hinterlands in Europe. Second, the measured modularity scores of no-transfer shipment networks (Fig. 4) are more stationary than that of transfer shipment networks in both Europe and China (Fig. 5). Thus, *Infomap* cannot detect community structure in a transfer network with good stability, so it is difficult for this algorithm to divide the network in the same way at every run. Last, China has a more volatile distribution of modularity scores than that of Europe. When we compare the results in the whole network (Fig. 3) and in the transfer network (Fig. 5), Europe shows more consistent modularity scores than does China. This implies that it is easier to delineate the community structure in Europe than China, meaning once again that Europe has more distinctions in community structure than China. In other words, the partitioning in economic territories is clearer in Europe.

These two tests provide evidence of the validity of each algorithm as well as the reliability of the results. We find that the *Walktrap* algorithm can clearly delineate a certain community structure in both world regions regardless of shipment types. However, the outcomes of the *Infomap* algorithm are distributed more haphazardly, which leads to some difficulty in deciding which outcome effectively reflects the real logistic community structure and port hinterlands. In next section, we discuss how we choose one of multiple outcomes as a representative result and scrutinize and also compare the distributions of port-based communities in the two world regions.

4 Community Detection

The tests reported previously show that the algorithms may generate inconsistent outcomes that cannot always be relied upon. In this stage, we should ask how we can choose one algorithm as the representative one to use. The results of the two tests can be a reference for selecting the representative outcomes. For the *Walktrap* algorithm, the result where the detected community structure starts to converge is chosen to be a representative outcome. If the number of communities does not change for the next two steps, we consider it as the start of convergence.

For the *Infomap* algorithm, we choose the outcome with the highest modularity among 100 outcomes. Because the outcome of maximum modularity would present a distinct division of community structure within the network, we regard the outcome with the greatest modularity as the best representative outcome to show the most effective community structure.

4.1 The European and Chinese Economic Spaces

The following analysis can be made on the basis of the community detection results. Let us first compare the European and Chinese economic spaces. The geographic distribution of logistics communities is more complex in Europe than in China. The number of communities detected through the process established for both algorithms in Europe is greater than that in China (Table 1, Figs. 6 and 7). China has 3–5 communities depending on the algorithm and the transfer situation, whereas Europe has more diversified community structures, 11–28 communities. This reflects the greater number of forwarding ports engaged in foreign maritime trade in Europe, as well as other factors discussed further later in the text, such as the more complex physical geography of this region, as well as the greater heterogeneity in the degree of contestability of shipping markets in sub-regions of the European economic space, compared with China.

The spatial division of community structure in Europe (Fig. 6) is found to be clearer than in China (Fig. 7) and a better match to port hinterlands. Overall, European communities have a well-delineated territorial expression; these territories are often contiguous spatial expanses. In all transfer situations and with both algorithms, Europe's community structure strongly matches port hinterlands. For instance, as shown by Table 2 in the Appendix (*Walktrap* communities), communities 4 (Spain) and 6 (France) are primarily served by Algeciras and Le Havre, respectively. In the same table, we see that several other communities are more contestable because they are served by two main ports instead of a single one: community 1 is served by Barcelona and Valencia, 2 by Naples and Gioia Tauro, 3 by Antwerp and Rotterdam, 7 by Southampton, Liverpool, and Felixstowe, 8 by Bremerhaven and Hamburg, and 9 by La Spezia, Genoa, and Leghorn.

Table 1 Number of detected communities

		Walktrap	*Infomap*
Europe	Whole	11 (9)	25 (15)
	No-transfer	20 (13)	22 (12)
	Transfer	11 (9)	28 (14)
China	Whole	4 (4)	4 (3)
	No-transfer	7 (5)	4 (3)
	Transfer	6 (5)	4 (3)

Note Number in parentheses is the number of communities whose share of total TEU shipments is >1‰

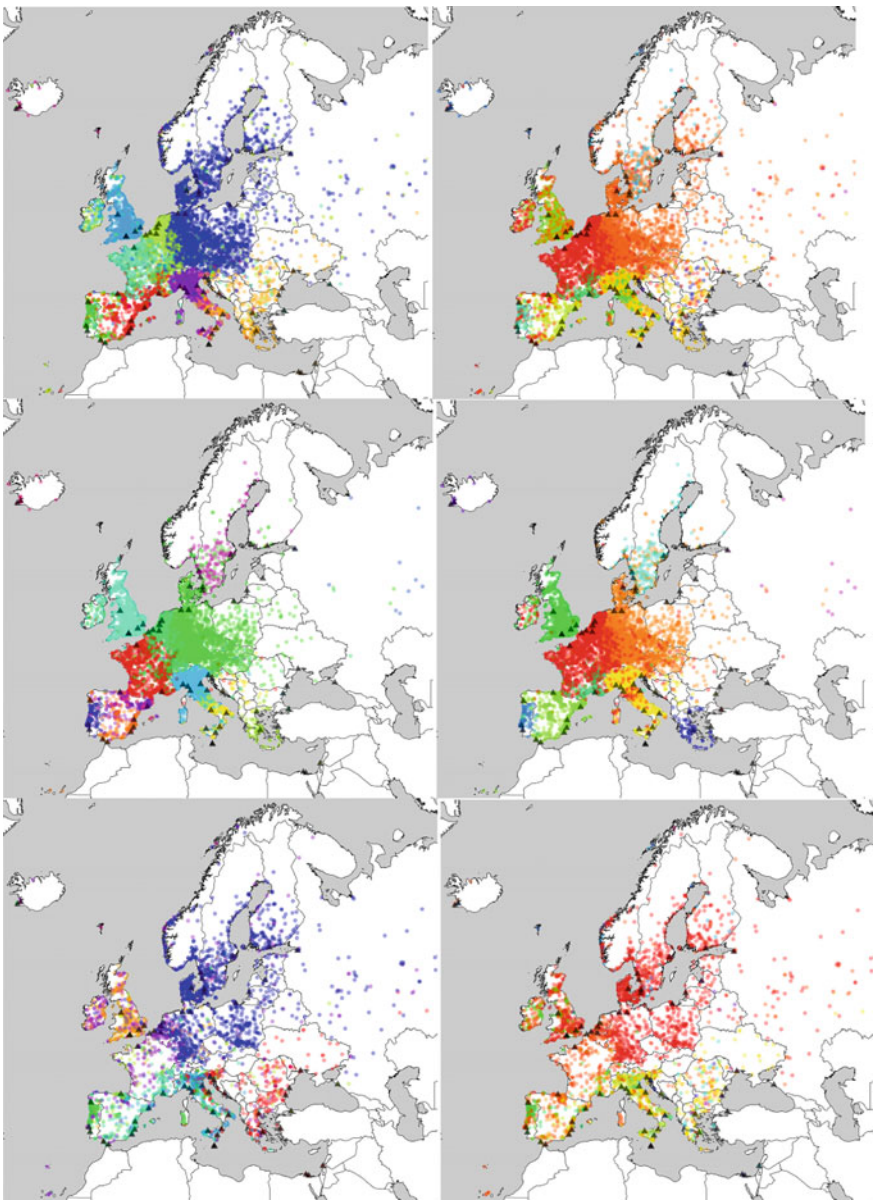


Fig. 6 Community detection in Europe. (*Left panel*) *Walktrap*. (*Right panel*) *Infomap*. (*Top row*) The whole network. (*Middle row*) No-transfer network. (*Bottom row*) Transfer network. Each shipment source is depicted by *a dot color* coded according to the community to which it belongs. *Triangles* ports

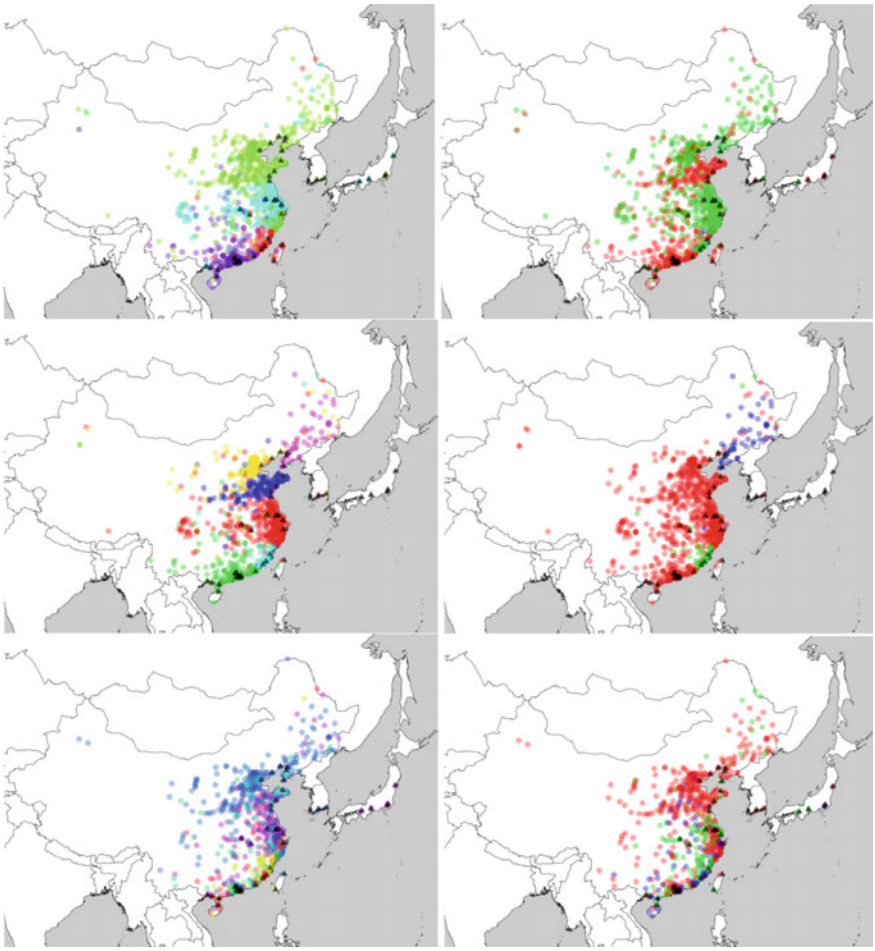


Fig. 7 (*Left panel*) *Walktrap*. (*Right panel*) *Infomap*. (*Top row*) The whole network. (*Middle row*) No-transfer network. (*Bottom row*) Transfer network. Each shipment source is depicted by a dot color coded according to the community to which it belongs. *Triangles* ports

The communities revealed by *Infomap* (Table 3 in the Appendix) are less contestable, more often dominated by a single maritime port, and thus frequently delineating hinterlands. When applying the *Walktrap* and *Infomap* algorithms to the whole network of Europe (Fig. 6a, b), we can observe the clear demarcation between the northern Italian community and the French and German communities because the Alps cut through the two regions and thus form a boundary. Conversely, in China, communities are more interspersed spatially than in Europe, and we can see a hodgepodge in the overall community structure. Detected community structures of the whole network of China (Fig. 7a, b) show exclaves in the spatial distribution and show port hinterlands that are far from traditionally mutually exclusive and compact geographic territories (Tables 4 and 5 in Appendix).

The topological complexity of coastlines seems to affect the spatial division of port hinterlands. Europe has a complex coastline and is exposed to seas on three sides, and this brings with it a diversified logistics community structure owing to the plurality of possible shipping routes. In contrast, China has a relatively simple coastline and only faces seas on the east and south, which mold the communities and associated port hinterlands to have a linear east–west shape.

4.2 Community Structure by Transfer or No-Transfer

The whole-shipping networks can also be compared with the no-transfer networks and with the transfer networks in each region. The community structure of the no-transfer network shows more exclusive territorial delineation in the geographic space and thus is more effective at showing port hinterlands, which are found to be less contestable than when trans-shipment is involved. Especially, the division of communities in Europe follows country borders to a much greater extent than does the whole network (Fig. 6). The same can be said of the no-transfer network of China. The no-transfer shipment seems to be transported to the nearest ports, which means that it is more affected by geographic restrictions. In Europe, the no-transfer shipment seems to be transported within country and hardly crosses country boundaries.

In both Europe and China, shipping communities with transfer are highly intertwined so that the spatialized version of these communities hardly allows to identify the hinterlands. Contestability is pervasive, which is the reason for shipping lines to offer trans-shipment services in the first place.

4.3 Algorithmic Performance of Infomap and Walktrap

The performance comparison between the *Infomap* and the *Walktrap* algorithms can be summarized as follows. In Europe, *Walktrap* captures fewer communities, but it shows clear spatial division of community structures. In the whole network, communities detected by *Infomap* are spatially overlapped without clear borders between them. For example, in the region formed by the Low Countries–Eastern France–Western Germany, the *Walktrap* algorithm differentiates three communities, each one directed toward specific ports, whereas *Infomap* fails to identify the Low Countries as an individual community. *Walktrap* performs better than *Infomap* in terms of recognizing the multiple hinterlands than dot this economic region (in sharp contrast to *Infomap*'s lower discriminating power).

In China, *Walktrap* and *Infomap* detect a similar number of communities. However, the spatial distribution of community structures is quite different; spatial discrimination is notably clearer with the *Walktrap* algorithm.

5 Conclusions

In this research, we investigated community structures in large data sets of disaggregated flows of cargo shipments between source locations and forwarding ports in two large economic regions of the world, namely, Europe and China. Questions we set out to answer included the following:

- How effective are the popular algorithms *Walktrap* and *Infomap* at detecting communities in such highly asymmetric networks?.
- Does their performance depend on the geographic context and type of flows (shipments with or without maritime transshipments)?
- When examined from the perspective of their geographic footprint, do revealed community structures correlate with port hinterlands, and how does market contestability account for heterogeneity across the networks?

Our main findings can be summarized as follows:

1. *Walktrap* performs better in detecting logistics communities and ultimately economic territories within a cross-country region than *Infomap*.
2. *Infomap* has been presented as the best algorithm in network science research based on testing on synthetic data sets, but our research concludes differently on the basis on a real-world network of asymmetric relationships between source locations and forwarding ports.
3. In both Europe and China, the spatial division of community clusters by *Walktrap* is more "crisp" than *Infomap*.
4. *Walktrap* communities are balanced in size and more readily reflect the economic landscape of the two regions than *Infomap*, which has the tendency to extract a few large heterogeneous communities as well as a number of smaller communities that can be hyper-focused in some cases.

Mapping community structures by their shipping types (trans-shipment *vs*. no trans-shipment) can clearly show topological traits between regions within a cross-country region. The mappings exhibit sharp contrasts in their spatial distribution and the boundaries between communities. No-transfer networks show more distinct spatial division between communities, which is indicative of lesser contestability, whereas community structure in transfer networks displays overlaps over large geographic expanses. If we further specify the detailed type of shipment, we can explore how shipping patterns determine the community structures in terms of shipping and logistics industries.

The validity check on the algorithms is suggested as a method to pick a representative outcome among multiple results. Sensitivity analysis and robustness testing can effectively measure consistency of the algorithms. There has been no clear way to set up the length of a random walk for the *Walktrap* algorithm, and the limitations of the search engine of *Infomap* make it difficult to rely on direct results. Our suggested method can solve these problems and suggest a way to determine a representative result.

Spatialization of community structures effectively reveals how geographic traits affect the distribution of economic territories and their shapes. We confirm that the complex coastline of Europe diversifies the compositions of economic territories. A simple community structure is found in China due to the simplicity of its coastline. Combined with geographic information systems, the network-science approach helps to trace the economic landscape inside of one country's economy and within a cross-country region.

Future work will consist of expanding the comparison with other algorithms on the basis of a larger number of performance metrics. We will also test this approach on other world regions.

Appendix: Port-Shipment Volume in Communities

Tables 2, 3, 4, and 5.

Table 2 Port-shipment volume in each detected community (all shipments from Europe [*Walktrap*]) (%)

	Algeciras	Antwerp	Barcelona	Bremerhaven	Felixstowe	Fos	Genoa	Gioia Tauro	Hamburg	La Spezia	Le Havre	Leghorn	Liverpool	Naples	Rotterdam	Southampton	Valencia	Others
	4	3	1	8	7	1	9	2	8	9	6	9	7	2	3	7	1	
1	2.5	0.4	30.5	0.1	–	19.3	0.1	0.2	–	1.0	2.5	–	–	0.9	0.6	–	41.2	0.6
2	1.2	0.6	1.2	1.1	0.1	–	1.4	37.1	0.3	1.9	1.0	1.3	–	32.0	0.4	–	–	20.4
3	0.2	41.2	0.1	3.7	0.3	0.3	–	–	1.0	0.1	1.1	–	0.5	–	50.5	0.2	0.1	0.6
4	45.8	2.2	4.4	0.1	0.1	0.1	–	0.5	–	0.9	0.3	0.7	–	0.2	6.0	0.1	0.8	37.7
6	0.1	4.8	0.6	–	0.1	3.7	–	0.5	–	0.3	84.6	–	–	0.2	4.0	0.1	0.3	0.5
7	0.1	2.0	–	1.2	39.1	–	–	–	0.3	–	0.4	–	26.5	–	3.1	20.2	–	7.0
8	0.1	2.9	–	71.0	1.0	–	–	0.1	17.8	–	0.1	–	–	–	3.1	–	–	3.7
9	1.0	0.1	0.7	0.1	–	–	25.3	1.9	–	48.6	–	20.5	–	1.3	0.3	–	–	0.2
10	–	–	–	–	–	–	–	–	–	–	–	–	0.2	–	0.5	–	–	99.3
Others	–	–	–	3.7	–	–	–	–	1.2	–	–	1.2	–	–	2.4	–	–	91.6

Note The row number in the first column is the number label of each community, and the number in the second row identifies the community to which each port belongs. The percentage shows the proportion of shipping volume each port forwards to U.S. ports among the total shipping volume of localities in each community, i.e., Barcelona forwards 30.5% of shipments sourced from localities in community 1. The ports presented in this table account for 95% of total forwarded TEUs. The communities listed in this table are communities that ship at least 1‰ of total TEUs sourced in Europe

Table 3 Port-shipment volume in each detected community (all shipments from Europe [*Infomap*]) (%)

	Algeciras	Antwerp	Barcelona	Bremerhaven	Felixstowe	Fos	Genoa	Gioia Tauro	Hamburg	La Spezia	Le Havre	Leghorn	Liverpool	Naples	Rotterdam	Southampton	Valencia	Others
	1	1	4	2	2	7	3	3	2	3	1	6	5	3	1	5	4	
1	3.5	32.7	0.7	2.8	0.3	1.0	0.1	0.2	0.8	0.4	13.6	0.2	0.7	0.1	40.2	0.3	0.4	2.0
2	0.1	2.8	–	61.6	10.1	–	–	0.4	14.9	0.1	0.2	–	2.0	–	2.9	2.2	–	2.6
3	1.0	0.1	0.8	0.1	–	–	22.2	9.9	–	41.7	0.1	10.9	–	9.6	0.1	–	–	3.4
4	2.3	0.1	36.8	–	–	1.9	0.2	0.2	–	0.4	1.1	–	–	0.7	0.3	–	55.1	0.8
5	0.1	0.7	–	0.1	7.9	–	–	–	0.1	–	0.5	–	49.7	–	1.1	31.8	–	8.0
6	0.7	–	0.3	–	–	–	6.8	1.4	–	2–	0.1	69.2	–	1.3	–	–	–	0.1
7	1.5	1.2	12.0	0.1	–	76.6	–	–	–	2.5	1.3	–	–	2.2	0.1	–	2.6	–
8	2.1	4.9	0.8	0.1	–	–	–	–	–	–	–	–	–	–	3.3	0.1	0.7	88.0
9	–	2.3	–	23.3	–	–	–	–	6.7	–	0.3	–	–	–	8.6	–	–	58.9
10	–	–	–	–	–	–	–	–	–	–	–	–	0.2	–	0.5	–	–	99.3
11	–	1.1	–	–	–	–	–	4.1	1.5	–	4.1	0.5	–	–	–	–	0.3	88.4
13	0.4	0.2	–	1.5	–	–	–	30.1	–	0.6	4.2	–	–	8.8	–	–	–	54.2
14	–	2.9	–	22.7	–	–	–	–	41.8	–	–	–	–	–	9.1	–	–	23.5
18	–	0.8	–	28.0	–	–	–	–	16.3	–	–	–	–	–	3.4	–	–	51.4
23	0.9	1.9	–	–	–	–	–	90.0	–	–	4.9	–	–	–	–	–	–	2.3
Others	–	6.3	3.2	0.7	0.2	–	0.6	7.0	0.2	1.0	3.0	1.1	–	0.1	1.3	–	–	75.4

Note The row number in the first column is the number label of each community, and the number in the second row identifies the community to which each port belongs. The percentage shows the proportion of shipping volume each port forwards to U.S. ports among the total shipping volume of localities in each community, i.e., Barcelona forwards 30.5% of shipments sourced from localities in community 1. The ports presented in this table account for 95% of total forwarded TEUs. The communities listed in this table are communities that ship at least 1‰ of total TEUs sourced in Europe

Table 4 Port-shipment volume in each detected community (all shipments from China [*Walktrap*]) (%)

	Busan	Chiwan	Hong Kong	Ningbo	Qingdao	Shanghai	Shekou	Tianjin	Xiamen	Yantian	Others
	2	4	4	2	2	3	4	2	1	4	
1	1.5	0.7	11.5	2.5	0.2	2.0	0.3	0.1	58.3	0.7	22.1
2	17.2	0.1	2.3	26.4	22.4	8.8	0.1	13.5	0.2	0.5	8.7
3	3.2	0.2	1.2	4.9	0.5	87.1	0.1	0.2	0.1	0.6	1.9
4	1.1	5.5	16.1	1.1	0.2	1.2	6.4	0.1	0.4	66.9	1.2

Note The row number in the first column is the number label of each community, and the number in the second row identifies the community to which each port belongs. The percentage shows the proportion of shipping volume each port forwards to U.S. ports among the total shipping volume of localities in each community, i.e., Barcelona forwards 30.5% of shipments sourced from localities in community 1. The ports presented in this table account for 95% of total forwarded TEUs. The communities listed in this table are communities that ship at least 1‰ of total TEUs sourced in Europe

Table 5 Port-shipment volume in each detected community (all shipments from China [*Infomap*]) (%)

	Busan	Chiwan	Hong Kong	Ningbo	Qingdao	Shanghai	Shekou	Tianjin	Xiamen	Yantian	Others
	1	2	1	1	2	1	1	1	1	2	
1	4.4	4.2	13.0	2.9	9.4	2.7	5.1	2.2	0.4	53.2	2.5
2	6.6	0.4	3.0	14.0	0.6	57.5	0.1	4.2	7.7	0.6	5.3
3	1.6	1.1	8.5	3.9	0.1	3.0	0.3	0.2	24.5	0.5	56.1
4	–	–	–	–	–	–	–	–	–	–	100

Note The row number in the first column is the number label of each community, and the number in the second row identifies the community to which each port belongs. The percentage shows the proportion of shipping volume each port forwards to U.S. ports among the total shipping volume of localities in each community, i.e., Barcelona forwards 30.5% of shipments sourced from localities in community 1. The ports presented in this table account for 95% of total forwarded TEUs. The communities listed in this table are communities that ship at least 1‰ of total TEUs sourced in Europe

References

1. Chi G, Thill J-C, Tong D, Shi L, Liu Y (2014) Uncovering regional characteristics from mobile phone data: a network science approach. Pap Reg Sci 95(3):613–631
2. Csardi G (2015) Package 'igraph'. The Comprehensive R Archive Network. http://cran.r-project.org/web/packages/igraph/igraph.pdf
3. Csardi G, Nepusz (2006) The igraph software package for complex network research. Int J Complex Syst 1695(5):1–9
4. Fortunato S (2010) Community detection in graphs. Phys Rep 486(3–5):75–174
5. Hanjoul P, Beguin H, Thill J-C (1989) Advances in the theory of market areas. Geogr Anal 21 (3):185–196
6. Kashiha M, Thill J-C (2016) Spatial competition and contestability based on choice histories of consumers. Pap Reg Sci 95:877–895
7. Kashiha M, Thill J-C, Depken CA II (2016) Shipping route choice across geographies: coastal versus landlocked countries. Transp Res Part E 91:1–14
8. Lancichinetti A, Fortunato S (2009) Community detection algorithms: a comparative analysis. Phys Rev E 80(5):056117
9. Liu X, Gong L, Gong Y, Liu Y (2015) Revealing travel patterns and city structure with taxi trip data. J Transp Geogr 43:78–90
10. Orman GK, Labatut V, Cherifi H (2011) On accuracy of community structure discovery algorithms. J Convergence Inf Technol 6(11):283–292
11. Pons P, Latapy M (2006) Computing communities in large networks using random walks. J Graph Algorithms Appl 10(2):191–218
12. Rosvall M, Axelsson D, Bergstrom CT (2009) The map equation. Eur Phys J Spec Top 178(1): 13–23
13. Sun Y, Mburu L, Wang S (2016) Analysis of community properties and node properties to understand the structure of the bus transport network. Physica A 450:523–530
14. Wasserman S, Faust K (1994) Social network analysis: methods and applications. Cambridge University Press, New York

Simulation Modeling of Maritime Monitoring Systems with the Application of an Information Technology Complex

Pavel Volgin and Vladimir Deveterikov

Abstract In this chapter, capabilities of the simulation modeling methods based on modern informational technology are considered in order to help increase decision validity, efficiency of organization, and functioning of monitoring systems and of sea environmental control. The necessity and possibility of application of such informational technologies as geoinformational technologies and expert systems in the modeling structure is determined. It is demonstrated that application of an information technology complex allows to significantly expand the modeling capabilities of processes control and monitoring while taking into account the solution to harmonization problem as well as the integration and fusion of data as a part of simulation modeling problems.

Keywords Simulation modeling · Expert systems · Geoinformation technology · Decision efficiency · Maritime-monitoring systems

1 Introduction

At present, it is difficult to name an area of human activities where modeling methods are not applied in some capacity. This especially concerns the area of the development of complex information-management systems [1]. During the organization of design and the creation of maritime situation monitoring systems, the problem of validity and efficiency of decision-making arises during their management in the process determining of their functioning and development.

P. Volgin (✉)
SPIIRAS Hi Tech Research and Development Office Ltd, 14 Line, St. Petersburg, Russia
e-mail: volgin@oogis.ru

V. Deveterikov
The Military Educational and Research Centre Naval Academy, Admiral of the Fleet of the Soviet Union N.G. Kuznetsov, Petrodvorets-4, Razvodnaya Street, 15, 198514 Saint-Petersburg, Russia
e-mail: delphin76@inbox.ru

V. Popovich et al. (eds.), *Information Fusion and Intelligent Geographic Information Systems (IF&IGIS'17)*, Lecture Notes in Geoinformation and Cartography, DOI 10.1007/978-3-319-59539-9_14

The processes, realized by complex spatially distributed dynamic systems such as maritime situation and activity monitoring and control systems, are very complicated, dynamic, and, as a rule, stochastic and large-scale in nature. Exactly the complexity of such systems defines the absence of an opportunity to develop and perceive them on the basis of analytical (computational) methods within the limits of a single mathematical apparatus. Meanwhile, major specificities of decision-making conditions include a critical time shortage, the necessity for close collaboration with other information systems, and, usually, the existence of an antagonistic (competing) side [2]. All of the above-listed conditions demand objective methods of justification of decision-making methods for solving management tasks at different stages of a monitoring system's life cycle. In situations when mathematical formalization can not be applied to ensure an analytical solution to a problem, the major approach is to use simulation modeling methods based on modern information technologies.

The selection and application of a simulation modeling method for the mathematical description of a maritime situation monitoring processes is given in first section of this chapter. In the second section, approaches to creating a modeling shell for users are described. Possibilities of adaptation of a simulation model for performing experiments concerning the study and estimation of the effectiveness of functioning of various maritime monitoring systems are given in the third section.

2 Application of Simulation Modeling Methods for the Monitoring Research Process in a Sea-Zone Environment

Application of the existing objective estimation methods for justification of the decisions, plans, and management of their realization process can significantly increase the efficiency and quality of the organizational and technical functioning of the system [3]. These methods should be applied both for the estimation of different parameters connected to the implemented process as well as for the estimation of the influence of the whole process and its components and external conditions on the quality of system's task solving.

Currently simulation modeling has become one of the major methods of exploring such dependencies. As a rule, real processes and systems are investigated using two types of mathematical models: analytical ones and simulating ones. In the analytical models, the behavior of real processes and systems are specified by explicit functional relationships (linear and nonlinear equations, differential and integral systems of equations). However, obtaining these dependences is possible only for comparatively simple real processes and systems. When phenomena are complex, diversified, dynamical, a researcher is bound to simplify them.

That is why simulation modeling is necessary for the investigation and discovery of dependencies that are distinctive features of processes of monitoring and

controlling of situation and/or activity in the sea zone. It is well known that simulation modeling starts from the study of a modeled system and from its description in terms of logical schemata and functionality networks [4]. This provides the necessary correlation between a simulation model's mathematical apparatus and its subject domain (modeled system).

The informational technology development opens new opportunities in complex distributive systems research. One of them is an operative decision-making opportunity based on spatial data analysis with the application of GIS and with integration of the modeling results of a monitoring system in underwater, above-water, and air environments. The main capabilities of a geoinformational system are the following:

- spatial and time integration of applied data and necessary visualization level of the realizable process;
- adaptation of the simulation model with consideration of the particularities of the real monitoring system under study and its discovered functioning patterns; and
- necessary visualization level of the realized process [5].

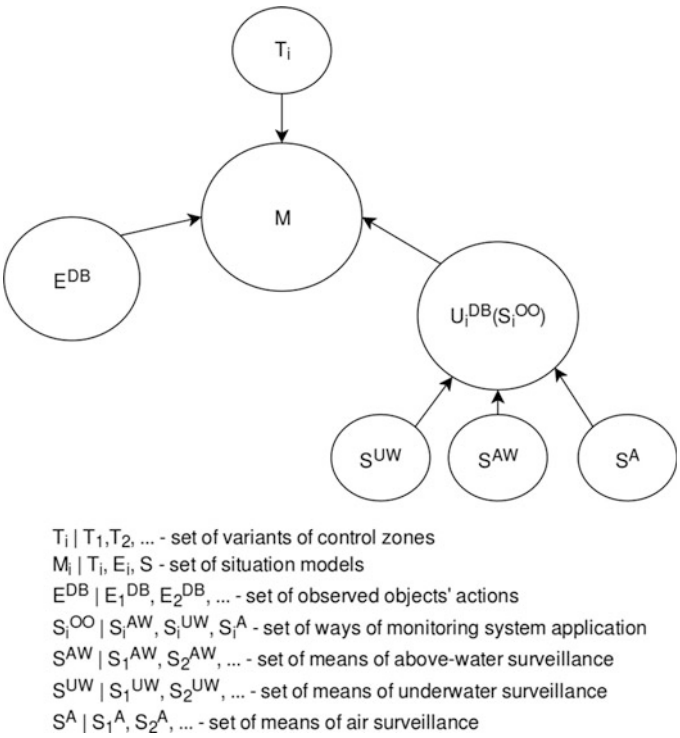


Fig. 1 The diagram of the situation models' set formation in a maritime monitoring system

With that, the efficiency estimation of the monitoring process is a multi-criteria problem connected to the effectiveness estimation of every way of using the monitoring system that realizes this process. This factor leads to the necessity of sorting through a set of results in order to assess the practicability of monitoring system's structure for the given functioning conditions. A logical diagram of situation models' set formation in a maritime monitoring system is given in Fig. 1.

The diagram shows an aggregate of sets of elementary events, actions, and connections between them. Together they form the basis of any maritime-monitoring process. Such sorting through the set of ways of the monitoring system's functioning and through the efficiency estimation results can be performed based on specified rules with the application of a special procedure of logical inference by expert system.

3 Principles of Designing the Modeling Shell of a Monitoring System

The foundation of a modeling shell is composed from the modeling sub-system and intellectual kernel, which has a structure of a standard static expert system. The modeling shell consists of the following elements (Fig. 2).

1. user interface;
2. database;
3. knowledge base; and
4. inference engine.

The user's interface (UI) contains a language processor for communication between the user and the computer. This communication can be organized by means of a

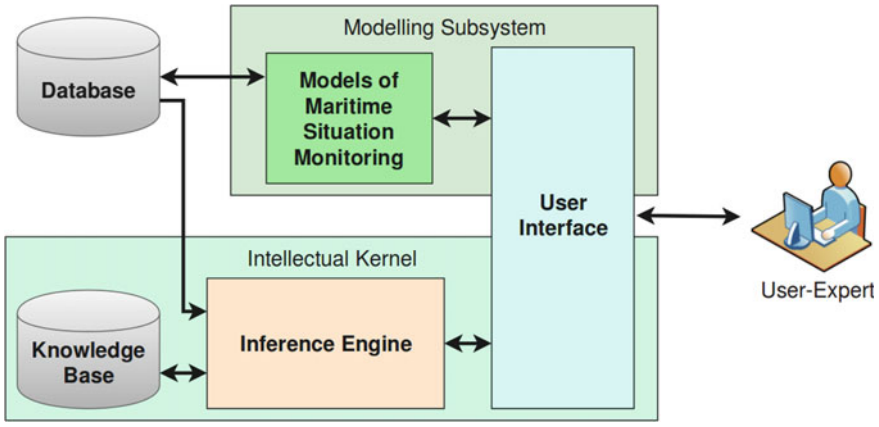


Fig. 2 Structure diagram of the modeling shell

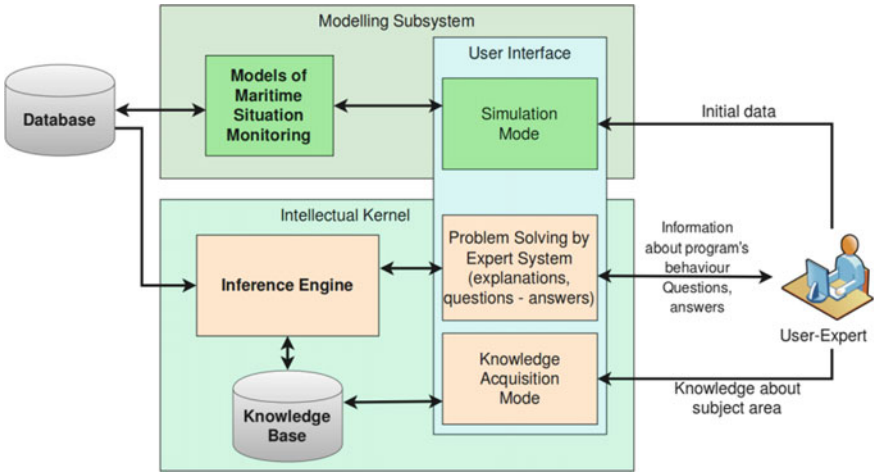


Fig. 3 Interaction scheme between the user-expert and the modeling shell

natural language and can be accompanied by graphics or a multi-window menu. Interaction between the user-expert and modeling shell is performed by choosing the operating modes of the modeling shell. The interaction scheme between the user-expert and the modeling shell is shown in Fig. 3.

The user's interface should provide three operating modes: (1) a knowledge acquisition mode, (2) a problem solving mode (explanation, questions-answers), and (3) a design, development, documentation, simulation mode. In the knowledge-acquisition mode, the process of filling the expert system with knowledge by the expert-user and the adaptation of the database to its functioning conditions are executed. The expert-system's adaptation to the changes in subject area is accomplished by replacement of the rules and facts in the knowledge base. In the problem-solving mode, the expert system acts as data storage for the expert-user and also allows to achieve a reasonable way to monitor the system's functioning and to explain the way of its achievement. For solving these tasks, there are the following elements in the user's interface: (1) a knowledge acquisition sub-system; (2) a dialogue sub-system, and (3) an explanatory sub-system. The user's interface functional diagram is shown in Fig. 4.

The database (DB) is meant for initial and intermediate storage of data obtained in the process of the functioning of the model and the means of environmental monitoring. The knowledge base (KB) is the basis of the intellectual core and acts as a storage of sets of facts and rules obtained from the experts for a given subject area. The inference engine (IE) is designed to obtain a reasonable sea environment monitoring method based on the comparison of the initial data from the database with knowledge from the knowledge base. It manipulates the knowledge base information and discovers the order in which to detect correlations and draw conclusions. The inference engine is used for the modeling of reasonings, processing of questions, and preparation of answers.

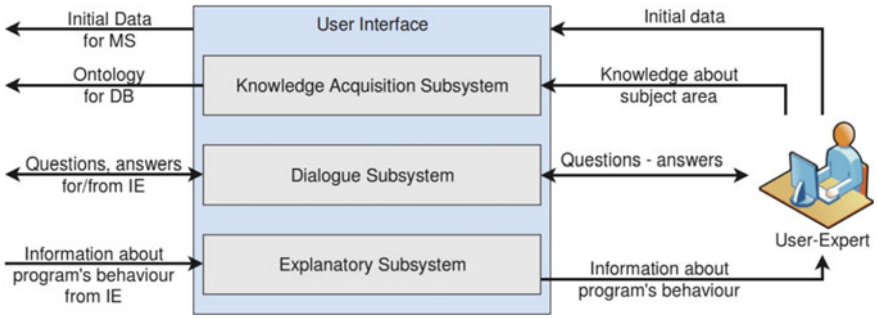


Fig. 4 The user's interface functional diagram

4 Principles of Constructing the Modeling Sub System

The modeling subsystem is intended for design, development, model documenting, and computer simulations. The foundation of the model is constructed according to the following principles [4, 5]:

- representation of a complex system using an object-oriented method;
- analysis and representation of the real process of monitoring system functioning in form of the integral group of results of above-water, air, and underwater environmental observations (current or final);
- analysis and representation of the real process of the monitoring system functioning in form of the element group (i.e., phenomena, utilization methods, subsystems and their elements realizing these phenomena, and sub-system management methods of the each environmental type under regard);
- necessary analysis of the conditions under which the monitoring system operates for each environmental type and during information integration on higher level;
- mathematical description based on logic-mathematical interpretation of a group of discovered primitive components and conditions with consideration of their space–time interdependency and interconnection as well as the informational consolidation of two levels according to environmental type and an integral one;
- representation of the model as a set of interacting functions in parallel activities (objects); and
- possibility of visualization that provides demonstrable acumen at creating both models and simulation levels (in the process of simulation modeling).

The following structure of the modeling subsystem is proposed:

1. graphic editor,
2. model-constructor,
3. compiler,

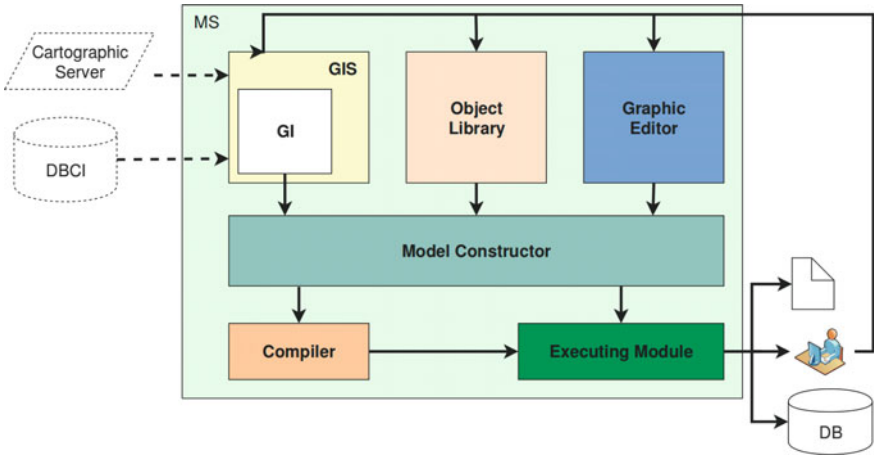


Fig. 5 Subsystem modeling structure

4. executing module,
5. geoinformation system, and
6. object library of subject domain.

The proposed structure of the modeling subsystem is shown in Fig. 5. Such a structure provides flexible adaptation of the simulation model to the corresponding subject area in the process of preparation or execution of the experiment.

- GIS (geoinformation system) is used as source of initial environmental information; for space–time integration of data that describes objects' behavior; and for visualization of the results of simulation modeling on a map [6]. Support of GIS space carries out object search opportunities, putting objects in new places, obtaining information about current location, moving objects from the current location into a new one with a given speed, execution of the certain actions on arrival, object animation representation (static or in motion), and detection of correlation between objects according to their location.
- The compiler performs translation of the machine program from an object-oriented language to a machine-oriented one.
- The model constructor is intended for the interactive modeling of complex dynamical systems. One of the conditions set for a model constructor is that it should be designed in terms of object-oriented technology and provide a comfortable, intuitive visual interface for users.
- The object library for the subject area should allow to save objects and provide the model's developers with access to a set of repeatedly used classes of agents and Java classes for a given subject area. The object library, for convenience, can be opened on the palette in user's interface.
- The graphic editor should allow the following:

(a) create and edit graphical primitives;
(b) create and edit experiment diagrams;
(c) draw a presentation for the experiment; the presentation should be displayed in the presentation window during the experiment start-up;
(d) create interactive elements of model's control to the presentation; and
(e) add data-obtaining elements to the presentation.

These experiments are performed by an executing module that displays animation during the modeling process. The experimental results are loaded into the database of the modeling shell as initial data for further processing in the inference engine. Therefore, the modeling sub-system provides an opportunity for the user-expert to develop models and retain them in code.

Practical application of the simulation model described in this chapter has shown capabilities of its application for the improvement of maritime situation monitoring systems. Experiments with application of the simulation model were performed in order to establish the appropriate directions of improvement of maritime situation monitoring systems for the White Sea and the Gulf of Finland.

5 Conclusion

In this chapter, an approach to the intellectualization of research of a complex, dynamical, distributed system, such as a maritime monitoring system, was described. The proposed approach is based on already implemented projects and is a logical continuation of the process of improving methods of maritime observation systems research. It is anticipated that application of the informational technologies, as described in this paper, will allow to considerably improve the researcher's opportunities to increase the quality of decisions made in the stages of designing and organizing such complex systems.

References

1. Sirota AA (2006) Computer modeling and effectiveness estimation of complex systems. Technosphera, Moscow
2. Volguin PN, Ivakin YA (2011) Simulation modelling and basic informational technologies for the monitoring and sea environmenatal control systems. In: J Scientific and technical conference «Integrated and automated control systems,» Ulyanovsk, FSPC OAS «SPS» «Mars» Report collection
3. Volgin NS (1999) Operation research. Part 1. St. Petersburg, Naval Academy
4. Shannon RE (1975) Systems Simulation the art and science. Prentice-Hall, Englewood Cliffs, NJ
5. Volguin PN (1999) Operations research. Part 1. Naval Academy, St. Petersburg
6. Volguin PN, Deveterikov VV (2014) Integrated simulation model of the monitoring system in the sea zone. In: Proceedings of the 7-th Russian control multiconference «Informational control technologies, SSC RF OAS «Concern «CSII» «Electropribor»»

Part V
Intelligent GIS for Land-Based Research

ST-PF: Spatio-Temporal Particle Filter for Floating-Car Data Pre-processing

Xiliang Liu, Li Yu, Kang Liu, Peng Peng, Shifen Cheng, Mengdi Liao and Feng Lu

Abstract Floating-car data (FCD) are never perfectly accurate due to various noises. To make FCD available, we propose a novel spatio-temporal particle filter ST-PF for FCD pre-processing. First we analyze the causes of errors and the shortcomings of previous studies. Second, we introduce the spatio-temporal constraints into the modeling of ST-PF. We also devise a novel iterating strategy for the recurrence of particle filtering based on sequential-importance sampling (SIS). We further design a series of experiments and compare the performances with that of other four traditional filters, namely, the mean filter, the median filter, the Kalman

X. Liu · L. Yu · K. Liu · P. Peng · S. Cheng · F. Lu (✉)
State Key Laboratory of Resources and Environmental Information System, IGSNRR, Chinese Academy of Sciences, Beijing, China
e-mail: luf@lreis.ac.cn

X. Liu
e-mail: liuxl@lreis.ac.cn

L. Yu
e-mail: yul@lreis.ac.cn

K. Liu
e-mail: liukang@lreis.ac.cn

P. Peng
e-mail: pengp@lreis.ac.cn

S. Cheng
e-mail: chengsf@lreis.ac.cn

X. Liu
Fujian Collaborative Innovation Center for Big Data Applications in Governments, Fuzhou, China

F. Lu
Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing, China

M. Liao (✉)
College of Geomatics, Shandong University of Science and Technology, Qingdao, China
e-mail: 93710@163.com

V. Popovich et al. (eds.), *Information Fusion and Intelligent Geographic Information Systems (IF&IGIS'17)*, Lecture Notes in Geoinformation and Cartography, DOI 10.1007/978-3-319-59539-9_15

filter, and the original particle filter. The final results show ST-PF is much more effective for noise reduction and improvement of map-matching performance and shows a promising direction for FCD pre-processing.

Keywords Particle filter · Data pre-processing · Geo-spatial · Spatio-temporal

1 Introduction

During the past few years, global positioning system (GPS) devices have become a major positioning technology for providing location data [1]. Currently real-time floating-car data (FCD) of city road networks collected by operating vehicles (taxicabs, probe cars, buses, private cars, etc.) equipped with GPS-enabled devices has become mainstream in the research of intelligent-transport system (ITS) and many other GIS-related research and services (LBS) because of its cost-effectiveness and flexibility compared with other traffic-data sources [2]. The prevalence of FCD has made the collection of a huge number of GPS trajectories possible, thus providing a promising opportunity to upgrade the service level amongst a list of trajectory-based applications.

However, the raw FCD cannot be used directly. One reason lies in the relatively low spatial-positioning accuracy of consumer GPS devices, which varies from 15 to 30 m [3]. This situation is even worse due to the blocking by tall buildings, trees, and channels [4] where the noisy GPS observations are always drifting from the actual vehicle trajectories on the road networks. The other one is rooted in the fault of electric-communication devices along with the constraints of transmission bandwidth and energy consumption [5]. These two reasons degrade the quality of FCD, especially in the context of the low sampling rate of GPS devices. Therefore, the observed FCD must be pre-processed before use [4–6].

Many pre-processing methods have been proposed for filtering the raw errors and noises of the original FCD. These methods can generally be classified as statistical methods and state-space–based methods [7]. The traditional statistical methods involve techniques such as the mean filter and the median filter [4]. These methods first model the GPS observations of the FCD as a linear system and then employ a sliding window to smooth the original data. Although effective in processing speed, these methods are only suitable for smooth distributions and cannot deal with lots of fluctuations (especially during peak hours) and outliers. Lags are also introduced into the final results [4]. To alleviate these problems, state-space–based methods, such as the Kalman filter and the particle filter, are proposed to incorporate more influencing factors (*e.g.*, speed, direction, etc.) into FCD-system modeling [7–9]. The Kalman filter (KF) and its many variants and generalizations have played a fundamental role in modern time-series analysis by allowing the study and estimation of complex dynamics and by drawing the attention of

researchers and practitioners to the rich class of state-space models (*i.e.*, the dynamic models). Compared with the mean and median filters, the Kalman filter models the state space of the original GPS observations as a stochastic system according to state equation and observation equation based on Bayesian inference and estimates a joint probability distribution over the GPS observations for each time frame. However, the Kalman filter still calls for the Gaussian distribution of the original data without many outliers. The result of raw-error filtering based on the Kalman filter cannot be guaranteed especially in the transportation domain [10]. Because the calculation of the Kalman filter require the support of all of the previous steps, some researchers argue that time lags exist in on-line applications [11].

To deal with the nonlinear non-Gaussian filtering problem, one popular solution is to use sequential Monte Carlo methods (SMC), also known as "particle filters" [12]. The particle filters consist of estimating the internal states in dynamical systems when partial observations are made and random perturbations are present in the GPS sensors as well as in the dynamic-observation system. The objective of the particle filter is to compute the conditional probability (a.k.a. posterior distributions) of the GPS observations given that some noisy and partial GPS observations and outliers existed in the raw errors. Compared with the Kalman filter, a particle filter does not make a priori assumption that the noises in the original GPS observations follow the Gaussian distribution. The particle-filter method exhibits a wider tolerance for outliers in the raw errors, thus making it more suitable for pre-processing of the observed GPS sequences under various stochastic traffic scenarios.

In this chapter, based on the particle-filter method, we propose a novel filter method called the spatio-temporal particle filter (ST-PF) for GPS observations pre-processing. In the modeling of ST-PF, we consider the spatio-temporal continuity of the original GPS observations. In the practical processing of the FCD, we find that the driving speed of a given floating car fluctuates within a controllable interval. In addition, we also find that the driving direction of a floating car with GPS devices does not change dramatically in a given time slot, thus indicating that the deviation of driving directions should also be added into the modeling of spatio-temporal continuity. With spatio-temporal constraints (*i.e.*, the speed and directional deviations), we devise a novel iterating strategy for the recurrence of particle filtering based on sequential-importance sampling (SIS). We further design a series of experiments based on real FCD collected in Beijing and compare the performances with that of four other traditional filters, namely, the mean filter, the median filter, the Kalman filter, and the original particle filter. The final results show that the ST-PF method not only haves better performance and robustness than the other popular models, but it also holds a competitive time complexity in the real-time processing of mass FCD.

The rest of this chapter is organized as follows. Section 2 describes the problem confronted in the pre-processing of FCD and formulizes the preliminaries in this chapter. Section 3 discusses the ST-PF model in detail including the modeling of

driving speeds and directional changes, sequential importance sampling (SIS) of ST-PF, and iterating-strategy implementation. Section 4 repots the experiments with real personal and industrial FCD from Beijing. Section 5 discusses the modeling of ST-PF and draws conclusions.

2 Preliminaries and Problem Statement

2.1 Preliminaries

Definition: Road network

A road network is a directed graph $R = < V, E >$ where V is a set of road vertices; and E is a set of road segments. A road vertex is a network node that corresponds to longitude, latitude, and road-segment index. A road segment is a directed edge that is associated with two terminal vertices, length and a direction symbol.

Definition: Actual GPS position

The actual GPS position of a given floating car along the road network R is denoted as $X_i = (x_i, y_i)^T$. The index i represents time increments with $i = 1, 2, \ldots, N$. The actual GPS position X_i is a two-element vector representing the real coordinates at time stamp i. Generally speaking, the actual GPS position is unknown due to sensor noises and raw errors.

Definition: GPS observation

A GPS observation Y_i is a GPS-observation sequence lined by the timestamps of the actual GPS position X_i. Here $i = 1, 2, \ldots, N$.

Definition: GPS trajectory

A GPS trajectory Tr is a GPS-observation sequence lined by the timestamps of the GPS observations. Here $Tr = \{Y_i\}_{i=1,2,\ldots,N}$.

Definition: Actual GPS sequence

According to a given GPS trajectory Tr, the actual GPS sequence along the road network R is denoted as $Tr = \{Y_i\}_{i=1,2,\ldots,N}$.

Definition: Spatial-positioning error

Due to the sensor noises and raw errors in the measurements, the actual GPS sequence X is always not exact to the observed GPS trajectory Tr. The spatial-positioning error v_i, between an actual GPS position X_i and its corresponding GPS observation Y_i, is usually assumed to be drawn from a two-dimensional Gaussian probability density with zero mean and diagonal covariance matrix E, *i.e.*,

$$v_i \approx N(0, E), i = 1, 2, \ldots, N \quad (1)$$

For GPS, the Gaussian noise model above is reasonable [13].With the diagonal-covariance matrix, this is the same as adding random noise from two different one-dimensional Gaussian densities, x_i and y_i, separately, each with a zero mean and SD σ.

2.2 Problem Statement

Except for the raw errors in the GPS observations, there are mainly two problems to be solved during the pre-processing stage of the FCD, namely, the *floating* problem and the *bouncing* problem of the GPS signals.

The floating problem refers to that when a floating car stops or runs at a low speed, the GPS signal of the floating car will float randomly within a certain error circle. This problem is modeled as the spatial-positioning error as mentioned in Sect. 2.1. The spatial-positioning error may behave heterogeneously at different parts of the city's road network [4]. Moreover, the spatial-positioning error also varies at different time periods (*e.g.*, in the daytime and in the night). For simplicity, here we model the spatial-positioning error with the same SD σ.

The bouncing problem of the GPS signals also influences the spatio-temporal continuity of the GPS observations. During the operation of a floating car, the speed and the driving direction of the floating car at time stamp *i* do not match the speed and the driving direction of the floating car at time stamp *i-1*. It's often observed that the current GPS position is "bounced" to the previous position, thus leading to spatio-temporal inconsistencies of the GPS observations. The two problems are illustrated in Fig. 1.

In Fig. 1, the cyan line represents a real GPS trajectory in our floating-car data set in Beijing. The floating car stopped at an intersection for a while in the morning peak hour. The black circles marked with 1, 2,…, 5 are GPS observations during this time slot, and the red dotted circle is the spatial positioning error with a radius as the SD. The actual position of the floating car is at black circle 1. Black circles 2 and 3 are the results of the floating problem. Black circle 4 is caused by the bouncing problem, which results in a wrong temporal sequence and a greater spatial positioning error. Black circle 5 is the raw error in the GPS observation. The task for FCD pre-processing is to remove as many raw errors, such as black circle 5, as possible and decrease the consequences of these two problems as much as possible. The FCD preprocessing is non-trivial. The results of the FCD pre-processing directly influence the quality of various subsequent analysis tasks. How to deal with such problems has become the challenge in FCD pre-processing.

Fig. 1 Problems confronted in the pre-processing of FCD

3 The Formulation of ST-PF

3.1 State-Space Modeling of a Floating-Car System

The state space of the floating car system consists of two separate parts: (1) the system model, which refers to modeling the consequence of the actual GPS sequence X; and (2) the observation model, which describes the relationship between the GPS observation and the actual GPS position. The state-space model based on a particle filter is assumed to follow the first-order Markov process [8], thus indicating that the following:

1. The elements in the GPS trajectory $\{Y_1, \ldots Y_i\}$ are independent of each other; and
2. the current actual GPS position X_i is only related to the former position X_{i-1} and is conditionally independent of the previous actual GPS sequence $\{X_1, \ldots, X_{i-2}\}$.

According to this consumption, the system model in a particle filter is modeled as a linear system $X_i = f(X_{i-1}, u_{i-1})$ where f is the transformation function; and u stands for the external factors in the location process. The observation model is formulated

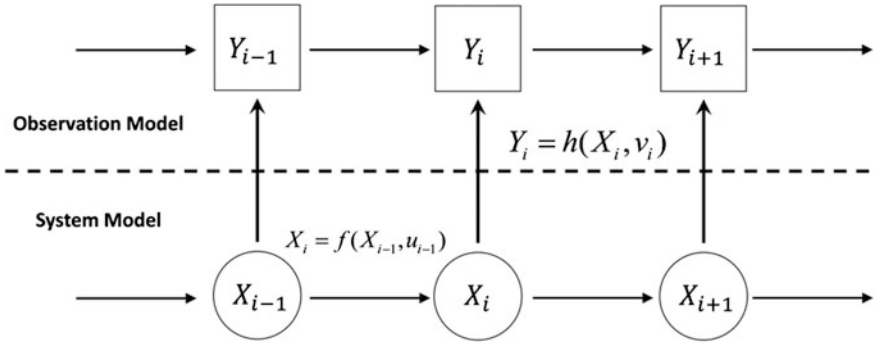


Fig. 2 The framework of state-space modeling

as non-linear system $Y_i = h(X_i, v_i)$ where function h stands for the corresponding relationship between Y_i and X_i. Figure 2 shows the process of state-space modeling.

Based on state-space modeling, the pre-processing task of FCD can be expressed as follows: given the GPS trajectory $\{Y_1, \ldots, Y_i\}$ in the context of raw error and spatial-positioning error, estimate the maximum likelihood of the conditional probability $p(X_i|Y_{1:i})$ with the help of the initial value X_0 and corresponding external factors u. In view of the first-order Markov process consumption, this objective function can be rewritten as follows:

$$p(X_i|Y_{1:i}) = \frac{p(Y_i|X_i)p(X_i|Y_{1:i-1})}{p(Y_i|Y_{1:i-1})} \tag{2}$$

Here $p(X_i|Y_{1:i-1})$ stands for the state-transition probability-density function of the state-space model, and $p(Y_i|X_i)$ is the likelihood probability-density function of the observation model. In the modeling of a particle filter, $p(X_i|Y_{1:i-1})$ is the predicting part, and $p(X_i|Y_{1:i})$ is the update of $p(X_i|Y_{1:i-1})$. For a linear system, $p(X_i|Y_{1:i})$ and $p(X_i|Y_{1:i-1})$ can be easily solved by their integral forms (the integral form of $p(X_i|Y_{1:i})$ is $p(X_i|Y_{1:i}) = \int p(X_{0:i}|Y_{1:i})dX_{0:i-1}$). However, the FCD system is non-linear. How to approximate $p(X_i|Y_{1:i})$ under conditions of raw errors and external factors is discussed in Sect. 3.2.

3.2 The Sequential-Importance Sampling Procedure

As mentioned in Sect. 3.1, the FCD system is non-linear. Traditional approaches employ Monte-Carlo simulation (MCS). First, K samples $\{X^{(j)}\}$ that satisfy the *i.i.d* condition are randomly selected from $p(X_i|Y_{1:i})$, where $j = 1, 2, \ldots, K$. Each sample is called as a "particle." The posterior-probability density $p(X_i|Y_{1:i})$ then can be rewritten as follows:

$$p_K(X|Y) = \frac{1}{K}\sum_{j=1}^{K}\delta\left(X - X^{(j)}\right) \tag{3}$$

where δ is the Dirac delta function. When K increases, the expectation of the observation model $Y_i = h(X_i, v_i)$ gravitates toward its exact value asymptotically. However, the MCS does not work well under the condition of a non-linear system [7, 10, 11]. We employ the sequential-importance sampling (SIS) strategy to solve this problem according to [14].

The SIS strategy introduces $\tilde{p}(X_i|Y_{1:i})$ and $q(X_i|Y_{1:i})$ into the calculation of the expectation $E(h)$:

$$E(h) = \frac{1}{\int q(X|Y)\omega(X|Y)dX}\int h(X,u)q(X|Y)\omega(X|Y)dX \tag{4}$$

where $\tilde{p}(X_i|Y_{1:i})$ is the unnormalized version of $p(X_i|Y_{1:i})$; $q(X_i|Y_{1:i})$ is the proposal-probability density of $p(X_i|Y_{1:i})$; and ω is the weight vector. For Eq. 4, the MCS with K samples then can be employed to calculate the approximate value of $E(h)$ as follows:

$$\hat{E}_K(h) = \frac{1}{\int q_K(X|Y)\omega(X|Y)dX}\int h(X,u)q_N(X|Y)\omega(X|Y)dX = \sum_{j=1}^{K} h\left(X^{(j)},u\right)\omega^{(j)} \tag{5}$$

According to Eqs. 4 and 5, the value of $p(X_i|Y_{1:i})$ can be approximated as follows:

$$\hat{p}_K(X|Y) = \frac{1}{K}\sum_{j=1}^{K}\delta\left(X - X^{(j)}\right)\omega\left(X^{(j)}\right) \tag{6}$$

To obtain the value of ω, the SIS strategy also models the proposal-probability density $p(X_i|Y_{1:i})$ as the first-order Markov process. The ω value of the jth particle at time stamp i then can be expressed iteratively as shown in Eq. 7.

$$\tilde{\omega}_i^{(j)} = \tilde{\omega}_{i-1}^{(j)}\frac{p\left(Y_i\middle|X_i^{(j)}\right)p\left(X_i^{(j)}\middle|X_{i-1}^{(j)}\right)}{q\left(X_i^{(j)}\middle|X_{i-1}^{(j)},Y_{1:i}\right)} \tag{7}$$

Based on Eqs. 3, 5, and 7, the posterior-probability density $p(X_i|Y_{1:i})$ of the predicting part in the state-space model then can be approximately formulated as follows:

$$\hat{p}_K(X_i|Y_{1:i}) = \frac{1}{K}\sum_{j=1}^{K}\delta\left(X_i - X_i^{(j)}\right)\omega_i^{(j)} \tag{8}$$

where $\omega_i^{(j)}$ is the normalized value:

$$\omega_i^{(j)} = \frac{\hat{\omega}_i^{(j)}}{\sum_{j=1}^{K}\hat{\omega}_i^{(j)}} \tag{9}$$

The whole iterative procedure of SIS is as follows:

Input:

The posterior-probability density at time stamp $i-1$ with K sampling is as follows:

$$p(X_{i-1}|Y_{1:i-1}) \approx \left\{X_{i-1}^{(j)}, \omega_{i-1}^{j}\right\}, j = 1, 2, \ldots, K$$

The initial values of actual GPS position and GPS observation are as follows:

$$\{X_0, Y_0\}$$

Output:

The posterior-probability density at time stamp i:

$$p(X_i|Y_{1:i}), i = 1, 2, \ldots, N$$

Step:

1. For the current particle $j = 1, 2, \ldots, K$
   (1) Monte-Carlo simulation: $X_i^{(j)} \sim q\left(X_i \middle| X_{0:i-1}^{(j)}, Y_{1:i}\right)$
   (2) Calculate the unnormalized iterative value of ω

$$\tilde{\omega}_i^{(j)} = \tilde{\omega}_i^{(j)}\frac{p\left(Y_i\middle|X_i^{(j)}\right)p\left(X_i^{(j)}\middle|X_{i-1}^{(j)}\right)}{q\left(X_i^{(j)}\middle|X_{i-1}^{(j)}, Y_{1:i}\right)}$$

2. Normalize ω

$$\omega_i^{(j)} = \frac{\hat{\omega}_i^{(j)}}{\sum_{j=1}^{K} \hat{\omega}_i^{(j)}}$$

3. Calculate the approximate value of $p(X_i|Y_{1:i})$ at time stamp i:

$$p(X_i|Y_{1:i}) \approx \left\{X_i^{(j)}, \omega_i^j\right\}, j = 1, 2, \ldots, K$$

After the entire iterative procedure of SIS, each GPS observation is extended with K samplings, and the weights for all of the K samplings are also calculated. In our experiments, we only consider the sampling with the highest weight as the corresponding GPS position of the given GPS observation.

3.3 The Modeling of ST-PF

Traditional particle filters only consider the influence of speed in the system model of the original space state [4, 10, 14], whereas if neglects other influence factors that affect the spatio-temporal continuity of the FCD system. In the practical experiments, we found that the deviation in driving directions of the consecutive GPS observations also plays an important role. The system model is expressed as follows:

$$X_i = \left(x_i, y_i, s_i^{(x)}, s_i^{(y)}, A_i^{(x)}, A_i^{(y)}\right)^T \tag{10}$$

where $A_i^{(x)}$, $A_i^{(y)}$ stand for the directional deviations in longitude and latitude; and $s_i^{(x)}$, $s_i^{(y)}$ stand for the speed deviations both in longitude and latitude.

The directional deviations is modeled as the Gaussian distribution with a zero mean and a variance of σ_A^2. In the same way, the speed deviation is expressed by way of the Gaussian distribution with a zero mean and a variance of σ_s^2. Given the initial values $\{X_0, Y_0\}$, the iterative K samplings in the SIS is expressed as follows:

$$\begin{cases} X_{i+1} = X_i + v_i^{(x)} \Delta t_i \\ Y_{i+1} = Y_i + v_i^{(y)} \Delta t_i \\ s_{i+1}^{(x)} = s_i^{(x)} + \beta_i^{(x)}, \beta_i^{(x)} \sim N\left(0, \sigma_s^2\right) \\ s_{i+1}^{(y)} = s_i^{(y)} + \beta_i^{(y)}, \beta_i^{(y)} \sim N\left(0, \sigma_s^2\right) \\ A_{i+1}^{(x)} = A_i^{(x)} + \alpha_i^{(x)}, \alpha_i^{(x)} \sim N\left(0, \sigma_A^2\right) \\ A_{i+1}^{(y)} = A_i^{(y)} + \alpha_i^{(y)}, \alpha_i^{(y)} \sim N\left(0, \sigma_A^2\right) \end{cases} \tag{11}$$

where α, β, are the longitude and latitude disturbances for directional and speed deviations, respectively. Consequently, the initial diagonal covariance matrix E_0 is defined as follows:

$$E_0 = \begin{bmatrix} \sigma^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_s^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_s^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_A^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_A^2 \end{bmatrix} \tag{12}$$

With the help of $\{X_0, Y_0\}$ and E_0, all of the GPS trajectories in FCD are pre-processed successfully by way of the SIS procedure mentioned in Sect. 3.2. The FCD pre-processing work is performed after the SIS procedure.

4 Experiments

4.1 Study Area and Data-Set Description

The study area covers Beijing's main urban area. The road network of Beijing contains 14,614 vertices and 26,621 segments. Its vertical length is approximately 43.6 km, and the horizontal length is approximately 41.7 km.

Beijing's FCD is a human-labeled data set from 2013.1.7 to 2014.7.10 containing both the original GPS observations and the corresponding mapped results on the road network. The total number of trajectories is 462 including approximately 350,000 GPS observations. The sampling rate is 1 s, and the average GPS positioning accuracy is 15 m.

4.2 Baselines and Evaluation Metrics

Based on the literature review, we selected four different filters as our baselines, namely, the mean filter, the median filter, the Kalman filter [4], and the traditional particle filter [4, 10]. The Kalman filter and the traditional particle filter only consider the influence of speed in space-state modeling.

According to travel experiences in Beijing, the speed disturbance β is set as 5 m/s, and the directional disturbance α is 30 degree/s. The sliding window size is set as 5 for the mean filter and the median filter. For the Kalman filter and the traditional particle filter, we set the number of particles K as 1000 in accordance with previous studies [4, 10].

With the help of the original GPS observations and the corresponding mapped results in Beijing's FCD, we choose the noisy-data ratio (NDR) as an evaluation

metric for comparison of the pre-processing performance. The term "noisy data" refers to data with raw errors, a floating problem, or a bouncing problem. The NDR is calculated as follows:

$$NDR = \frac{the\ number\ of\ noisy\ data}{the\ number\ of\ all\ the\ GPS\ observations} \times 100\% \tag{13}$$

The other evaluation metric we employ in our experiments is the point-correct matching percentage (PCMP), which illustrates the map-matching performances both with and without the FCD pre-processing based on different filters. The PCMP is expressed as follows:

$$PCMP = \frac{correct\ matching\ points}{number\ of\ points\ to\ be\ matching} \times 100\% \tag{14}$$

4.3 Experiments Results

All of the experiments were coded in Python 2.7 and run on a 64-bit Windows operating system equipped with eight Intel(R) Core(TM) i7-4720HQ processors clocked at 2.6 GHz. The RAM of the computing environment is 16 G.

To test the performances of the proposed ST-PF method at different temporal resolutions, we down-sampled the original FCD at different sampling rates, namely 5, 10, 15, 30, 60, 90, and 120 s. To guarantee the reliability, all of the experiments were repeated 10 times, and the average values were regarded as the final outputs.

Noise-Reduction Comparison

The original FCD is not perfect due to various errors. We first calculated the NDR for each down-sampled FCD, and then we performed a comparison of the noise-reduction ability of different filters. The final results are listed in Table 1.

Table 1 NDR results of all of the filters

Time (s)	ORI*	Mean	Median	Kalman	PF*	ST-PF
1	0.2847	0.2642	0.2537	0.1504	0.1643	0.1056
5	0.2549	0.2202	0.2325	0.1742	0.1555	0.0982
10	0.2434	0.1844	0.1798	0.1334	0.1249	0.0754
15	0.2212	0.1628	0.1555	0.1223	0.1051	0.0682
30	0.2096	0.1523	0.1232	0.1055	0.1129	0.0512
60	0.1588	0.1020	0.0933	0.0872	0.0832	0.0032
90	0.1211	0.0810	0.0823	0.0556	0.0524	0.0000
120	0.0960	0.0544	0.0342	0.0012	0.0008	0.0000

*ORI = NDR of the sampled FCD; *PF = traditional particle filter

It is clear the proposed ST-PF performs the best under all of the time resolutions. Moreover, ST-PF is much more effective for FCD with a low sampling rate (*e.g.*, 60 s, 90 s, 120 s). These findings show the advantage of the proposed ST-PF in terms of noise reduction for FCD pre-processing.

Time-Complexity Comparison

For each filter, we compared the running time to process the original FCD. The result is shown in Fig. 3.

Figure 3 indicates that along with more influencing factors adding up to the space-state modeling, the time complexity increases accordingly. Although the proposed ST-PF holds the maximum running time compared with other filters, the time complexity of ST-PF is still acceptable for practical applications (350,000 + points/min).

Map Matching–Performance Comparison

In this experiment, we test the ability of the proposed ST-PF for the improvement of map-matching performance. The noise-reduction ability of ST-PF has been proven. So here we employ only ST-PF for the FCD pre-processing. We compare the map-matching performances under different time resolutions with and without the ST-PF pre-processing. For a specific time resolution, we first matched the sampled FCD without pre-processing, and then we processed the sampled FCD again with ST-PF. We employed ST-CRF [5] for all of the map-matching calculations. The result of map matching–performance comparison is shown in Fig. 4.

As shown in Fig. 4, all of the map-matching performances are increased with the help of ST-PF pre-processing. The maximum improvement of PCMP value is 5.33% at a time resolution of 90 s. Figure 4 shows the ability of ST-PF for map-matching. It also shows the necessity of FCD pre-processing for map-matching performance.

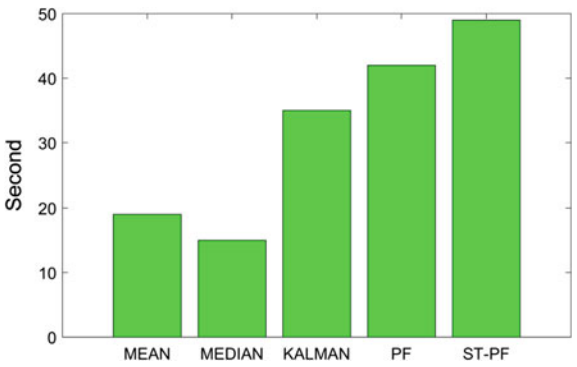


Fig. 3 The result of the time-complexity comparison

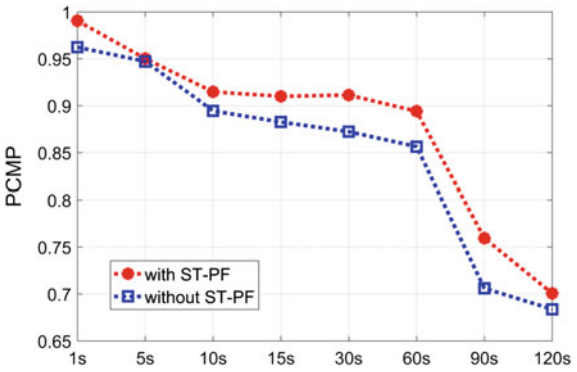


Fig. 4 The result of map matching–performance comparison

5 Conclusions

Floating-car data pre-processing is a non-trivial task meriting further analyses. In this chapter, we propose a novel particle-filter method called "ST-PF" to solve the frequently encountered problems (*i.e.*, raw errors, floating problem, and bouncing problem) during FCD pre-processing. We consider both the speed and directional deviations in the modeling of ST-PF and adjust the sequential importance sampling procedure accordingly. A series of experiments based on Beijing's road network and FCD are performed including noise reduction, time complexity, and improvement of map-matching performance. The final results show that although the time complexity of ST-PF is greater than that of the other four baselines (*i.e.*, mean filter, median filter, Kalman filter, traditional particle filter), it is still acceptable for practical applications (350,000 + points/min). Moreover, ST-PF is much more effective for noise reduction and improvement of map-matching performance, thus showing a promising direction for FCD pre-processing. In further work, we will investigate the acceleration of ST-PF and open-source ST-PF as a general package for FCD pre-processing.

Acknowledgements This work was supported in part by the National Natural Science Foundation of China under Grants No. 41601421 and 41271408 and in part by the China Postdoctoral Science Foundation under Grant No. 2015M581158.

References

1. Xiliang L, Feng L, Hengcai Z, Peiyuan Q (2013) Intersection delay estimation from floating car data by way of principal curves: a case study on Beijing's road network. Front Earth Sci 7 (2):206–216
2. Hofleitner A, Herring R, Abbeel P, Bayen A (2012) Learning the dynamics of arterial traffic from probe data using a dynamic Bayesian network. IEEE Trans Intell Transp Syst 13 (4):1679–1693

3. Cui Y, Ge S (2003) Autonomous vehicle positioning with GPS in urban canyon environments. IEEE Trans Robot Autom 19(1):15–25
4. Zheng Y, Zhou X (2011) Computing with spatial trajectories. Springer-Verlag, New York, pp 1–10
5. Xiliang L, Kang L, Mingxiao L, Feng L (2016) A ST-CRF map-matching method for low-frequency floating car data. IEEE Trans Intell Transp Syst 99:1–14
6. Quddus MA, Ochieng WY, Noland RB (2007) Current map matching algorithm for transport applications: State-of-the art and future research direction. Transport Res C-Emerg 15: 312–328
7. Liangquan L, Weixin X, Zongxiang L (2016) A novel quadrature particle filtering based on fuzzy c-means clustering. Knowl Based Syst 106:105–115
8. Hedibert FL, Ruey ST (2011) Particle filters and Bayesian inference in financial econometrics. J Forecast 30:168–209
9. Gustafsson F, Hendeby G (2012) Some relations between extended and unscented kalman filters. IEEE Trans Signal Proc 60(2):545–555
10. Lyudmila M, Andreas H, Amadou G, René KB (2012) Parallelized particle and Gaussian sum particle filters for large-scale freeway traffic systems. IEEE Trans Intell Transp Syst 13(1): 36–48
11. Hossein TN, Akihiro T, Seiichi M, David MA (2012) On-road multivehicle tracking using deformable object model and particle filter with improved likelihood estimation. IEEE Trans Intell Transp Syst 13(2):748–758
12. Carmia AY, Mihaylova L, Septierc F (2016) Subgradient-based Markov Chain Monte Carlo particle methods for discrete-time nonlinear filtering. Signal Proc 120(3):532–536
13. Nabil MD, Haitham MA, Otman AB (2013) GPS localization accuracy classification: A context-based approach. IEEE Trans Intell Transp Syst 14(1):262–273
14. Blanco JL, Gonzalez J, Fernandez-Madrigal JA (2010) Optimal Filtering for non-parametric observation models: Applications to localization and SLAM. Int J Robot Res 29(14): 1726–1742

A Framework for Emergency-Evacuation Planning Using GIS and DSS

Reham Ebada Mohamed, Essam Kosba, Khaled Mahar and Saleh Mesbah

Abstract There has been a growing need for the use of information and decision-making systems in evacuation planning as a part of emergency management in order to reduce as, many losses as possible. To minimize damage, an accurate and effective evacuation plan that gives more than one evacuation path considering the changing road conditions in minimal time is imperative. The use of Geographic Information Systems (GIS), Decision-Support Systems (DSS), and shortest-path algorithms as a solution for this problem is the subject of this chapter. A framework for providing preparedness and response plans after an emergency event occurs is proposed. The plan provided by the proposed framework includes a calculated degree of hazard posed by that event (three emergency models are incorporated), the radius of affected area (a spatial overlay to draw a buffer zone is conducted), identification of all safe destinations, and the best alternative paths for evacuation from inside the buffer zone to safe destinations based on the dynamic road conditions displayed on the map. To identify all of the safe destinations and obtain the best alternatives, a graph theory–based model is proposed based on a developed algorithm to get everyone the closest safe destinations. Dijkstra's algorithm is executed from a single source inside the buffer to all identified safe destinations resulting in the minimum travel time path and other alternative paths displayed on the map. To evaluate the proposed framework, a GIS-based evacuation-planning (G-BEP) prototype is implemented. The prototype is tested with different emergency types, different variables, and different street maps. The prototype is also found to respond differently based on the dynamic road conditions.

R.E. Mohamed (✉) · E. Kosba · K. Mahar · S. Mesbah
College of Computing and Information Technology, Arab Academy for Science, Technology and Maritime Transport (AASTMT), Alexandria, Egypt
e-mail: reham.ebada@gmail.com

E. Kosba
e-mail: ekosba@aast.edu

K. Mahar
e-mail: khmahar@aast.edu

S. Mesbah
e-mail: saleh.mesbah@gmail.com

V. Popovich et al. (eds.), *Information Fusion and Intelligent Geographic Information Systems (IF&IGIS'17)*, Lecture Notes in Geoinformation and Cartography, DOI 10.1007/978-3-319-59539-9_16

Keywords GIS-based DSS · Evacuation planning · Emergency models · Dijkstra's shortest path · Evacuation alternative paths · Dynamic road conditions · Safe destinations

1 Introduction

An emergency event is uncertain, sudden, and complex for analysis. Emergencies are either caused by a natural disaster or a deliberate act. Natural disasters include earthquakes, floods, fires, tsunamis, tornadoes, hurricanes, volcanic eruptions, or landslides. Deliberate acts include explosions, chemical spills, radiation, terrorism attacks, and others.

In recent years, the world has experienced many crises related to disaster and emergency occurrences. Those crises usually cause a great loss of population and wealth. It has become imperative to establish a scientific and effective plan to manage the disaster and to reduce losses as much as possible. Such plans are the fundamental purpose of emergency management in order to evacuate as many victims as possible in the shortest time. The essential mapping of the spatial relationships between natural-hazard phenomena (earthquakes, fires, cyclones, etc.) and the elements at risk (people, buildings, and infrastructure) require the use of tools such as GIS. The most significant relationships in risk analysis and modeling are largely spatial, and although GIS did play an earlier role, but the plans were relatively primitive, rarely commercially available, and often experimental or research-focused rather than operational.

The use of DSS would improve the efficiency and effectiveness of evacuation. The focus on emergency management, especially the evacuation-planning field, is to prevent disasters—or mitigate them when they occur—as quickly as possible. Accurate and reliable information and spatial data on disaster, as well as how to quickly deal with the statistical summary and analysis, requires efficiency and effectiveness; thus, the use of DSS is required.

This chapter is an extended work of a previously published paper by the same authors [1]. This work focuses on using GIS for obtaining information about the emergency event, calculating its danger rating, performing buffer analysis and overlay, identifying the closest safe destinations, and solving the shortest path (based on Dijkstra's model) to obtain the best evacuation path and other alternatives based on travel time. Despite the abundance of research in this area, most research papers calculate the best optimal path to our human knowledge. Compared with related work presented in this chapter, one of its contributions is presented as the second-best and third-best alternative paths rather than calculating one optimal evacuation path.

The reminder of this chapter is organized as follows: related work is presented in Sect. 2, which covers related studies of evacuation planning using GIS, DSS, and graph-theory algorithms. Section 3 covers the methodology showing the proposed

framework, the parameters used to select the best paths, the developed algorithms, and the emergency models and equations. Section 4 presents the implementation of a G-BEP prototype and the results by covering the functionality of buffer analysis and overlay as well as evacuation-path generation. Finally, the conclusion and future work are presented in Sect. 5.

2 Literature Review and Related Work

The value of GIS in emergency management arises directly from the benefits of integrating a technology designed to support spatial decision-making into a field with a strong need to address numerous critical spatial decisions [2]. An important step in examining the role of GIS in emergency management is selecting a conceptual framework to help organize the existing research-and-development activities. One such framework is comprehensive emergency management (CEM) [3]. The use of GIS in providing the platform for CEM is discussed in [4].

CEM activities can be grouped into four phases; however, in examining the GIS literature, it is more appropriate to reduce them into three phases: (1) mitigation, (2) preparedness and response, and (3) recovery. This is because many GIS developed in the preparedness phase are used in the response phase [2]. Challenges for GIS in emergency preparedness and response are presented in [5].

Studies related to emergency response based on GIS, computational-simulation models, mathematical-assessment models, and dynamic algorithm to obtain the optimal solution are presented in [6–9]. Other studies focus not only on natural phenomena but also on traffic accidents [10, 11].

Some studies have focused based on GIS and DSS for evacuating large-scale buildings or public squares [12, 13]. Other studies focus on real-time evacuation systems in open areas based on GIS and DSS such as the Blue-Arrow system [14] and the emergency decision-making system developed in [15]. Shortcomings of the Blue-Arrow system are that target shelters should be pre-defined in the system because the radius of affected area might not be known to the user, and the system does not consider road conditions. The shortcoming in [15] is that it does not provide other alternatives to the evacuation path.

An improved Dijkstra's algorithm applicable to the GIS drilling accident–emergency rescue system is proposed in [16]. The shortcoming of the system is that the search for the evacuation path is based on defining a source and a single destination node. A dynamic vehicle routing during a peak-hour traffic system is implemented in [17] using Krushkal's algorithm to obtain the best route based on different parameters, but its shortcoming is that it only generates the optimum path.

Such studies contributed as a motivation to contemplate this research to overcome the mentioned shortcomings by integrating some graph-theory algorithms with GIS and DSS to compute the best alternative evacuation paths.

3 Methodology

This section presents an explanation of the proposed framework including a description of each component and the chosen algorithms. Additionally, it discusses how the framework adopts both the previously mentioned issues that must be considered regarding the development and limitations of previous work. Moreover, it covers the background of the used emergency models.

3.1 G-BEP: The Proposed Framework

G-BEP is proposed and developed in this chapter. It overcomes the shortcomings of previous research by the calculating radius of affected area and providing preparedness and response plans after the incident has occurred. Along with its degree of hazard, it identifies all safe destinations (algorithm), obtains alternative evacuation paths (not just the best one), and considers dynamic road-condition data that contribute as parameters for the selection of the best alternative paths. Figure 1 shows the main framework components of G-BEP. Note that developed components appear in the shaded area.

G-BEP is based on using GIS for obtaining information about the emergency event, calculating its danger rating, and performing buffer analysis and overlay on a street map in a Geo-Database. Three emergency types are represented: a fire event, a hurricane event, or an earthquake event. The spread area is calculated for each event based on three emergency models: the McArthur Fire Danger Index Model, the Carvill Hurricane Index (CHI), and the Earthquake Damage Estimation, respectively.

To obtain the best alternative, a graph theory-based model to obtain the shortest path (based on Dijkstra's) is developed in G-BEP, and alternatives based on the travel time are generated. For this to be applicable, the original street map with the

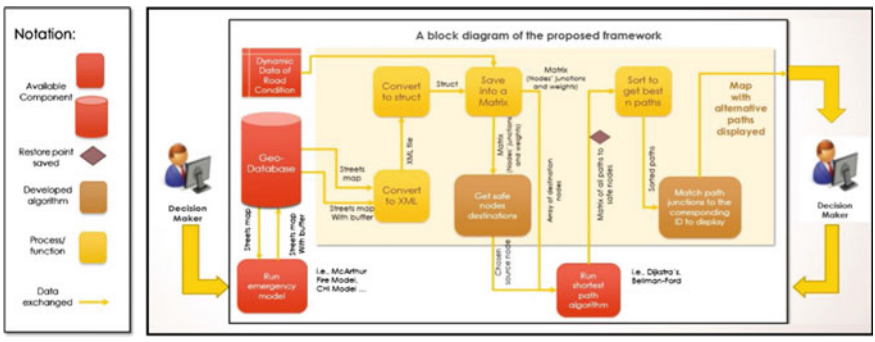


Fig. 1 G-BEP framework

drawn buffer zone must be converted to a graph showing streets as edges, street junctions as nodes, and the travel time between each node as an edge weight or cost. This is performed by executing some processes and functions that appear in the shaded area in the framework. The two developed algorithms appearing in that area are explained in the third sub-section of this section. Dijkstra's algorithm is chosen to be the shortest-path algorithm used in this research because according to [18, 19], it has been recognized as the best algorithm for the shortest-path problem since 1959.

After obtaining a result, the decision-maker commands a rescue team on the emergency spot with an evacuee vehicle through the generated paths; in that case, some roads' density would reach its maximum value: The best paths generated at a moment are not necessarily the best paths in the moment after. In addition to the traffic flow, other dynamic parameters could exist (these parameters are elaborated in the next section). Thus, dynamic data of the road conditions should be obtained and read into the framework so that weights/costs of the edges in the graph are recalculated and the result is updated. Feedback from the decision-maker would be a different time input for the rate of spread or a new decision-making process based on the dynamic data of the road conditions.

3.2 Selecting the Best Path

Some particular parameters contribute to selecting the best path: the current traffic flow, the safety of the road, and the existence of sudden blocks. The first parameter, traffic flow, determines the mean speed at any moment. The flow is calculated by knowing the current density of the road traffic. The number of lanes of the road and the road's capacity determine the number of vehicles passing a road at the certain time at which the flow is being measured. To calculate the flow, equations in [20] state the relationship between numbers of vehicles, length of the road, road density, and traffic flow. When the density reaches a maximum jam density, flow must be zero because vehicles will line up end to end. Thus, the road with a flow equals to zero must be excluded because its capacity will not allow extra new vehicles.

After obtaining the dynamic data of the road conditions, the new speed based on the traffic flow is updated and considered as new weights for the edges in the built graph. If the flow equals zero, the speed equals zero as well; thus, the travel time will equal infinity. In that case, the edge of that weight/cost is a nonexistent edge for the shortest-path algorithm because it searches for edges with minimal weights. Selection is based on obtaining the fastest route with the minimum travel time and other alternatives.

The second parameter, safety of the road, means that road's conditions must be considered safe, i.e., has the emergency's hazard already affected the road? If any road is affected by the emergency, it should not be considered as an existent edge for the decision maker to select as a potential alternative. Thus, as mentioned in the previous point, the weight/cost of that edge must equal to infinity.

The third parameter is the existence of any sudden blocks in the road. This means that if vehicles are already guided to select a road, a sudden block due to the emergency's consequences might appear at any segment of that road. In that case, all vehicles lined in that road should be moved back from that road to the nearest exit to other road. In that case, no new vehicles shall enter that road while the other vehicles are moving back, nor shall the road be considered as a potential alternative in the following decision-making process. Thus, the weight/cost of that edge must equal to infinity as well.

3.3 *Developed Algorithms in G-BEP*

The major feature of G-BEP is developing two algorithms (the brown squares in Fig. 1). One of them is for identifying safe destinations (mentioned in step 7), which are the closest safe nodes outside the buffer's boundary, which are sent to the shortest-path algorithm as destinations; the pseudo-code is shown in Fig. 2.

The algorithm is run after an overlay analysis (an intersect operation) is performed for the streets map after generating a buffer zone of the affected area. Some nodes and edges are inside the buffer, and some nodes and edges are outside its boundary. All of the features inside the buffer zones boundary that intersect with the outer graph are saved as a new feature. The algorithm marks nodes' junctions connecting those edges with a flag = 1. All other nodes outside the boundary take a zero flag. The algorithm visits each node inside the buffer, checks the flag of all edges of its neighbors, and then it checks neighbor nodes. If the algorithm finds no flag with value equaling 0, no safe node has been found yet. Looping on all nodes inside the buffer, the algorithm visits one of the neighbors of a last visited node. After checking the flags, if the algorithm finds one neighbor with a different flag, which is a safe node, it is saved into the safe_nodes array. The same process applies

```
1  [safe_nodes] = FUNCTION GetSafeNodes (weights, buffer_nodes, indx, flagged_nodes_array)

2  safe_nodes=[ ]                                  //Initialize
3  subset_w=weights(indx)                          //indx= indices of all nodes flagged with 1

4  FOR i=1 to length(buffer_nodes)                 //Loop for each node inside the buffer
5     nz_indx =find(subset_weights (i) >0)         //Search each row and save indices where
                                                   elements in the row > 0

6     current =find (flagged_nodes_array(nz_indx)==0)  //The index of a neighbor node outside the buffer
                                                   (flag=0), is saved into current
7     safe_nodes=[safe_nodes; nz_indx(current)]    //Add the nonzero element with index = current to the
                                                   safe nodes array
8  END

9  safe_nodes = unique(safe_nodes)                 //Get unique safe nodes without repetition
10 arc_safe_nodes = unique_nodes(safe_nodes)       //Map indices of the safe nodes to the main matrix of
                                                   nodes to get their values
11 END
```

Fig. 2 Identifying the safe-destinations pseudo-code

```
1  [Obj_ID] = FUNCTION GetIDs (nodes_info, unique_nodes, PATH)
2  obj_ID=[ ];                                          //Initialize
3  FOR i=1 to length (nodes_info(:,1))                  //Loop for each
                                                        from_junction
4      FOR j=1 to length (unique_nodes(PATH))           //Loop for path junctions
5        IF ( nodes_info(i,1) == unique_nodes (PATH (j)))  //Compare both to check if
                                                        there's intersection
6              FOR k=1 to length (unique_nodes (PATH))  //when found, each path
                                                        junction is rechecked for
                                                        matching with to_junctions
7                    IF ( unique_nodes (PATH (k))==nodes_info(i,2))
8                          obj_ID =[ obj_ID; nodes_info(i,4)];  //object_ID for matched
                                                        junctions is added to the
                                                        obj_ID array
9                    END
10             END
11       END
12     END
13 END
14 END
```

Fig. 3 Matching the path junctions with the object_ID pseudo-code

to all other nodes inside the buffer, and all safe nodes are saved. The safe nodes, which are junctions and edges connected to these nodes, represent the end of every path that will be found by the shortest-path algorithm from a source node to any of these safe-node destinations.

The second algorithm is for matching the path junctions (a path is the output of the shortest-path algorithm) with the saved streets' junctions to obtain the corresponding object_ID, which is sent to the map show the function to display the generated path. The pseudo-code is shown in Fig. 3.

The algorithm matches each value appearing in path with the saved junctions to obtain the ID when an intersection between a path and an (F_Junc, T_Junc) is found. It begins with searching F_Junc to match each element with all the elements in path.

Once it finds a matching element in F_Junc, it checks the value of T_Junc for the same element. Next, it searches if the same value in T_Junc appears in path as well. It loops from the beginning of PATH until a matching element is found. When an F_Junc paired with a T_Junc appears in the path, the algorithm saves the ID of the corresponding junctions.

3.4 Emergency Models and Equations

The models presented are used in the framework to determine the degree of hazard posed by the emergency event and to calculate the buffer of the affected area.

McArthur Fire Danger Index Model

The criteria of choosing a fire-danger rating are the area of validity, the availability of data for the model input, and the basis of the model. The chosen model is the

McArthur Forest Fire Danger Index. It works on open conditions on flat ground; input data are available; and it is an empirical model. Certain equations are added to the G-BEP system to calculate the Fire Danger Index (FDI), thus indicating which Fire Danger Rating (FDR) the fire has generated, and FROS determines the radius of affected area (buffer zone). The equations used to calculate the rate of spread on which the buffer zone area is defined are as follows [21, 22]:

$$\mathrm{FDI} = 2.0(10A)^{0.987}\mathrm{e}^{(-0.450+0.0338T-0.0345H+0.0234V)} \tag{1}$$

$$\mathrm{FROS} = 0.0012\mathrm{FDI} \times \mathrm{FL} \times \mathrm{e}^{(0.069S)} \tag{2}$$

where FDI is the fire danger index; A is the fuel-availability index (numerical value ranges from 0 to 1); T is the air temperature (°C); H is the relative humidity (%); V is the average wind velocity at a height of 10 m (km/h); FROS is the forward rate of spread (km/hr); FL is the fuel weight (t/ha); and S is the ground slope (degrees). Valid ranges include $A \in (0, 1)$, $T \in (5, 45)$, $H \in (5, 70)$, and $V \in (0, 65)$.

Carvill Hurricane Index

An equation to calculate CHI is added to G-BEP to calculate the buffer zone and output the hurricane index to the user, thus indicating the danger ranking of the hurricane according to inputs added by the G-BEP user. The CHI is calculated as follows [23]:

$$\mathrm{CHI} = \left(\frac{V}{V_0}\right)^3 + \frac{2}{3}\frac{R}{R_0}\left(\frac{V}{V_0}\right)^2 \tag{3}$$

where V is the maximum sustained wind speed of a hurricane in miles per hour; V_0 is constant and equals 74 mph; R is the radius of hurricane force winds of a hurricane in miles (i.e., how far from the center of hurricane winds $\geq$74 mph or greater are experienced); and R_0 is the radius of hurricane force winds at 74 mph, which equals 60 miles.

Earthquake Damage Estimation

In order to add an earthquake modeling feature to G-BEP system, it is required to have data regarding how far the damage of an earthquake may reach based on its magnitude. The radius of affected area as found by [24] estimated based on the Richter magnitude as follows (the buffer zone is based on the earthquake-magnitude input):

- Magnitude 5.0–6.0 = 2.36–6.14 km
- Magnitude 6.0–7.0 = 6.14–15.92 km
- Magnitude 7.0–8.0 = 15.92–41.30 km
- Magnitudes >8.6 = >41.30 km.

4 Implementation and Results

To evaluate the proposed framework explained in Chapter "Fuzzy-Vector Structures for Transient-Phenomenon Representation", the G-BEP prototype is implemented, and the results are displayed and compared. The emergency model is incorporated into the proposed solution by developing a new toolbar called Evacuation-Analysis Tool, which is easily added by the user into the program inside ArcMap®. The shortest-route analysis is carried out after performing the buffer analysis and displaying the buffer zone on the map, thus showing to the user the calculated affected zone of any previously added one of the three emergency events. The shortest-route analysis is performed in the second feature.

4.1 G-BEP Functionality of Buffer Analysis and Overlay

Using the newly developed Evacuation-Analysis Toolbar, shown in Fig. 4, the user can choose to add the event's data.

According to the mathematical model used, the model calculates two outputs: (1) the degree of hazard posed by the emergency event and (2) the radius of affected area, which in turn is sent to the ArcObjects® function that creates a buffer layer with the same projection and displays the buffer on the streets map.

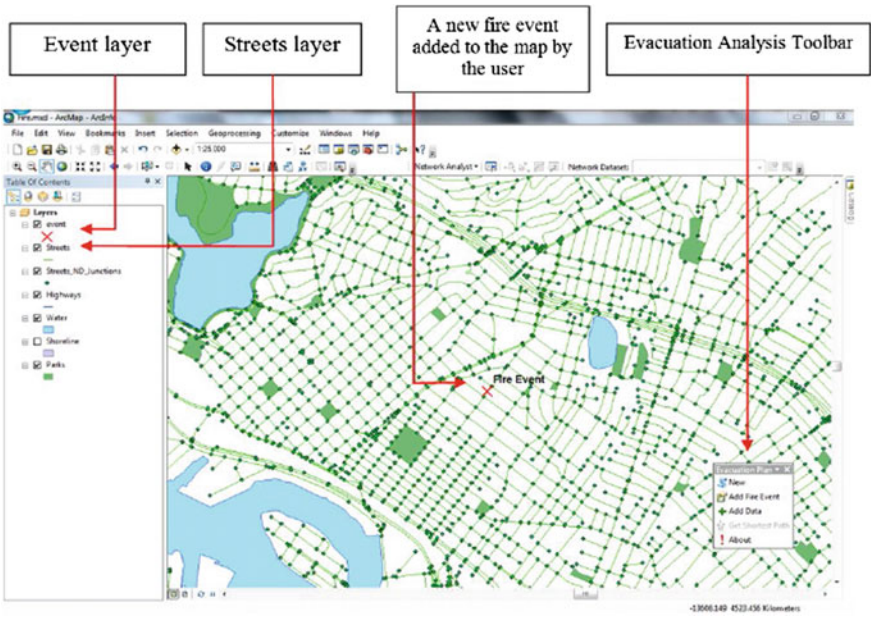


Fig. 4 An ArcMap document with an added streets layer and fire event

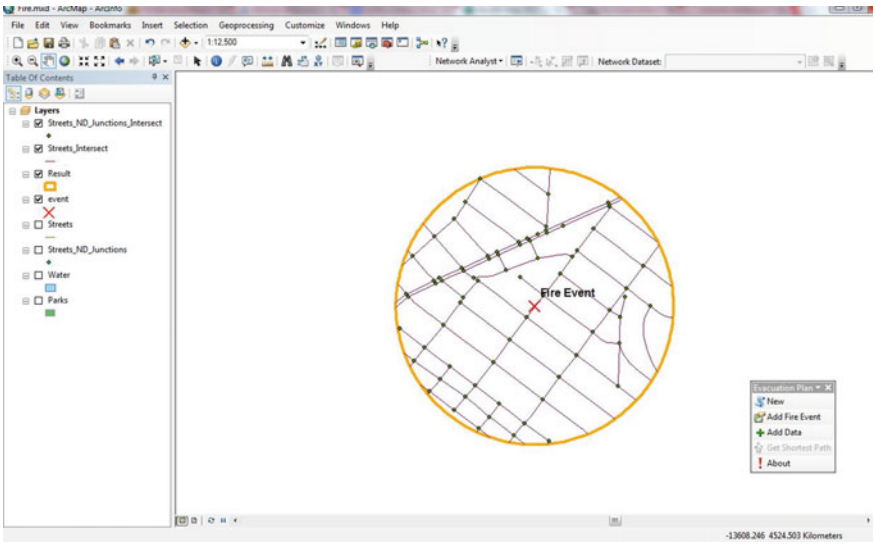


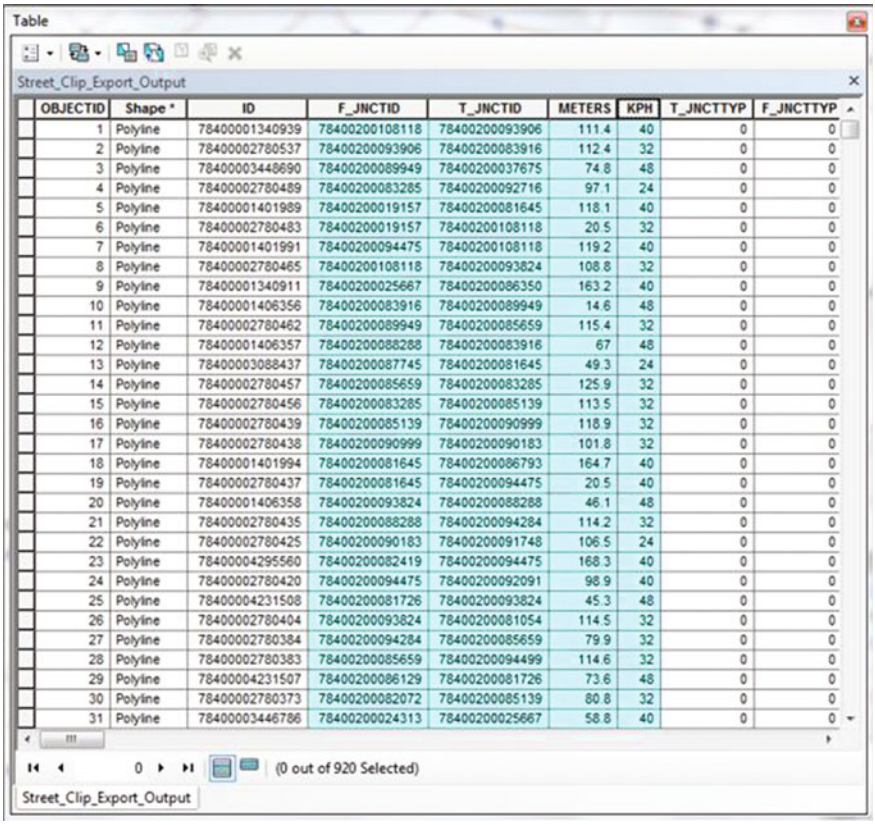
Fig. 5 Intersect layer after overlay analysis

An overlay analysis is performed in order to save the intersection between the streets layer and streets inside the buffer layer. A map with the displayed output of the overlay-analysis tool is shown in Fig. 5.

4.2 G-BEP Evacuation-Path Generation

A network dataset is needed as an input to G-BEP (streets map/layer). The network dataset [25] stores symbology for junctions, edges, and other needed attributes such as distance and speed (to calculate the travel time). The graph (sparse matrix) is constructed with these four attributes (shown in Fig. 6). To run any shortest-path algorithm, a source and a destination should be predefined, except in the case of using Bellman–Ford algorithm, in which no destinations are required. Thus, it computes the shortest paths to all other nodes in the graph, which would be meaningless and time-consuming in our case. To solve that problem, an algorithm for identifying all of the safe nodes outside the buffer is called up (as explained in Sect. 3), and all safe destinations are stored inside an array. Running both Dijkstra's and Bellman–Ford algorithms produces same results. At this point, the work is saved for historical records.

The best three alternatives from a centric source node after obtaining the corresponding Object_IDs are drawn on the shapefile being read into MATLAB® as shown in Fig. 7.



Table

Street_Clip_Export_Output

OBJECTID	Shape *	ID	F_JNCTID	T_JNCTID	METERS	KPH	T_JNCTTYP	F_JNCTTYP
1	Polyline	78400001340939	78400200108118	78400200093906	111.4	40	0	0
2	Polyline	78400002780537	78400200093906	78400200083916	112.4	32	0	0
3	Polyline	78400003448690	78400200089949	78400200037675	74.8	48	0	0
4	Polyline	78400002780489	78400200083285	78400200092716	97.1	24	0	0
5	Polyline	78400001401989	78400200019157	78400200081645	118.1	40	0	0
6	Polyline	78400002780483	78400200019157	78400200108118	20.5	32	0	0
7	Polyline	78400001401991	78400200094475	78400200108118	119.2	40	0	0
8	Polyline	78400002780465	78400200108118	78400200093824	108.8	32	0	0
9	Polyline	78400001340911	78400200025667	78400200086350	163.2	40	0	0
10	Polyline	78400001406356	78400200083916	78400200089949	14.6	48	0	0
11	Polyline	78400002780462	78400200089949	78400200085659	115.4	32	0	0
12	Polyline	78400001406357	78400200088288	78400200083916	67	48	0	0
13	Polyline	78400003088437	78400200087745	78400200081645	49.3	24	0	0
14	Polyline	78400002780457	78400200085659	78400200083285	125.9	32	0	0
15	Polyline	78400002780456	78400200083285	78400200085139	113.5	32	0	0
16	Polyline	78400002780439	78400200085139	78400200090999	118.9	32	0	0
17	Polyline	78400002780438	78400200090999	78400200090183	101.8	32	0	0
18	Polyline	78400001401994	78400200081645	78400200086793	164.7	40	0	0
19	Polyline	78400002780437	78400200081645	78400200094475	20.5	40	0	0
20	Polyline	78400001406358	78400200093824	78400200088288	46.1	48	0	0
21	Polyline	78400002780435	78400200088288	78400200094284	114.2	32	0	0
22	Polyline	78400002780425	78400200090183	78400200091748	106.5	24	0	0
23	Polyline	78400004295560	78400200082419	78400200094475	168.3	40	0	0
24	Polyline	78400002780420	78400200094475	78400200092091	98.9	40	0	0
25	Polyline	78400004231508	78400200081726	78400200093824	45.3	48	0	0
26	Polyline	78400002780404	78400200093824	78400200081054	114.5	32	0	0
27	Polyline	78400002780384	78400200094284	78400200085659	79.9	32	0	0
28	Polyline	78400002780383	78400200085659	78400200094499	114.6	32	0	0
29	Polyline	78400004231507	78400200086129	78400200081726	73.6	48	0	0
30	Polyline	78400002780373	78400200082072	78400200085139	80.8	32	0	0
31	Polyline	78400003446786	78400200024313	78400200025667	58.8	40	0	0

0 (0 out of 920 Selected)

Street_Clip_Export_Output

Fig. 6 Needed attributes to build the graph

To evaluate how the prototype responds to updates from the dynamic data, the weights of some edges are updated, and the shortest-path algorithm is run again to generate new results. In Fig. 8, it is assumed that two road segments (marked with X) have one of the parameters and thus must be blocked and ignored by the algorithm. Their speed value is set to zero, and thus travel time will equal to infinity. Because the shortest-path algorithm only considers edges with minimal weights, these two edges are ignored, and alternative new paths are generated. Note that paths from the same source node before updating the weights are shown in Fig. 7.

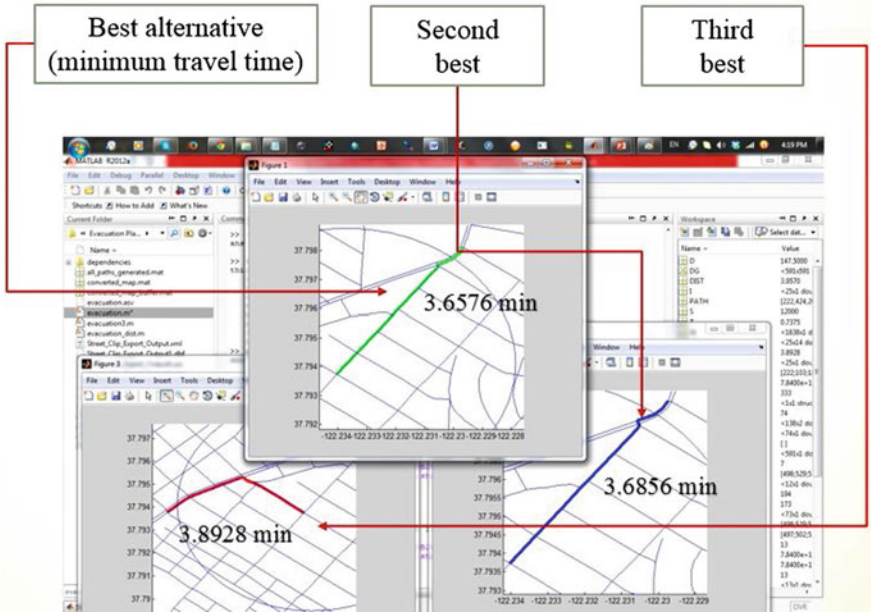


Fig. 7 The best three alternatives from one single source

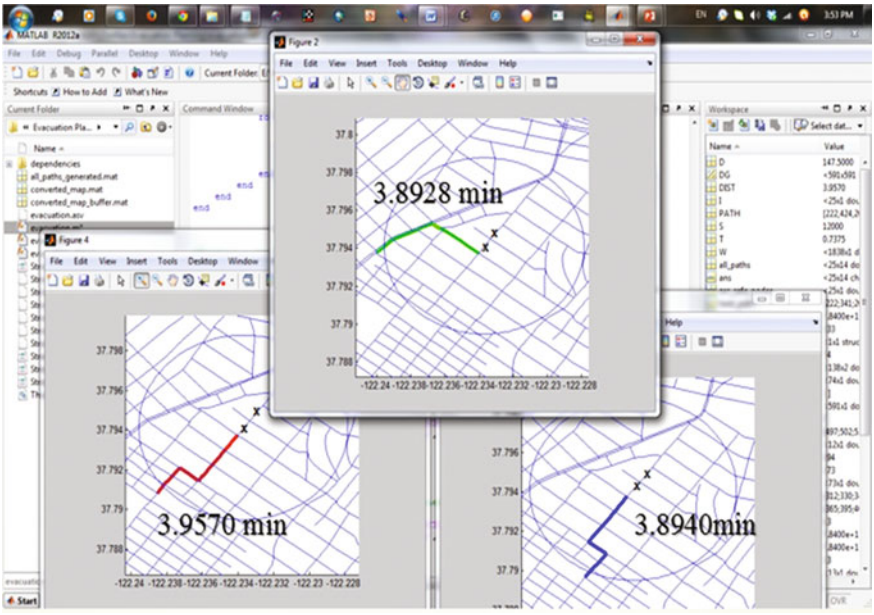


Fig. 8 The best three alternatives after updating the weights

5 Conclusion

This research presents a decision-making framework on the ArcGIS® platform that combines spatial-analysis and graph-theory algorithms with the three presented emergency models. Obtaining alternatives gives the decision-maker more options during a crisis in order to evacuate as many people as possible within an optimum time. The main contributions of this research are the integration of (1) three distinct emergency danger–prediction models incorporating features of the ArcGIS® Spatial Analyst; (2) the addition of a new toolbar in ArcGIS® called Evacuation-Analysis Tool; (3) the proposition of a framework that integrates GIS capabilities, emergency models, and shortest-paths algorithms and considers parameters based on dynamic road conditions, which contribute to the selection of best alternative paths; and (4) identification of the nearest safe destinations outside the buffer zone, not just one destination, but all of the best alternative routes based on graph-theory algorithms (multiple algorithms can be applied, and any number of alternatives can be displayed because all paths are saved).

The findings of this study have a number of important implications for future practice and policy. It is recommended to attach a live network of webcams and streaming-video cameras to the streets and to obtain online data updates through GPX (GPS-Exchange Format) files sent from vehicles or electronic devices at the emergency location, thus producing an evacuation path that depends on the evacuees' status, the weather variables, and the road conditions such as crowd and traffic congestion. Using real-time data would be applicable because all data would be dynamically read through streaming channels.

References

1. Ebada R, Mesbah S, Kosba E, Mahar K (2012) A GIS-based DSS for evacuation planning. ICCTA. Alexandria, Egypt, pp 35–40
2. Cova TJ (1999) GIS in emergency management. In: Longley PA, Goodchild MF, Maguire DJ, Rhind DW (eds) Geographical information systems: principles, techniques, applications, and management. Wiley, New York, pp 845–858, Published
3. Drabek TE, Hoetmer GJ (1991) Emergency management: principles and practice for local government. International City Management Association, Washington DC
4. ESRI (2008) Geographic information systems providing the platform for comprehensive emergency management. ESRI, New York, Published
5. ESRI (2000) Challenges for GIS in emergency preparedness and response. ESRI, New York, Published
6. Bo Y, Yong-gang W, Cheng W (2009) A multi-agent and GIS based Simulation for emergency evacuation in park and public square. Huazhong University of Science and Technology, Wuhan, China
7. Tao C, Hong-yong Y, Rui Y, Jianguo C (2008) Integration of GIS and computational models for emergency management. Center for Public Safety Research, Department of Engineering Physics, Tsinghua University, Beijing, P.R.China

8. Li Y, Yao L, Hu H (2010) Development and application of GIS-based multi-model system for risk assessment of geological hazards. Institute of Geology China Earthquake Administration, Beijing, China
9. Song W, Zhu L, Li Q, Liu Y, Dong Y (2009) Evacuation model and application for emergency events. Tianjin Polytechnical University
10. Jian-guo W, Jian-long Z (2009) Research and development on the expressway emergency response system based on GIS. School of Traffic and Transportation Engineering, Changsha University of Science& Technology, Changsha, China
11. Zhao H, Mao H (2009) Highway emergency response system based on GIS-T. School of Automobile Engineering, Harbin Institute of Technology, Weihai, China
12. Hormdee D, Kanarkard W, Taweepworadej W (2006) Risk management for chemical emergency system based on GIS and Decision Support System (DSS). Department of Computer Engineering, Khon Kaen University, Thailand
13. Zhichong Z, Yaowu W (2009) Framework of spatial decision support system for large-scale public building evacuation. Harbin Institute of Technology, Harbin, China
14. Ling A, Li X, Fan W, An N, Zhan J, Li L, Sha Y (2009) Blue arrow: a web-based spatially-enabled decision support system for emergency evacuation planning. School of Information Science and Engineering, Lanzhou University, Lanzhou, China
15. Zhu L, Song W, Li Q (2009) Construction of emergency decision system based on GIS. Tianjin Polytechnical University, Tianjin, China
16. Wenjing M, Yingzhuo X, Hui X (2009) The optimal path algorithm for emergency rescue for drilling accidents. Institute of Computer Technology, Xi'an Shiyou University, Shannxi, China
17. Shashikiran V, Kumar TS, Kumar NS, Venkateswaran V, Balaji S (2011) Dynamic road traffic management based on Krushkal's algorithm. Department Of Computer Science, Sri Venkateswara College of Engineering, Chennai, India
18. Yuanzhe X, Zixun W, Qingqing Z, Zhiyong H (2012) The application of Dijkstra's algorithm in the intelligent fire evacuation system. Polytechnic Engineering, QiongZhou University, Sanya, China
19. Pettie W (2002) A new approach to all-pairs shortest paths on real-weighted graphs. Theoret Comput Sci 312(1):47–74
20. Banks JH (1991) Two capacity phenomenon at freeway bottlenecks: a basis for ramp metering? Transp Res Rec 1320:83–90
21. Rothermel RC (1972) A Matrematical model for predicting fire spread in wildland fuels. Intermountain Forest And Range Experiment Station, Forest Service, U.S. Department of Agriculture, Ogden, Utah 84401, USA
22. Willis C, Wilgen BV, Tolhurst K, Everson C, D'Abreton P, Pero L, Fleming G (2001) The development of a national fire danger rating system for South Africa. Department of Water Affairs and Forestry Pretoria, Pretoria
23. Carvill Hurrican Index Whitepaper, http://www.cmegroup.com/trading/weather/files/Hurricane-CHIWhitepaper.pdf. Last accessed 25 Aug 2014
24. Saradjian MR, Akhoondzadeh M (2011) Prediction of the date, magnitude and affected area of impending strong earthquakes using integration of multi precursors earthquake parameters. Remote Sensing Division, Surveying and Geomatics Engineering Department, University College of Engineering, University of Tehran, Iran, Copernicus Publications on behalf of the European Geosciences Union, doi:10.5194/nhess-11-1109-2011
25. ESRI (2004) Network dataset defined. Available: http://help.arcgis.com/en/arcgisdesktop/10.0/help./index.html#//004700000007000000. Last accessed 25 Aug 2014

Optimized Conflation of Authoritative and Crowd-Sourced Geographic Data: Creating an Integrated Bike Map

Linna Li and Jose Valdovinos

Abstract A complete and accurate geographic dataset is critical for relevant analysis and decision-making. This chapter proposes a four-step geographic data-conflation system: preprocessing, automatic conflation, evaluation, and manual adjustments. The automatic-conflation component uses an optimization approach to find matched features and a rubber-sheeting approach to complete spatial transformation. This system was tested using two bikeway datasets in Los Angeles County, California, from an authoritative source (Los Angeles County Metropolitan Transportation Authority) and an open source (OpenStreetMap). While bikeways that are already in both datasets are improved in terms of positional accuracy and attribute completeness, the conflated bikeway dataset also integrates complementary data in either of the input datasets. Experiments demonstrate the advantages of using crowd-sourced data to improve official bikeway data, which is important for building and maintaining high-quality bicycle-infrastructure datasets. The framework described in this chapter can be adapted to conflate other types of data themes.

Keywords Geographic information fusion · Conflation · Bikeway data · Optimization · Algorithm · Uncertainty · Spatial-data quality

1 Introduction

The 21st century is the era of big geospatial data: Geographic information is generated at an unprecedented pace, thus providing a great potential to study physical and human phenomena on the Earth's surface. The large number of Earth-Observing

L. Li (✉) · J. Valdovinos
Department of Geography, California State University, 1250 Bellflower Blvd, Long Beach, CA 90840, USA
e-mail: linna.li@csulb.edu

J. Valdovinos
e-mail: Jose.Valdovinos@student.csulb.edu

V. Popovich et al. (eds.), *Information Fusion and Intelligent Geographic Information Systems (IF&IGIS'17)*, Lecture Notes in Geoinformation and Cartography, DOI 10.1007/978-3-319-59539-9_17

satellites and widespread availability of sensor networks have made vast volumes of georeferenced information available. For example, the leading Earth-Observing agency, the National Aeronautics and Space Administration (NASA), has collected billions of gigabytes of data through its approximately 30 operating satellite missions. Furthermore, benefitting from the readily available location-aware communications and the recent advancement of social-media technologies, billions of people, serving as citizen sensors [1], are now capable of creating large amounts of geo-tagged data on various aspects of human life. Although we can almost always obtain data about the same phenomena from different sources, it is not always straightforward to take advantage of this abundance due to inconsistency, incompatibility, and heterogeneity among various datasets. The activity of combining data from different sources for further data management and analyses is called "conflation".

According to the Oxford English Dictionary, the word "conflation" means "the action of blowing or fusing together; composition or blending of different things into a whole". It implies the merging of different things into an integrated whole and was first used to refer to the process of combining two geographical data sources in a semi-automatic map-merging system developed at the U.S. Census Bureau [2]. They defined conflation of maps as "a combining of two digital map files to produce a third map file which is 'better' than each of the component source maps". In addition to map conflation, other communities are also interested in combining data from different sources. A closely related area is image fusion in remote sensing [3] which mostly focuses on the combination of data at the pixel level. "Data fusion," as a broader concept, refers to techniques used for merging data from different sources for better inferences in a wide range of applications including automated target recognition, battlefield surveillance, medical diagnosis, and robotics [4]. In the database literature, conflation is referred to as a type of data integration that focuses on the combination of heterogeneous databases to give a unified view of input data by mapping different local schemas to a global schema [5]. For datasets with the same data schema but different values for the same attribute, the problem is referred to as "record linkage," "record matching," and "duplicate-record detection" in the statistics community with a main goal to identify entries in different files that represent the same entity in reality [6].

Human brains are very good at integrating information from different sources. In conducting scientific research and making everyday decisions, people usually gather relevant information from different sources, take all of them into consideration, and draw a conclusion according to a weighting scheme assigned to each piece of information either explicitly or implicitly. In terms of scientific research, in ancient times geographers worked on paper maps, put them next to each other, or stacked several maps of the same region to do spatial analyses. Before the emergence of computers and digital datasets, the need for geographic data conflation was

not so obvious. In the early stages of computers, data conflation was also not a big issue because very few geographical datasets were created, and there was rarely more than one dataset of the same theme for a particular region. As computer hardware and software improved dramatically, especially the development of geographic information systems (GISs) and the advancement of all sorts of data-collection methods, large volumes of geospatial datasets were created at relatively low prices. In the past two decades, the rapid growth of the Internet and the dissemination of free software have provided scientists and policy-makers with easier access to geospatial data and thus the possibility to improve research and decisions. For example, volunteered geographic information (VGI), such as OpenStreetMap and Ushahidi, has grown rapidly as an alternative source for authoritative geographic information generated by government agencies and commercial companies [1]. Although abundant geographic data enable us to obtain knowledge in ways that were not possible before, it also presents significant new challenges for appropriately using data from multiple sources, produced for different purposes by various agencies, in a consistent and reliable way. Therefore, conflation of different geographic datasets is critical for us to use data generated through various mechanisms including authoritative datasets usually provided by government agencies and crowd-sourced datasets created by potentially anyone with geospatial technologies. This chapter proposes a geographic data-conflation system that relies on an optimization approach to integrate various data sources. Experiments with two bikeway datasets in the Los Angeles County of California, U.S. show the success of this system by creating an integrated bike map with more complete coverage of bicycle paths than either of the two input datasets and enriched attributes.

The rest of the chapter is organized as follows. Section 2 reviews relevant work on geographic data conflation. Section 3 presents an optimized approach to conflating authoritative and volunteered geographic data in a four-step framework. Section 4 describes experimental results on creating an integrated bike map for Los Angeles County, using our approach. The chapter ends with conclusions and future research suggestions in Sect. 5.

2 Related Work

Geographic-data conflation is necessary because no geographic dataset can possibly represent the complex world without any information loss or simplification, which means that uncertainty is certain in any geographic dataset. When people create geographic data, they use different filters to select and conceptualize reality. The uncertainty of geographic information has received increasing attention in the GIScience literature [7–11]. With respect to sources of data uncertainty, it could be introduced in the data-creation process due to limited knowledge, measurement

errors, etc., and it may also be propagated in data processing. Heuvelink [12] discussed several error models and uncertainty propagation in local and global GIS operations. Shi [13] summarized the error models for basic GIS features and investigated uncertainty in overlay and buffer analyses as well as quality control of spatial data. Furthermore, Kyriakidis et al. [14] investigated the use of geostatistics to quantify uncertainty in conflation of DEMs and to obtain elevation measurements of greater accuracy.

Understanding the omnipresence of uncertainty is important for studying the nature and challenges of conflation. To characterize uncertainty introduced in conflation, Saalfeld [15] discussed three types of errors as well as their implications and detections: (1) false-positive: a map feature is identified as having a match when it does not in fact have one; (2) mismatch: a feature is correctly identified as having a match, but its matching feature is incorrectly selected; (3) false-negative: a feature is identified as not having a match when it actually does. Saalfeld also pointed out that the occurrences of false-positives and mismatches in the early stage of conflation are more serious than false-negatives because they can precipitate more errors and are usually hard to correct in later stages.

In information retrieval, precision and recall are two widely used measures for evaluating the quality of results. Precision is a measure of exactness, whereas recall is a measure of completeness. In literature that provides information on conflation quality, precision and recall are usually the two indicators [16–18].

As mentioned previously, conflation is the process of combining information from multiple geospatial sources to generate new knowledge to meet the needs of particular applications. In the simplest case, it involves a one-to-one match between objects in two datasets that both represent the same entity. In other situations, there may be one-to-none or one-to-many correspondences. For example, in one dataset, a local street is represented as a polyline, whereas in another dataset, it is not stored because of the difference of the resolution or human generalization. This complicates the problem because decisions must be made about which map source should be treated as more reliable when there is a conflict. In some cases, such decisions must be made on a feature-by-feature basis.

According to the data model of input datasets, conflation can be classified into three categories: field-to-field conflation, field-to-object conflation, and object-to-object conflation. Field-to-field conflation and field-to-object conflation are usually referred to as “image registration” or “image fusion”. These two kinds of conflation mostly involve the identification of corresponding control points at the pixel level followed by spatial transformation and resampling. For a detailed review of image-registration methods, see the works by Brown [19] and Zitova [20]. Particularly, Pohl and Van Genderen [3] described and compared major techniques in multi-sensor remote-sensing image processing in a variety of applications. When features are extracted from images using segmentation procedures or other techniques, or created by surveying and digitization, the combination of multiple data sources is object-to-object conflation, which is the focus of this chapter.

3 A Geographic Data Conflation System

The workflow of a conflation system is demonstrated in Fig. 1. The input to this system is the datasets to be conflated. Each dataset is a database of features and associated locations, attributes, and metadata.

The first step is the pre-processing of input datasets, which ensures that individual datasets are internally consistent, that they are under the same coordinate system and projection, and that they have overlapped spatial coverages, etc. To optimize the matching process, data preparation may be performed on the source datasets. For example, the global geometric deviation between multiple datasets must be reduced before conflation [21]. For network data, it is desirable to have linear features in different datasets with as many 1:1 correspondences as possible. In addition, source datasets should have some minimum similarity in the objects, which is the foundation of conflation. For instance, we cannot conflate a road network to a stream network.

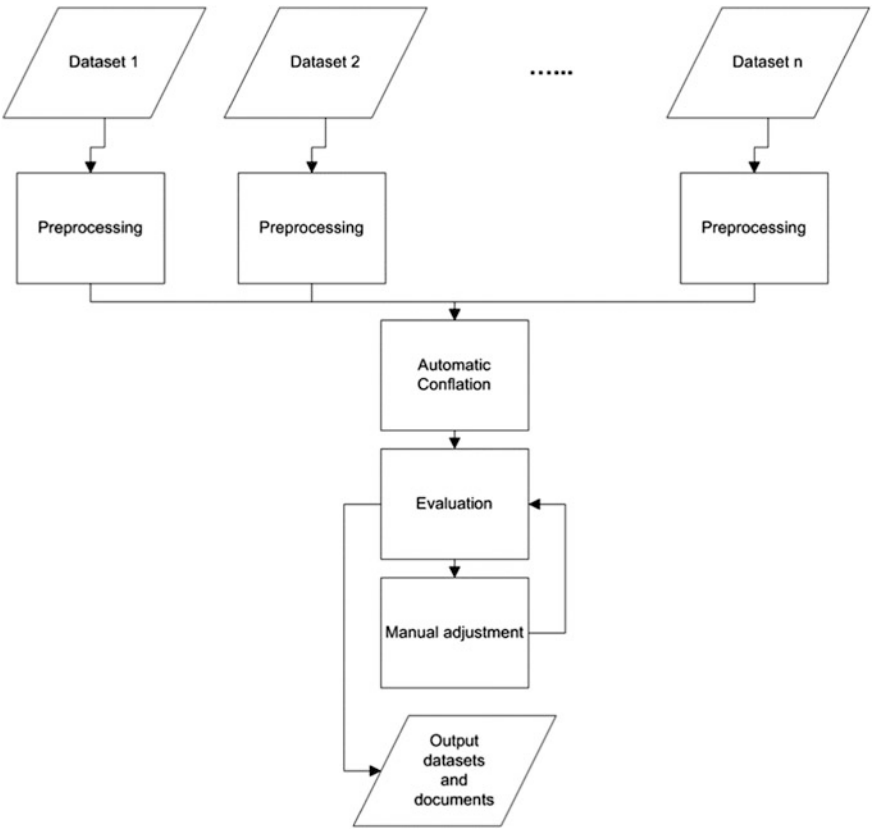


Fig. 1 Workflow of a conflation system

The second step is automatic conflation, which creates new data. This step is composed of two major parts: The first one is the definition of a schema for the new data, and the second part is the application of various techniques to execute conflation. The conceptual schema is created based on the objective of a particular application. It describes the general data structure for the conflated datasets, the spatial coverage, the necessary objects, and all of the desirable attributes, etc.

The third step is the evaluation of the conflation results based on the quality report generated in the previous step including the uncertainty of feature matching, etc. Another important assessment here is visual examination of the conflated datasets. Measurements and a statistical report alone may not provide sufficiently reliable confidence of accurate conflation. Because of the complexity of the spatial object representations and the limitations of the computer technology, human intervention may be needed at this stage to check the mismatched objects and uncertainty.

The fourth step is the manual adjustment of conflation. With uncertainty evaluations from the previous step, more user attention can be directed to areas with high uncertainty. Users are allowed to correct mismatched features and associated attributes. Furthermore, users can adjust conflation parameters such as different weights of various criteria and the threshold for the proximity. The output of this process includes datasets required by a particular project and a document that records the quality of the conflated dataset, the relationship between new datasets and original ones, as well as the lineage information.

3.1 Automatic Feature-Matching Using Optimization

After pre-processing, one major step in conflation is feature matching, which identifies features representing the same object of the world in different data sources. Different methods of feature-matching have been developed during the years (e.g., [2, 15, 21–27]). Our system adapts the optimization model for linear feature-matching proposed by Li and Goodchild [25, 26]. Two important components in this approach include a similarity score and an optimized strategy to find matched pairs in input datasets based on the defined similarity score.

Similarity score here is defined as a weighted measurement of directed Hausdorff distance, angle, and dissimilarity between feature names. Directed Hausdorff distance is defined as follows:

$$d_{i\to j}^{DH} = \max\left\{d(x, L_j)\right\} \tag{1}$$

where $d(x, L) = \min\{d(x, y): y \in L\}$ is the shortest distance between a point x and a line L. We expect two corresponding linear features to have similar direction, so if the angle between them is larger than a certain degree, they should not be matched. The similarity score from feature i to feature j is defined as follows:

$$s_{i\to j} = \begin{cases} 0, & if\ d_{i\to j}^{DH} > a \\ a - d_{i\to j}^{DH} & if\ d_{i\to j}^{DH} < a \text{and} D_{ij}^{n} \text{is not available} \\ a - \left(D_{ij}^{n} + d_{i\to j}^{DH}\right)\Big/ 2 & if\ d_{i\to j}^{DH} < a \text{and} D_{ij}^{n} \text{is available} \end{cases} \quad (2)$$

where $d_{i\to j}^{DH}$ is the directed Hausdorff distance; D_{ij}^{n} is the distance between two feature names; and a is a distance threshold beyond which two features are considered too far away to be matched.

Furthermore, feature-matching is formulated as an optimization problem with the following objective function:

$$\text{Maximize} \sum_{i=1}^{k} \sum_{j=1}^{j} s_{i\to j} z_{i\to j} \quad (3)$$

where i, j are indices for the features in the first and second dataset, respectively; k and l are the number of features in each dataset; and $s_{i\to j}$ is the directed similarity score defined in Eq. (2). The variable $z_{i\to j}$ represents a match from feature i to feature j, taking a value of 1 if a match is made and 0 otherwise. A solution to this problem gives matched feature pairs in the two datasets.

Based on the conceptual model described previously, a feature-matching tool has been developed in ArcMap (Fig. 2). This tool takes two linear datasets as input and requires two parameters (default values of the parameters are specified). The first one is the cut-off distance d. If the distances between a feature i in dataset 1 and features in dataset 2 are all larger than d, the feature i is not matched to any feature in dataset 2. The default value of d is set to 25 m and can be changed based on the

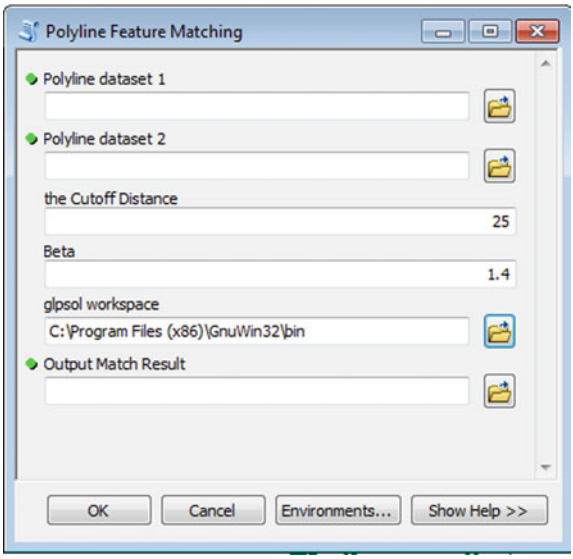


Fig. 2 Interface of an optimized feature-matching tool in ArcMap

uncertainty and discrepancy of the input datasets. The second parameter is *beta*. It is used to ensure that the total length of all features in dataset 1 that are matched to the same feature *j* in dataset 2 does not exceed the length of feature *j* times beta. The parameter *beta* is a tolerance factor that takes into account uncertainty in feature lengths in different datasets depending on the resolution of the two input datasets. For input datasets of similar resolution as that in our experiments, *beta* is a value a little larger than 1 (e.g., *beta* = 1.4). If dataset 2 has a much greater resolution than dataset 1, i.e., the same linear feature has a longer length in dataset 2, *beta* is smaller than 1. On the contrary, if dataset 2 has a lower resolution than dataset 1, *beta* should be assigned a value >1 based on the ratio of the resolution of two datasets. However, determination of the value for *beta* may need a closer examination of the data for more complicated situations.

3.2 Feature Transformation

Another major step in conflation is feature transformation, which is intended to improve spatial and non-spatial information or to calculate new values for certain attributes. Spatial transformation is used to improve the spatial accuracy of conflated datasets and has been studied mostly in the area of image registration [19, 20]. Two strategies may be adopted: First, the spatial information in one dataset is adjusted to be consistent with the spatial information in the other dataset, which usually has a greater spatial accuracy; second, the spatial information in the conflated dataset is derived from all input datasets and is supposed to have greater accuracy than any of them. Two basic steps in spatial transformation are the selection of control points and construction of the transformation functions according to control points. To enable local transformation, spatial partitioning is usually needed. Therefore, the three steps in spatial transformation are the identification of control points, spatial partitioning, and the establishment of transformation functions in each sub-region. Most research efforts concentrate on the first case, which adjusts a target dataset using a reference dataset, whereas rubber-sheeting is widely used to arrange the geometry of objects in the target dataset [15, 28, 29]. Our system adopts Saalfeld's method by defining rubber-sheeting as a map transformation that allows local distortions but preserves topological properties and is linear on sub-sets of the map. First, similar to point-to-point matching, corresponding control points are identified in different datasets according to a similarity criterion followed by the construction of corresponding sub-regions—usually triangles (especially Delaunay triangles) in both datasets—to stretch each sub-region by matched points to meet its predefined corresponding area. Then the positions of points or vertices that are within each sub-region in the conflated map are calculated according to locations of control points through a linear transformation.

Non-spatial transformation is concerned with the improvement of attributes in the conflated dataset. The problem in this task is mostly deconfliction of attribute

values because different kinds of attribute discrepancies can occur. Cobb et al. [30] summarized three kinds of situations: (1) different datasets have the same attributes; (2) one dataset has more attributes than the others; and (3) the datasets have completely different attributes. It happens quite often that values of the same attribute in different datasets are not consistent, so the determination of which value or derived value should be selected for the conflated dataset is a big issue. For case no. 2, one might want to choose common attributes or the superset of the attributes. Case no. 3 is relatively easy to deal with due to the lack of information because we can take the union of the attributes or those applicable to a particular project. In summary, there are two issues to consider in non-spatial transformation. First, which attributes should be included in conflated datasets? This is determined by the objective and the requirements of particular applications. Second, when there is a conflict in the attribute, how should we determine the value in the final dataset?

4 Experiments

In this section, we test our proposed geographic data-conflation system on two datasets (Fig. 3) that contain bikeway data of Los Angeles, California, U.S., obtained from two sources: Los Angeles County Metropolitan Transportation Authority (Metro) and OpenStreetMap (OSM). The former dataset was compiled and provided by the governmental agency in 2012 and is regarded as an authoritative data source, and the latter was created by the OSM community with the most recent update and is an example of volunteered geographic data. In total, there are 1383 bikeways in the Metro dataset and 723 features in the OSM dataset. As shown in Fig. 3, these two datasets contain some common features, and some bikeways are only present in one of the datasets. Our main goal is to conflate the two datasets into an integrated and consistent dataset that contains all of the features of the two original datasets but without any duplicates. Because the Metro dataset was created using quality-control procedures as an authoritative dataset, we use it as a reference dataset here, but some bikeways in the OSM dataset are not present in the Metro dataset, so those features will be integrated to update the Metro dataset. This type of dataset update can be scheduled periodically given that the volume of crowd-sourced data are ever growing, whereas the cycle of map updating in governmental agencies is long (which is why the most recent dataset was created in 2012). Preprocessing was performed on the two datasets to ensure their internal consistency and the same coordinate system NAD_1983_UTM_Zone_10N.

4.1 A Schema for the Conflated Dataset

Both input datasets use polylines to represent bikeways in LA, California, so we will keep the same geometric type in the conflated dataset. In contrast, the two

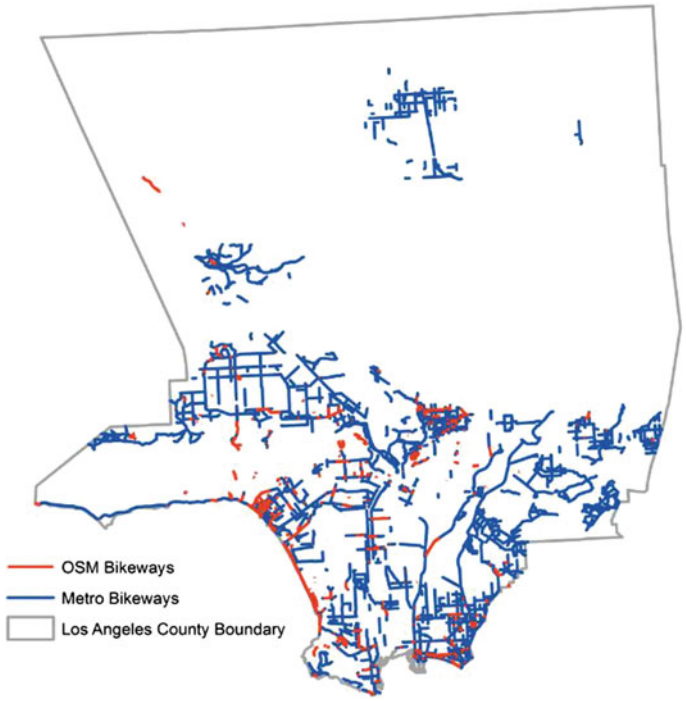


Fig. 3 Bikeway datasets from Metro and OSM in Los Angeles, California, U.S.

Metro

FID	Shape *	Name	FolderPath	length
0	Polyline ZM	EL DORADO PARK	2012 Bikeways of Los Angeles County/Class 1	682.707109
1	Polyline ZM	EL DORADO PARK	2012 Bikeways of Los Angeles County/Class 1	383.520295
2	Polyline ZM	EL DORADO PARK	2012 Bikeways of Los Angeles County/Class 1	378.772066
3	Polyline ZM	EL DORADO PARK	2012 Bikeways of Los Angeles County/Class 1	790.433058
4	Polyline ZM	EL DORADO PARK	2012 Bikeways of Los Angeles County/Class 1	220.076839

OSM

FID	Shape *	osm_id	name	highway	other_tags	Length
0	Polyline	3102198	Golden State Freeway	motorway	"hgv"=>"designated","ref"=>"I 5","lanes"=>"4","onewa	974.818
1	Polyline	3831354	Pacific Coast Highway	primary	"hgv"=>"designated","ref"=>"CA 1","oneway"=>"yes",	87.4486
2	Polyline	3987199	Maple Street	primary	"oneway"=>"yes","cycleway"=>"lane","source_ref"=>	1263.3
3	Polyline	4000078	Monterey Road	tertiary	"lanes"=>"3","oneway"=>"yes","cycleway"=>"lane","s	27.7023
4	Polyline	4082352	Stadium Way	tertiary	"lanes"=>"2","cycleway"=>"lane"	269.71

Fig. 4 Different attributes in the Metro and OSM datasets. (*Upper panel*) Attribute table of the Metro dataset. (*Lower panel*) attribute table of the OSM dataset

datasets include different sets of attributes (Fig. 4). The Metro dataset contains well-defined attributes including name, folderpath (containing classification information of bikeways), and length, whereas the OSM dataset contains some loosely defined attributes including name, highway, other_tags, and length. In the integrated dataset, we want to keep three main attributes: *name*, *type*, and *length*.

The attribute *name* stores information about the name of a bikeway such as El Dorado Park or Pacific Coast Highway. The attribute *type* stores information about the classification of a bikeway, and the attribute *length* is the length of a bikeway polyline. The attributes *name* and *length* are straightforward, but the attribute *type* must be matched between the two datasets, which is usually called "semantic matching" or "ontology matching" as a way to establish semantic correspondences between input schemas [31, 32].

4.2 Feature Matching

The two datasets were first automatically matched using the tool described in Sect. 3.1 and then manually adjusted based on the evaluation report of the automatic feature-matching results. When the similarity score between two features is smaller than a threshold value, user inspection validates the matching results. Finally, the matching result is compared with the true matched pairs, which are established by human visual inspection. We used precision to assess the matching results, which is defined as the number of correctly matched features divided by the sum of true- and false-positives. In the automatic-matching step, precision is as high as 92.54%, which greatly decreases our work in the manual-adjustment step. A matched pair of features indicates the same bikeway in the reality, whereas an unmatched feature in one dataset is a missing feature in the other dataset. The Metro dataset is an authoritative dataset, so all of the features have a consistent naming convention and have values in the *name* attribute. In contrast, some features in the OSM dataset have missing names. Therefore, our strategy is to keep all of the features in the Metro dataset and add complementary features from the OSM dataset. As a result, a set of 292 features in the OSM dataset that do not have corresponding features were added to the Metro dataset. In the conflated bikeway dataset, we have a total of 1675 features that contain all of the features from both input datasets without any duplicates.

4.3 Feature Transformation

As mentioned in Sect. 3.2, we use rubber-sheeting to do spatial transformation to ensure that the features from the OSM dataset are well aligned with those from the Metro dataset in the conflated dataset. This step was completed using the Rubbersheet Features Tool in ArcMap with vertices of matched bikeways selected as control points.

Regarding non-spatial transformation, among the three attributes in the defined schema of the conflated dataset, the definitions of various bikeway types are not

consistent in the two input datasets. In the Metro dataset, four types of bikeways are defined as follows:

- *Bike Path*: Off-street shared-use path for bicycles and pedestrians;
- *Bike Lane*: On-street striped and signed dedicated travel lane for bicycles;
- *Bike Route*: On-street travel lane shared by bicyclists and other vehicular traffic; and
- *Cycle Track*: On-street one or two-way bikeways that are physically separated from other vehicular traffic.

In the OSM dataset, the classification of bikeways is not as well-defined. Due to the nature of crowd-sourced geographic data, people may choose different ways to tag geographic features. A bikeway may have the value "cycleway" in the tag *highway* or may have a "cycleway" tag or a "bicycle" tag. For example, a bikeway may be marked as a cycleway in the *highway* tag as "track," "lane," or "shared_lane" in the *cycleway* tag or as "designated" in the *Bicycle* tag. More classes are used in the OSM dataset with various tags. After comparing different classes of bikeways in the two datasets using high-resolution Google Earth images, we decide to use the four classes in the Metro dataset as the classification system in our conflated dataset and mapped the classes in the OSM dataset to the four classes. Photos of the four classes are shown in Fig. 5. Figure 6 shows the semantic-matching relationship between the two datasets.

Fig. 5 Four classes of bikeways in the conflated dataset

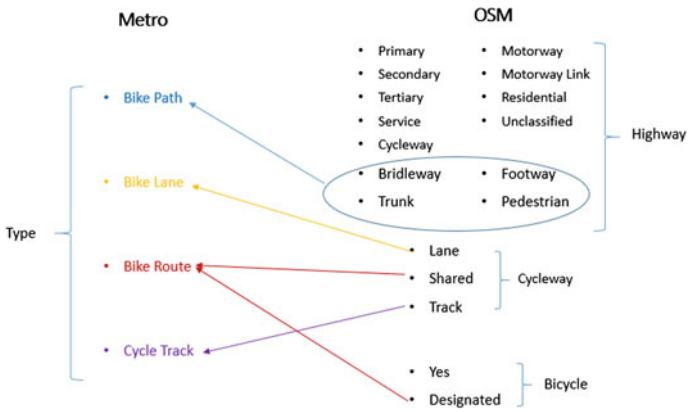


Fig. 6 Semantic mapping between two datasets

5 Conclusions and Future Work

This chapter presents a geographic data-conflation system to integrate diverse data sources into a consistent conflated dataset in order to improve the positional accuracy, attribute accuracy, and completeness of the geographic data. The system is composed of four steps: preprocessing, automatic conflation of different sources, evaluation, and manual adjustments. The automatic-conflation component uses an optimization approach to find matched features and a rubber-sheeting approach to complete spatial transformation. Feature matching is completed in a two-step procedure. First, a similarity score is calculated between the features in two datasets based on location and attributes. Second, the features are matched based on an optimization method. The optimized feature-matching method is superior to other sequential methods because it considers all matched pairs simultaneously. In a sequential feature-matching procedure, one matching mistake in a previous step can affect later steps. In addition, non-spatial attributes are transformed based on ontology matching. As demonstrated in the experiments with bikeway datasets obtained from Metro and OSM, this system successfully conflated two datasets into a coherent one with four well-defined classes of bikeways and a more complete coverage of 1675 features. The automatic-conflation component significantly decreases human intervention by achieving precision as high as 92.54%. This system can readily be used in other cases of linear-feature conflation and can be adapted for point and polygon datasets with new similarity measurements. Our next step is to design and develop a conflation system that is able to conflate more than two datasets. In the era of big data, a feature may be represented as multiple different versions. With sufficient amounts of data, these datasets may be conflated using a statistical method to create an average feature that is a better representation

of reality than any of the input datasets. For instance, multiple GPS trajectories of the same bikeway provided by different individuals can be generalized to obtain an accurate representation of the feature.

References

1. Goodchild MF (2007) Citizens as sensors: the world of volunteered geography. GeoJournal 69(4):211–221
2. Lynch MP, Saalfeld AJ (1985) Conflation: automated map compilation—a video game approach. In: Proceedings of AUTOCARTO 7, Washington, DC
3. Pohl C, Van Genderen JL (1998) Multisensor image fusion in remote sensing: concepts, methods and applications. Int J Remote Sens 19(5):823–854
4. Hall DL, Llinas J (1997) An introduction to multisensor data fusion. Proc IEEE 85(1):6–23
5. Lenzerini M (2002) Data integration: a theoretical perspective. In: Proceedings of the Twenty-First ACM SIGMOD-SIGACT-SIGART symposium on principles of database systems (Madison, Wisconsin, June 2003–2005, 2002). PODS '02. ACM, New York, NY, pp 233–246
6. Elmagarmid AK, Panagiotis GI, Vassilios SV (2007) Duplicate record detection: a survey. IEEE Trans Knowl Data Eng 19:1–16
7. Goodchild MF, Gopal S (eds) (1989) Accuracy of spatial databases. Taylor and Francis, New York
8. Heuvelink GBM, Burrough PA (2002) Developments in statistical approaches to spatial uncertainty and its propagation. Int J Geogr Inf Sci 16(2):111–113
9. Zhang J, Goodchild MF (2002) Uncertainty in geographical information. Taylor and Francis, New York
10. Couclelis H (2003) The certainty of uncertainty: GIS and the limits of geographic knowledge. Trans GIS 7(2):165–175
11. Goodchild MF, Zhang J, Kyriakidis P (2009) Discriminant models of uncertainty in nominal fields. Trans GIS 13(1):7–23
12. Heuvelink GBM (1998) Error propagation in environmental modelling with GIS. CRC
13. Shi W (2009) Principle of modeling uncertainties in spatial data and analyses. CRC
14. Kyriakidis PC, Shortridge AM, Goodchild MF (1999) Geostatistics for conflation and accuracy assessment of digital elevation models. Int J Geogr Inf Sci 13(7):677–707
15. Saalfeld A (1988) Conflation automated map compilation. Int J Geogr Info Syst 2(3):217–228
16. Safra E, Kanza Y, Sagiv Y, Doytsher Y (2006) Efficient integration of road maps. In: Proceedings of the 14th annual ACM international symposium on advances in geographic information systems. ACM, Arlington, Virginia, USA
17. Samal A, Seth S, Cueto K (2004) A feature-based approach to conflation of geospatial sources. Int J Geogr Inf Sci 18(5):459–489
18. Chen C-C, Knoblock C, Shahabi C (2006) Automatically conflating road vector data with orthoimagery. GeoInformatica 10(4):495–530
19. Brown LG (1992) A survey of image registration techniques. ACM Comput Surv 24(4): 325–376
20. Zitová B, Flusser J (2003) Image registration methods: a survey. Image Vis Comput 21 (11):977–1000
21. Walter V, Fritsch D (1999) Matching spatial data sets: a statistical approach. Int J Geogr Inf Sci 13(5):445–473
22. Filin S, Doytsher Y (1999) A linear mapping approach to map conflation: matching of polylines. Survey Land Info Syst 59(2):107–114

23. Doytsher Y, Filin S (2000) The detection of corresponding objects in a linear-based map conflation. Survey Land Info Syst 60(2):117–128
24. Olteanu Raimond A-M, Mustière S (2008) Data matching—a matter of belief. In: Headway in spatial data handling, pp 501–519
25. Li L, Goodchild MF (2010) Automatically and accurately matching objects in geospatial datasets. Adv Geo-Spat Inf Sci 10:71–79
26. Li L, Goodchild MF (2011) An optimisation model for linear feature matching in geographical data conflation. Int J Image Data Fusion 2(4):309–328
27. Huh Y, Kim J, Lee J, Yu K, Shi W (2014) Identification of multi-scale corresponding object-set pairs between two polygon datasets with hierarchical co-clustering. ISPRS J Photogrammetry Remote Sens 88:60–68
28. White MS, Griffin P (1985) Piecewise linear rubber-sheet map transformation. Am Cartographer 12:123–131
29. Saalfeld A (1985) A fast rubber-sheeting transformation using simplicial coordinates. Am Cartographer 12:169–173
30. Cobb MA, Chung MJ, F H III, Petry FE, Shaw KB, Miller HV (1998) A rule-based approach for the conflation of attributed vector data. Geoinformatica 2(1):7–35
31. Pottinger RA, Bernstein PA (2003) Merging models based on given correspondences. In: Proceedings of the 29th international conference on Very large data bases—vol 29. VLDB Endowment, Berlin, Germany
32. Noy NF (2004) Semantic integration: a survey of ontology-based approaches. ACM SIGMOD Record 33(4):65–70

Context-Aware Routing Service for an Intelligent Mobile-Tourist Guide

Alexander Smirnov, Nikolay Teslya and Alexey Kashevnik

Abstract The use of intelligent geographical information systems in the tourism industry allows for providing access to information about attractions during a tourist's trip. Navigation is a main concern for tourists. Attractions can be located far away from the tourist's current location, and the tourist usually reaches a desired attraction with the use of available transportation modes: on foot, by personal car, or on public transport. The present work describes a service for an intelligent mobile-tourist guide (TAIS) that supports routing between attractions using transportation modes inside the city. The tourist context is considered during the routing process. The service has been tested over the transport network of St. Petersburg, which is a Russian megalopolis with population of >5 million people.

Keywords Routing · IGIS · Multi-modal route · Tourism · Guide · Context

1 Introduction

Currently tourism is one of the fastest-growing economic sectors in the world. According to a report of the United Nations World Tourism Organization (UNWTO), the number of tourist arrivals increased by 4.6% (52 million people) in 2015 [1]. According to a UNWTO forecast, the number of tourists is expected to increase by an average of 3.3%/y during the period 2010–2030. The use of intelligent geographical information systems (GIS) in the tourism industry allows for

A. Smirnov (✉) · N. Teslya · A. Kashevnik
SPIIRAS, 14th Line, 39, St. Petersburg 199178, Russia
e-mail: smir@iias.spb.su

N. Teslya
e-mail: teslya@iias.spb.su

A. Kashevnik
e-mail: alexey@iias.spb.su

A. Smirnov · N. Teslya · A. Kashevnik
ITMO University, 49, Kronverksky Pr., St. Petersburg 197101, Russia

V. Popovich et al. (eds.), *Information Fusion and Intelligent Geographic Information Systems (IF&IGIS'17)*, Lecture Notes in Geoinformation and Cartography, DOI 10.1007/978-3-319-59539-9_18

providing access to information about attractions such as location, description, images, rating, and routes to the attraction. This information is used by tourists to plan a trip between attractions considering the context (tourist preferences, his or her available time, and transportation-related information). Systems that provide such functions are called an "intelligent tourist guide" and are usually developed as a distributed system with the user access point located on mobile devices (smart phone or tablet) and computational services on servers. Such systems can find and use more information about the user and his r her environment to provide cognitive assistance during the trip.

Routing is one of the main issues for a tourist during trip-planning. Available intelligent mobile-tourist guides use GIS modules that provide routing based on the pedestrian and personal car routes. However, attractions can be located far away from the tourist's current location, and a personal car may not be available for the tourist. In this case, the best option to reach an attraction is the use of public transport. Existing GIS provide information about public transport routes and can find routes based on this information. These routes can be multimodal, but usually they do not consider the existing public-transport timetables. This makes the trip-planning rather difficult because intelligent mobile-tourist guides cannot predict the waiting time for a public-transport vehicle. Therefore, the trip cannot be optimized in terms of time criteria.

This work presents a service for an intelligent mobile-tourist guide that allows routing between attractions using three modes of transportation: on foot, with a personal car, and on public transport inside the city and between cities. For public-transport routes, a current timetable is considered during the routing process. For this purpose, the service uses an approach to trip planning that is based on a multi-graph with dynamic edge weight for representing a public-transport network. Vertices of the multi-graph correspond to public transport stops, and edges correspond to routes between stops. Each edge's weight is calculated as the time needed to move along the edge according to the current situation on the road. The service allows the planning route in such a way as to minimize the whole trip time including transfer time.

The example presented in the paper is implemented based on the environment of Tourist Assistant (TAIS). It shows the work of the intelligent mobile-tourist guide with service for multi-modal route planning with timetable support over the public transport network of St. Petersburg, which is a Russian megalopolis with a population of >5 million people.

2 State-of-the-Art

Systems and studies in the field of context-aware routing can be divided into three main categories: (i) Personal-navigation systems are navigation services for pedestrians, vehicle drivers, and public transport passengers (e.g., Personal Travel Companion [2], Smart Travel Information Service (STIS) [3], Navitime [4]);

(ii) environments for integrated navigation resources (e.g., iTransit [5, 6], Highway Traveler Information System, Jiangsu Province of China [7], Arktrans [8]); and (iii) context/location-based platforms using communications over next-generation networks (e.g., LoL@ [9], MapWeb [10], OnRoute [11]). The most complicated systems are described below.

Personal Travel Companion includes personalized multimodal trip planning as well as management of multi-modal trips. The prototype provides multi-modal trip planning "from door to door" with a desktop or mobile-device application. The trip can integrate pedestrian, road-network, public transport routes and use. User settings (for example, preferred transport modes, mobility requirements, options, and time limits) are stored in the system for calculating personalized trips. The planned trip can be saved in Personal History of Trips.

Multi-modal extensions in car-navigation systems or personal assistants on user mobile devices provide information support during the trip. Continuous integral support is provided by mobile tools for multi-modal trip planning on a variety of devices. The response in real-time is not guaranteed, and its dependency on the context being considered only in terms of services-content adaptation to the user's start location, his or her preferences, and the availability of vehicles.

STIS offers travelers a multi-modal travel planning service aimed at bridging the gap in the coordination of existing transportation systems. A travel plan is created based on the preferences expressed by the end-users and integrates static and dynamic information about traffic jams and public transport.

STIS uses data provided by the iTransit environment for infomobility integration in order to provide users with personalized travel plans. The service is available on a personal computer or mobile phone. The context includes information about traffic and public transport, and the use of "user context" (e.g., his or her location) is not a major feature in the service because the service is focused on a trip-planning rather than traveler support during the trip. The information, in accordance with the mobility conditions, can be obtained in real time by the integration of traffic and transport monitoring through the iTransit environment. STIS does not support dynamic trip adjustment based on real-time information such as public transport delays and traffic jams.

The system provides a platform for combining services, thus allowing the provision of multimodal trip planning and assessing the current needs of the user and providing him or her with support in different ways in different situations. For example, when a user is in the car, alerts are provided in audio format instead of icons on a screen.

The presented systems allow the integration of multiple intelligent transportation systems that form a heterogeneous environment, which makes it possible to build multi-modal trips taking into account the current traffic situation. However, these systems require pre-processing of the route network, which imposes restrictions on using the context and adjusting the route dynamically.

3 Context-Aware Multi-modal Routing in Tourism

One of the main decisions made by tourists is "where to travel" [12]. This includes a destination as well as points of interests in the destination area. Although existing mobile tourist guides can help with recommending a destination and points of interests, tourists also need transportation support. Unfortunately, most mobile-tourist guides can only provide pedestrian or car routes to recommended destinations [13]. A general scheme for tourist navigation support can be proposed [14] based on the ideas of environments for the integration of state-of-the-art information services.

The scheme is presented on the Fig. 1. When a tourist checks a list of recommended attractions, he or she also finds available transportation to them. Most modern tourists would also like to get personalized trip recommendations that can satisfy his or her requirements. This can be provided by collecting information about the tourist's environment including information about tourist himself or herself: This information is called a "context".

This scheme assumes two scenarios of tourism travel. The first is individual tourism. For this scenario, it is reasonable to walk on foot to use local public transport to reach far points of interest. In special cases, a tourist may prefer to rent a car. The second scenario is group tourism, e.g., visiting attractions with family or friends. In this case, the tourist usually prefers to rent a car in order to be independent from the public-transport timetable. For short distances, it is reasonable to walk on foot. To support both scenarios, the tourist guide should provide at least three modules for navigation: pedestrian, car, and public transport.

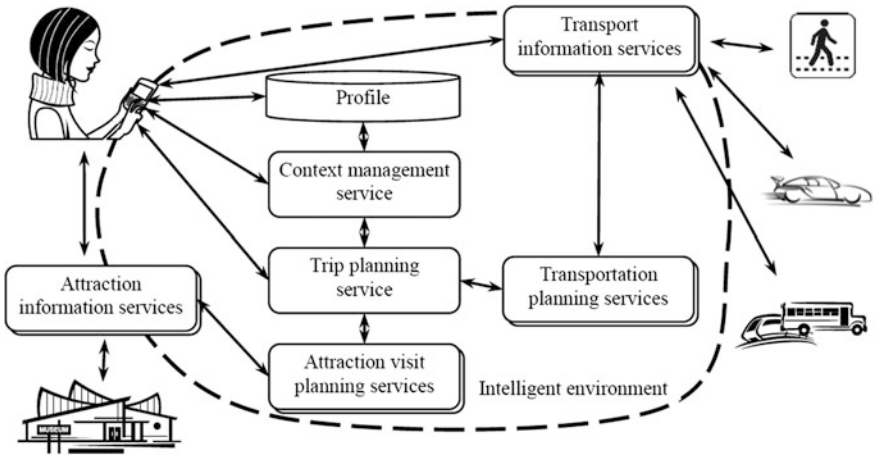


Fig. 1 Tourist-support scheme

3.1 Tourist Contextual Information

To support the tourist during the trip, the guide should constantly monitor the environment and the user status and react to their changes. The environment is described by the context (Fig. 2), which is usually divided on two classes: personal and environmental context [15]. Personal context is formed from the user profile using following elements:

- User type in the system (a tourist, a passenger, a driver, a fellow traveler, etc.). A tourist's type of transportation can change dynamically during the trip; therefore, the system should be adjusted to each type (e.g., use different notification types, show or hide different types of points of interest, etc.);
- User personal preferences. Attractions and transportation modes should be delivered based on the user's preferences: audio or video notifications, museums of a certain style, stores of certain brands and chains, etc.;
- User location. All information is gathered based on the user's current location and addressed to support him or her at the current point.

Various objects of the environment that are independent from the user form the environmental context:

- current time and time zone. The availability of specific vehicles, availability of visiting museums, shops, etc. that depend on time;
- public-transport location;
- weather conditions; and
- traffic status.

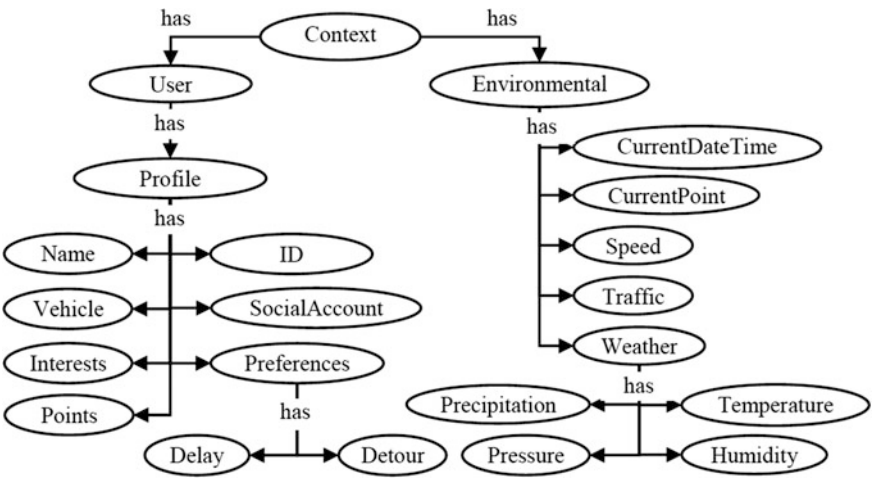


Fig. 2 Context structure

To provide timely support to the user, the gathering of contextual information should occur in real time without any additional actions from the user. This makes it possible to reduce the results waiting time and increase the responsiveness of the service. Services should track the traveler's evolving progress and gather additional information to offer services that are most likely to be useful in the current situation. This type of device behavior is called "proactive," and the support of a proactive mode is currently included in many information systems.

The contextual information is collected using sensors that are installed on user devices. This allows gathering of the physical parameters of the environment around the user. With the help of various GISs, the system can obtain information about different objects on the map, routes that can be planned using different modes of transportation, the weather forecast at specified locations, etc. Different social networks and information services provides ratings for various destinations.

3.2 *Multi-modal Routing on Public Transport with Context Awareness*

Tourists must find routes between their points of interest during a individual trip. The tourist assistant should provide support for any means of transportation: feet, car, or public transport. The foot or car routes are usually mono-modal, but a route on public transport can be multimodal. Moreover, all other types of transport can be used to transfer between public-transport stops. Multi-modal routing allows creating a greater interest in using public transport by travelers. Due to that, the number of vehicles in the streets can be reduced, and thus so can frequency and density of traffic jams.

The timetable of public transport, as well as the current situation on the roads, should be considered as a part of environmental context while planning a trip using public transport. Accounting for timetables allows estimating the waiting time for the vehicle in the case of a fixed and strictly abided timetable and making an initial trip plan. Accounting for the current situation refines the basic plan, thus allowing the estimation of waiting and transfer times at public-transport stops.

Information about city public-transport routes and timetables is usually distributed using the GTFS format (General Transit Feed Specification) [16] by the relevant local authorities. The timetable of public transport network in GTFS format can be represented as a directed weighted multi-graph $G(V, E, L)$, in which the vertices $V = \{v_1, \ldots, v_n\}$, $n \in N$ correspond to stops, and the edges $E = \{e_1, \ldots, e_m\}$, $m \in N$—are routes connecting these stops, $L = \{l_1, \ldots, l_k\}$, $k \in N$—weights of edges E (Fig. 3). The number of edges connecting the vertices corresponds to the number of routes between the respective stops.

In the case of transfer (public transport route changing), nodes are connected additionally with dummy edges to reflect the pedestrian or car routes. Each edges' weight can be defined by two parameters, time (l_t) or cost (l_c—ticket price), for

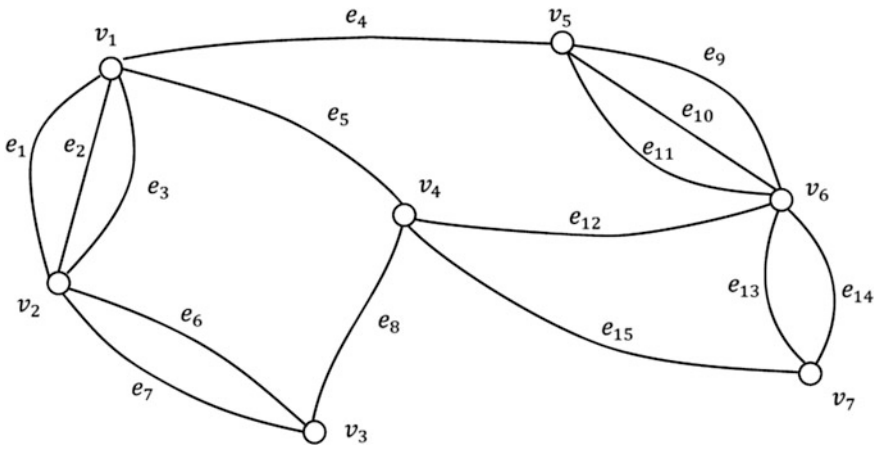


Fig. 3 Example of a multi-graph that represents a public-transport network

moving along the edge e_i. In both cases, the weights are dynamic, thus reflecting the current situation in the public-transport network.

With multi-modal route planning, the minimal distance criterion is important only if transfer requires the user to walk between transfer stops. In other cases, the following criteria for selecting the optimal trip can bee viewed: a minimal amount of travel time, a minimal cost, or a minimal number of transfers. However, a trip that meets one criterion may be inappropriate per the other criteria. Hence, it is needed to find a trip based on several criteria, thus solving the problem of multi-criteria optimization.

During trip-planning over the multi-graph, it is rather difficult to estimate how optimal a trip is accorded to a criterion due to the dynamic nature of the edges' weights. Therefore, it is hard to use the algorithm for a shortest-path search that uses heuristics such as like A* algorithm [17]. A modification of Dijkstra's algorithm [18] has been proposed to find the route through the multi-graph with dynamic weights based on the previously mentioned criteria. The choice of vertices at each step is carried out considering the type of vehicle, its route, and the timetable [19].

4 Case Study: Routing for a Tourist Attraction–Information Service

A tourist attraction–information service (TAIS) allows recommending attractions, which are better to be attended based on the tourist's preferences and the context information of the location region. The service allows the tourist to see a detailed description of the interesting attraction acquired from Internet sources (e.g.,

Wikipedia, Wikivoyage, Wikitravel, and Flickr). The recommendation system chooses an Internet source that provides a description of the interested attraction based on other tourists' ratings. TAIS includes a GIS based on data from the OpenStreetMap project. The GIS is based on the Mapnik render core and database in PostgreSQL, which stores OSM data for the area bounded by a rectangle. GIS allows map-rendering and object search. The GIS can be extended to provide a navigation support in the stored area.

For navigation purposes, the pgRouting library is used. It allows finding routes through the OSM data for pedestrian, bicycle, or car navigation. Raw data from the database must be converted to the topology of the road network presented by the weight-oriented graph before usage in pgRouting.

The topology is created by the osm2po library [20], which converts raw OSM/XML file to a database table. The conversion process extracts information about selected road classes; defines nodes (roads intersections, turns), which are converted to the graph nodes; and calculates edge weights. In addition, each edge is marked with the allowed maximal speed and offers the option of a reverse-traffic view.

Each entry in the topology database table contains an edge of the graph with the following characteristics:

- origin and destination ID;
- road class;
- path length in kilometers;
- average allowed speed;
- time spent to move along the edge; and
- backward cost (cost to move in a backward direction).

For routing along pedestrian and car routes, the A* algorithm is used. This algorithm was chosen due to the availability of a heuristic function that allows reducing the search area to speed up searching for the shortest path. The criteria "shortest path" can be set to distance or to travel time. The PgRouting function for the A* algorithm allows to set a graph orientation as well as a reverse-traffic view. In addition, the search area can be restricted by a new bounding box, and the search can be conducted over filtered road classes.

The filters allow selecting only those road classes accessible to the selected transport mode. For this purpose, the query for A* function can be extended by internal query for selecting only listed road types. Table 1 lists types of correspondence between road classes, class IDs in the database, and the transport types that were defined during the import process with use of the osm2po library [21].

4.1 Pedestrian Routing

In accordance to the classes listed in the Table 1, pedestrian filters should be configured to select roads of classes from 62 to 92. However, a pedestrian also can

Table 1 Road types in the OSM database

OSM class	Class in DB table	Transport type	Description
Motorway	11	Car	Highway and ramps
Motorway_link	12		
Trunk	13	Car	Inter-city routes, pass through the city
Trunk_link	14		
Primary	15	Car	State roads, city main highways, flyovers
Primary_link	16		
Secondary	21	Car	Regional roads, main highways of city areas, flyovers
Secondary_link	22		
Tertiary	31	Car, bicycle	Other main roads transit street, district transit streets, flyovers
Tertiary_link	32		
Residential	41	Car, bicycle	Roads inside residential areas
Road	42	Car, bicycle	Possible roads
Unclassified	43	Car, bicycle	Other district roads
Service	51	Car, bicycle	Service streets
Pedestrian	62	Pedestrian, bicycle	Streets and squares
Living_street	63	Car, bicycle, Pedestrian	Same as residential, except high-pedestrian priority
Track	71	Pedestrian, bicycle	Forest roads, unofficial Earth roads
Path	72	Pedestrian, bicycle	Paths
Cycleway	81	Bicycle	Bike path
Footway	91	Pedestrian	Pedestrian paths
Steps	92	Pedestrian	Stairs

pass roads from classes 31 the 51. Therefore, in navigation module for tourist-assistance pedestrian path is calculated through roads of classes 31 the 92. The routing function is configured for undirected graph usage with no additional reverse cost. At the left side of Fig. 4, a pedestrian route that has been built with the following query is shown:

```
SELECT * FROM pgr_astar(
'SELECT gid AS id, source::integer, target::integer,
length::double precision AS cost, reverse_cost, x1, y1, x2, y2
FROM ways WHERE class_id >= 31 and class_id <=92', 535, 44115, false, false)
```

In this query, the shortest-path A* function is used. In this function, the inner SQL selects the roads that will be used for routing; the first number is a start-point

Fig. 4 Route with restrictions for pedestrian navigation (*left*) versus car navigation (*right*)

ID; the second number is an end-point ID; and Boolean values set the road graph as directed with the edges having a reverse cost. The query returns 135 path points with a total distance of 8.4 km for pedestrian path.

4.2 Car Routing

Car routes have more restrictions compared with pedestrian routes because of stronger requirements for the roads. In accordance to Table 1, road classes from 11 to 51 as well as 63 can be selected for car navigation. However, there is no need to use service roads and living streets for a tourist's car navigation. Therefore, the tourist assistant is configured to use classes 11–43 for a route search. The right side of Fig. 4 shows a path with a restriction of roads types that cannot be traveled by a car. The routing function is configured for directed graph use with an additional reverse cost to take into consideration any one-way roads. A request that takes into consideration the listed road restrictions is presented below:

```
SELECT * FROM pgr_astar(
'SELECT gid AS id, source::integer, target::integer,
length::double precision AS cost, reverse_cost, x1, y1, x2, y2
FROM ways WHERE class_id <= 43',535, 44115, true, true)
```

A path has been found for the same origin and destination to show the difference between pedestrian and car routes. A request for car route returns a path with 112 points having a total distance of 8.8 km.

4.3 Public-Transport Routing

Information about public-transport routes is available in the OSM data, but there is no information about timetables. Currently most routes and timetables for public transport in the region are distributed in public GTFS format. This format contains files that simulate the structure of the required database tables. A full description of the format is presented in [16]. Due to its structure, the data in GTFS format can be imported into a database to allow a quick search of routes and additional information about them.

The city portal of public transport of St. Petersburg [22] provides information about public-transport routes in GTFS format. In addition to the timetable, the portal provides the current position of the vehicles, which can be used to predict the time of arrival of a vehicle to the selected stop. Routes for St. Petersburg have been imported to the PostgreSQL database of the TAIS routing module. In addition, the timetables for all routes have been imported to provide context for the routing module (Fig. 5).

A user usually selects a origin and destination on any place in map to find a route between them. Therefore, the first task of a multimodal route search is to find the nearest stops and create dummy edges between new vertices of the origin and destination and existing vertices of the multi-graph. The following SQL query is used for this task:

```
WITH closest_candidates AS (SELECT s1.stop_id, CAST (st_distance_sphere(s1.geom,
s2.geom) AS INT) AS distance, to_char(time '12:00:00' + (st_distance_sphere(s1.geom,
s2.geom)::int/1.389 || 'seconds')::interval, 'HH24:MI:SS')::varchar(255) AS walk_time
FROM stops as s1 LEFT JOIN stops as s2 on CAST (s2.stop_id AS
INTEGER) = current_id WHERE CAST (st_distance_sphere(s1.geom, s2.geom) AS
INT) < 300 ORDER BY s1.geom <-> s2.geom LIMIT 100); SELECT stop_id, distance
FROM closest_candidates WHERE distance < 200 ORDER BY distance;
```

This query selects vertices in the radius that the user agreed to walk for the public-transport stop (*Detour*) and calculates the distances and travel time from the new to the existing vertices, which is used as weights to the new edges. The query uses function *st_distance_sphere*(*geom1, geom2*) from the PostGIS extension,

Fig. 5 Public transport routes (*left*) and stops (*right*)

which calculates the distance between two geometries (here geometries are presented by points).

Each iteration of the modified Dijkstra algorithm [19] starts with the selection of a vertex with minimal travel time from the list of available graph vertices. Then the neighbor vertices that can be arrived by public transport or on foot are checked. Dummy edges with weights defined by distances and travel time between vertices are also created for vertices that can be reached on foot. This allows including the pedestrian parts of the route as new edges without additional algorithm modification. The path to each vertex is creating by including an edge with minimal weight and the edge's origin vertex. The resulting path is formed as a sequence of vertices and edges that the user should follow during the route. Vertices that are associated with public transport–route change or that are sources for a pedestrian route are marked as transfer vertices.

5 Evaluation

All evaluation tests were performed using the a virtual machine with a Debian 7.2 operating system running under the Hyper-V hypervisor. For the virtual machine, four cores of the Intel® Xeon(R) CPU E5620 2.40 GHz and 3 Gb DDR3 RAM were allocated. In addition, files of the virtual machine with a map database are held on RAID 1 to improve the read performance of the database.

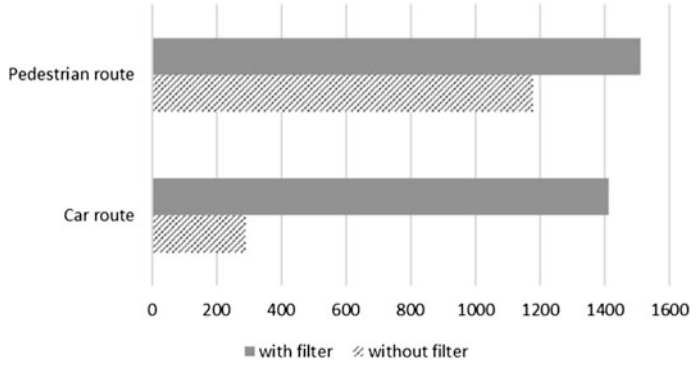


Fig. 6 Route-search time for available types of private transport

Routing functions work over the built topology and run rather fast. For example, approximately 1500 ms is required to the find route from the most northern point to the most southern point of St. Petersburg (the path consists of 363 points and has a total length of 31 km) with the A* shortest-path function. The path had been searched through the all types of roads including walk roads, footways, steps, and other road types that cannot be traveled by car.

The search time was decreased by 1.3 times for the pedestrian path and by 4.8 times for the car path after the filter configuration for each means of transportation was applied (Fig. 6). This difference can be explained by the decreased size of road graph. For St. Petersburg, the full road graph has 406,666 edges. Filters allow decreasing this number to 119,828 edges for the car-roads graph and 317,012 for the pedestrian-paths graph.

For public transport, the same analysis was carried out after the import of the public-transport routes and timetable data for St. Petersburg to a PostgreSQL database and creating the multi-graph. There are 6962 stops in the city, between which 965 routes run. The number of multi-graph edges used for route description is 19,773. The main difference with the road graph is the fact that a temporal component is used for timetable description. All routes include timetables; thus, the number of edges increases depending on the transport frequency of each route.

The performance of the used algorithm for public-transport routing was estimated by searching a fixed route with a gradual increase of area size from Vasilevskiy Island, where 1656 stops are located, to St. Petersburg, where (as previously mentioned) there are 6962 stops. At each of the stages, 100 measurements were carried out, and average value and error range were calculated. The analysis was performed based on the measurements defined by the 90th percentile (Fig. 7).

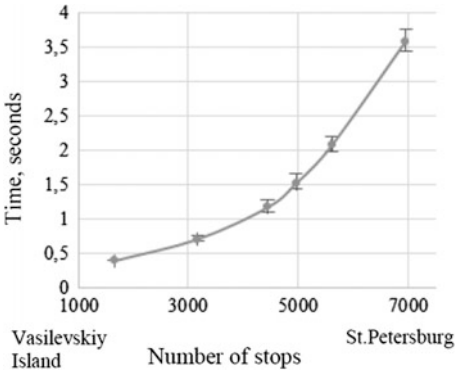


Fig. 7 Public-transport routing-time estimation

6 Conclusion

Support before and during travel is very important to tourists in new, unknown locations. Intelligent mobile-tourist guides can provide such support in an attraction search, which includes descriptions, photos, etc. The use context in a tourist-guidance systems allows providing personalized information about attractions as well as transportation between them. Contextual information is gathered from the user's device and from sources that are independent from the user. It allows defining user requirements in the particular situation, which contributes to high-quality information support.

Public transport is viewed as the main transportation type for a tourist to move between attractions. This choice is explained by its low cost, high speed, and availability in every region.

Routing service, as proposed in this chapter, unites three transportation modes—pedestrian, car, and public transport—to provide integrated support for a mobile tourist. Pedestrian and car routes are calculated based on the existing modules from the OpenStreetMap community. Road-type filters for each mode have been configured to provide a road network taking into consideration the chosen transportation mode.

The network of public-transport routes for user trip-planning was modeled with a multigraph. In the multi-graph, the vertices correspond to stops and the edges to the routes between stops. Each edge has a dynamic weight based on the available context values. A weight may represent the time of movement along the edge or a trip cost. To plan a route using the multi-graph, the modified Dijkstra's algorithm was used. Modifications allow to recalculate the weights of edges dynamically and to consider the presence of multiple routes between nodes and timetables of routes.

The tests performed showed that the service has acceptable efficiency for a megalopolis such as St. Petersburg. It provides routes for each transportation mode in an appropriate time, which depends on the available context and the size of a basic graph for each transportation mode.

Acknowledgements This work was supported by projects funded by Grants No. 15-07-08092, 15-07-08391, 16-07-00462, and 17-07-00327 from the Russian Foundation for Basic Research as well as Programs No. 0073-2015-0006 and 0073-2015-0007 from the Russian Academy of Sciences. This work was partially financially supported by the Government of Russian Federation, Grant No. 074-U01.

References

1. UNWTO (2016) UNWTO Tourism Highlights
2. Rehrl K, Bruntsch S, Mentz H-J (2007) Assisting multimodal travelers: design and prototypical implementation of a personal travel companion. IEEE Trans Intell Transp Syst 8:31–42
3. Brennan S, Meier R (2007) STIS: smart travel planning across multiple modes of transportation. IEEE Conf Intell Transp Syst Proceedings, ITSC, pp 666–671
4. Arikawa M, Konomi S, Ohnishi K (2007) Navitime: supporting pedestrian navigation in the real world. IEEE Pervasive Comput 6:21–29
5. Morenz T, Meier R (2008) An estimation-based automatic vehicle location system for public transport vehicles. IEEE Conf Intell Transp Syst Proceedings, ITSC, pp 850–856
6. Meier R, Harrington A, Cahill V (2005) A framework for integrating existing and novel intelligent transportation systems. IEEE Conf Intell Transp Syst Proceedings, ITSC 2005:650–655
7. Xiang QJ, Ma YF, Lu J, Xie JP, Sha HY (2007) Framework design of highway traveller information system of Jiangsu province in China. IET Intell Transp Syst 1:110
8. Natvig MK, Westerheim H (2007) National multimodal travel information—a strategy based on stakeholder involvement and intelligent transportation system architecture. IET Intell Transp Syst 1:102
9. Pospischil G, Umlauft M, Michlmayr E (2002) Designing LoL@, a mobile tourist guide for UMTS. Springer, Berlin Heidelberg, pp 140–154
10. Huang D, Liu F, Shi X, Yang G, Zheng L, Zhou Z (2006) MapWeb: a location-based converged communications platform. Bell Labs Tech J 11:159–171
11. García CR, Pérez R, Lorenzo A, Quesada-Arencibia A, Alayón F, Padrón G (2012) Architecture of a framework for providing information services for public transport. Sensors (Basel) 12:5290–5309
12. Nuraeni S, Arru AP, Novani S (2015) Understanding consumer decision-making in tourism sector: conjoint analysis. Procedia—Soc Behav Sci 169:312–317
13. Smirnov A, Kashevnik A, Shilov N, Teslya N, Shabaev A (2014) Mobile application for guiding tourist activities: Tourist Assistant—TAIS. In: Conf Open Innov Assoc Fruct IEEE Computer Society, pp 95–100
14. Smirnov A, Kashevnik A, Teslya N, Shilov N (2013) Virtual tourist hub for infomobility: service-oriented architecture and major components. ICEIS 2013 - Proc 15th Int Conf Enterp Inf Syst 1:459–466
15. Cheverst K, Davies N, Mitchell K, Friday A, Efstratiou C (2000) Developing a context-aware electronic tourist guide: some issues and experiences. Proc CHI2000 2:17–24
16. Google GTFS Static Overview. https://developers.google.com/transit/gtfs/. Accessed 10 Nov 2016
17. Hart P, Nilsson N, Raphael B (1968) A formal basis for the heuristic determination of minimum cost paths. IEEE Trans Syst Sci Cybern 4:100–107
18. Dijkstra EW (1959) A note on two problems in connexion with graphs. Numer Math 1: 269–271
19. Smirnov A, Teslya N, Shilov N, Kashevnik A (2016) Context-based trip planning in infomobility system for public transport. pp 361–371

20. Osm2po osm2po—openstreetmap converter and routing engine for java. http://osm2po.de/. Accessed 10 Nov 2016
21. OpenStreetMap Key:highway—OpenStreetMap Wiki. http://wiki.openstreetmap.org/wiki/Key:highway. Accessed 10 Nov 2016
22. St. Petersburg Public Transport Portal. http://transport.orgp.spb.ru/Portal/transport/main?lang=en. Accessed 10 Nov 2016

Geochronological Tracking: Specialized GIS-Analysis Tool for Historic Research

Yan Ivakin and Sergei Potapychev

Abstract Geo-information systems are widely applied in modern humanities. Such research studies are based on the use of geo-information technologies' universal functionality; however, there is an objective shortage of the specialized GIS-analysis tools intended for historic, ethnographic, and other research. Geo-chronological tracking gives us an example of the methodological and technological analysis tool specifically developed for solving a given class of historical problems. This chapter is dedicated to the analysis of the principle capabilities and specifics of such a GIS tool.

Keywords Geographic information systems · GIS technologies for historic research · Geo-chronological track · Modeling of geospatial historic process · GIS-based interdisciplinary research

1 Introduction

Today, geographic information systems (GIS) play the role of an effective analysis tool in humanities and, first and foremost, in historical research. However, the class of specialized methods and GIS tools intended for the intelligent support of researchers who are solving historical problems, and running computer simulations of various historical processes in geo-space, is somewhat insufficient. As a rule, such research studies are based on the use of GIS universal functionality, namely, on the use of geodesic, topographic, and universal geographic applications. Geo-chronological tracking gives us an example of specialized GIS-analysis tools

Y. Ivakin (✉) · S. Potapychev
Institute for Informatics and Automation of the Russian Academy of Sciences, Federal State Budgetary Institution of Science of St. Petersburg, 39, 14 Linia, V. O., St. Petersburg 199178, Russia
e-mail: ivakin@oogis.ru

S. Potapychev
e-mail: potapychev@oogis.ru

V. Popovich et al. (eds.), *Information Fusion and Intelligent Geographic Information Systems (IF&IGIS'17)*, Lecture Notes in Geoinformation and Cartography, DOI 10.1007/978-3-319-59539-9_19

aimed at solving the corresponding class of historical problems. The main form-factor of the proposed GIS-analysis tools is a mechanism of integration of chronological and geospatial data, in the form a geo-chronological track, that implements data methods presented in works [1, 2].

On the basis of an encyclopedic understanding of the "track" notion, that is, a number of points along the motion trajectory or chain of events, the Geo-Chronological Track could be interpreted as an aggregate of parameters (data) that describes series sequential events in the life of an individual (of a group, some historic community of people) bound to the time and location of these events' emergence. On the geographic map, such a track will be represented as a curve connecting the geographic points of the historical figure's (group's and other) whereabouts with color-gradient binding to the events' chronology [3]. Correspondingly, geo-chronological tracking is a procedure (method, process) meant for construction, generalization, and interpretation of the geo-chronological tracks' aggregate in accordance with a statistically significant social group, which allows to reveal new facts and regularities in the development of historic processes.

The development of algorithmic and software mechanisms for the construction and correct representation of geo-chronological tracks of some historical figures and members of small social groups on an electronic map, along with the integration of generalized historic and geographic information, form the core of this paper.

2 An Individual's Geo-Chronological Track as a Result of the Integration of Biographic and Geographic Information

Building of the geo-chronological track of an historic figure or a historic object based on the geospatial interpretation of their biographic information essentially is an integration of chronologic and geographic data in form of a graph that connects geographic points of the historical figure's (group's/other) whereabouts with color-gradient binding to given parameters of this individual's or historic events. At such, the nodes of such a graph have strong historical and geographic binding, and its arcs have a conditionally logical character. The essence of the above-described geo-chronological track concept is given in Fig. 1.

Development of the geo-chronological track of the historic figure (or object) can have a number of peculiar features in response to fragmented initial data. In this case, an adequate implementation of the proposed software and methodical analytical tools is provided by the combination of geo-information technologies and capabilities of modern geo-space/time–process development simulation modeling. For the mathematico-algorithmic and program realization of geo-chronological track-building in the form of a corresponding graph, under the objective conditions of fragmented historical and archival initial data, simulation-modeling methods allow to solve the following problems:

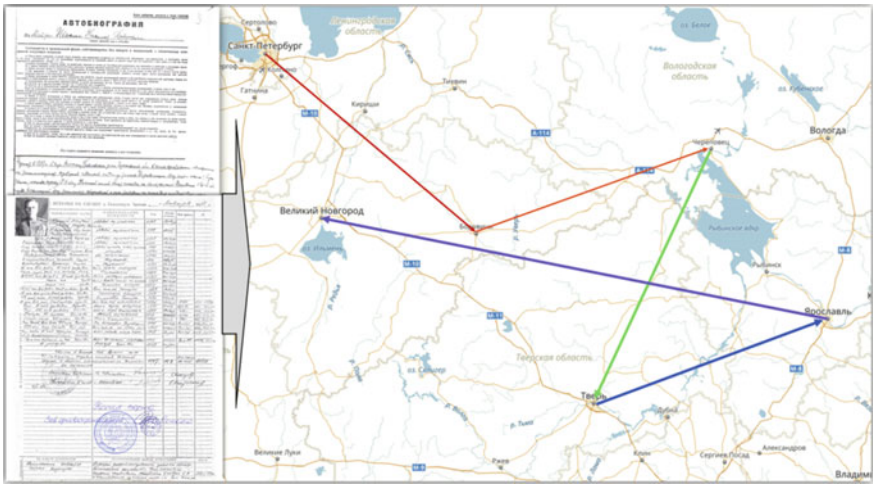


Fig. 1 An individual's geo-chronological track on a map

- probabilistic assessment and consideration of irregular movements of historic figures in time and space for track representation (mathematically, the absence of continuity and uniformity in increments of the historic figure or group);
- consideration of uncertainty and inaccuracy of the available historical data about the historic figures' movements in geographic space, of the location of given historical events represented as corresponding confidence intervals and confidence probabilities;
- consideration of the impact of changes in geographic space itself (landscape of historic processes' development) in time; and
- assessment of the impact of search specificity and preparation of initial historic and geospatial data necessary for track-building and a number of other similar problems.

Application of algorithmic and software mechanisms for building as well as correct representation of geo-chronological tracks in GIS for certain historical figures, members of small social groups, etc. allows to reduce the uncertainty and inaccuracy in historic knowledge while solving the following types of historical problems:

- determination of the encounter possibility, historic events relations, and others;
- detection and disavowal of historical falsifications; and
- rectification of computer reconstruction in the context of historic and geographic aspects, etc.

The process of accumulation and summation of geo-chronological tracks over a statistically significant sample of an individual's autobiographic data is of particular interest for researchers. Precisely, such a process is called "geo-chronological tracking."

3 Conceptual Model of Geo-Chronological Tracking

Geo-chronological tracking is understood as a process of accumulation and integration of data on a historical figure's (or object's) geographic movements over a given time period with as representation of the results in form of a generalized graph in GIS. As such, the nodes of such a generalized graph have strong geographic binding, and various characteristics of the arcs (color, thickness, shape, direction, etc.) represent the matching parameters of the historical figure's mass movements. Figure 2 schematically depicts, under the notation proposed in [4], a representation model of the arcs of a generalized graph.

Geo-chronological tracking allows to reveal, study, and visualize the hidden historic factors that concern the activities of the state, military, and other administrative agencies in the area of human-resources management as well as unobvious aspects of migratory, ethno-confessional, and other orientations. For instance, it can be used to trace the peculiarities and significant factors in the military-administration policy during recruitment arrangements for border military units in a pre-war period based on a statistically significant sample of service records. Also, an opportunity might be provided to analyze undocumented, although objectively existing, tendencies in repressive policy in the USSR during the 1930s to 1950s, etc.

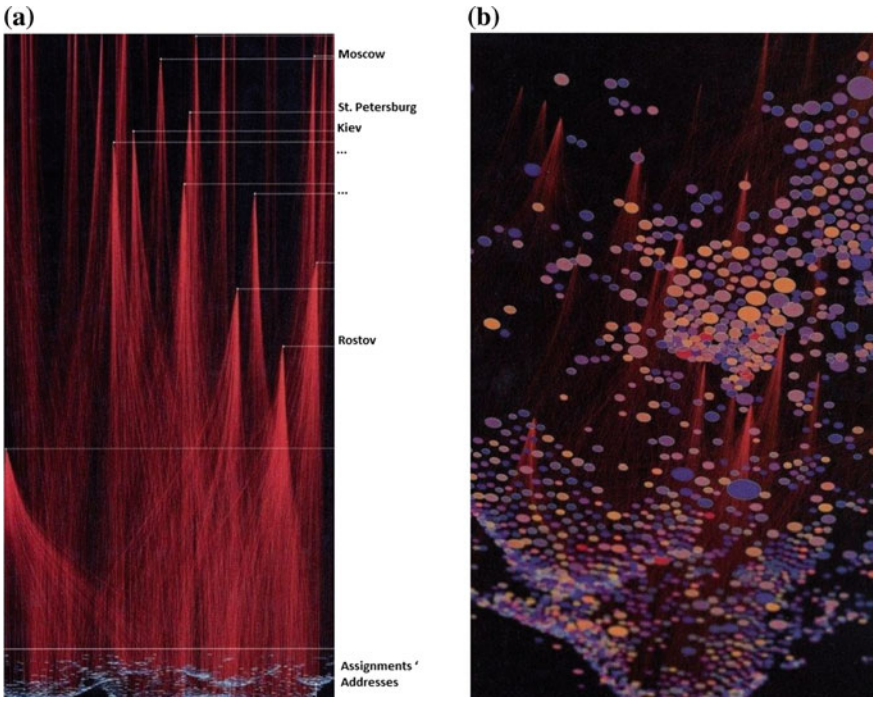


Fig. 2 Model of generalized graph's arcs (**a**) and the matching representation of geo-movement parameters (**b**)

Apparently, geo-chronological tracking, as a methodical research apparatus, is not of universal character; however, it provides a possibility of parameterization and development of nomenclature for undertaken research problems of historic and geographic spatial-processes analysis, i.e., a possibility of changing the process of acquisition and generalization of initial data with its further visual representation on map in a corresponding notation.

The subject of further research lies in identifying the limits of the proposed approach applicability as well as its boundary conditions, which is a challenge for both developers and potential users, namely, humanitarian researchers.

4 Generalized GIS Architecture for Geo-Chronological Tracking

Specialized GIS is the main analytical tool for the construction and analysis of geo-chronological tracks and for their generalization within the framework of tracking technology. The service-oriented architecture [1] allows to perform the most complete implementation of the geo-chronological tracking conceptual model. Figure 3 depicts the generalized service-oriented architecture of the proposed GIS in UML notation.

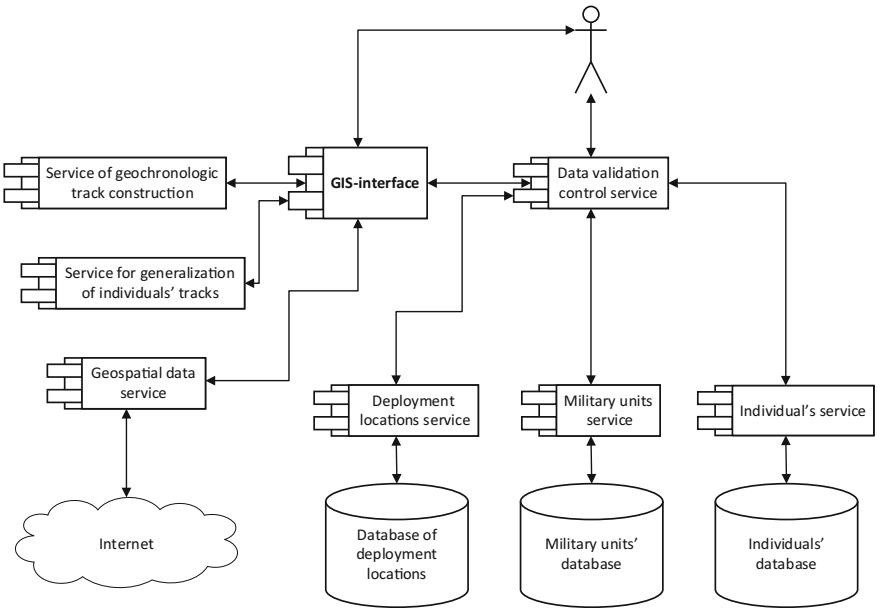


Fig. 3 GIS architecture of geo-chronological tracking

A geo-information system's interface (GIS interface) [1] is the main tool for the user's interaction with the software system. The GIS interface is a set of graphic-interface components that allow to perform system management through various information input devices. Such interface supports visualization and editing of data acquired from other GIS components in a user-friendly form.

GIS interface receives geospatial data through a corresponding service. The said service provides data as sets of raster files and sets of attributes for specific geospatial objects. All data are received by the server through the Internet, and OpenStreetMap (http://www.openstreetmap.org) is the main geospatial data set for the given service.

The data for geo-chronological track building are stored in the corresponding data bases; access to the data bases is realized through the set of services. Listing of the databases and services includes the following:

- database and service of military units deployment locations;
- database and service of military units; and
- database and service of individuals.

The data validation control service is intended for validation of the input data. In case the input data is incorrect, a special icon is displayed for the user suggesting these data correction.

The service of geo-chronological track building is the most important component of the given GIS. This service provides the following:

- building of geo-chronological tracks for different time intervals (months, years, centuries);
- adjustment of geo-chronological track visualisation (color, lines' thickness, etc.); and
- editing of geo-chronological tracks for better visual representation.

The service of generalizing individuals' tracks is necessary for the logical, mathematical, and visual representation of the generalized graph over the statistically significant sample of individual tracks within the studied historic (time) period.

5 Technology of Realizing Geo-Chronological Tracking

While considering the area of application of the proposed technology, the authors arrived at the conclusion that most representative information about migration of groups of population over long enough periods of time used to be stored in the military archives, so the decision was made to use these reliable sources for acquisition of the most accurate data. It became clear that geo-chronological tracking entails the following sequence of generalized steps of realization of its applied software functionality (drawn on the example of human resource management research in the military sector):

1. Initially, the user creates a listing of the human settlements with the military units' distribution in form of a separate database. Under the human settlement with the military units' distribution is understood the human settlement where the military unit headquarters were located, and in case of the headquarters absence, the place of the official distribution the unit's commander along with the group of assistant chiefs and assistants. At that, it is assumed that such human settlements are specified only once. The geographic location of the human settlement is specified by this settlement main post-office latitude and longitude (Table 1).

User detection of the data concerning a change in the status or in the name of the human settlement with static geographic coordinates requires corrections (additions) in the field "Human settlement name/Human settlement names used in the historic retrospective." The change of geographic coordinates with static name or status of the human settlement requires an introduction of a new entry in the field "Human settlement name/Human settlement names used in the historic retrospective" with a definition of the time period when these changes actually took place [e.g., City of Orenburg (1712–1728), currently City of Orsk]. The given table is not allowed to contain entries about human settlements with similar geographic coordinates. Fulfillment of this requirement is provided by an appropriate software function to control the accuracy of the entries in the table titled Military-Unit Deployment Locations.

The names of the human settlements that are not region centers should be entered under their full names (e.g., City of Dzhankoy, Crimea Region, Ukrainian SSR). The entered dates' accuracy equals 1 month.

To simplify the determination of geographic coordinates for the human settlement, an auxiliary function of these coordinated "chippings" from the user's work map is introduced.

2. In a separate modular window, a listing of those military units analyzed at the geo-chronologic tracks' construction is formed. The given sub-process also bears a character of introduction of the matching entries in a special tabular form as represented in Table 2.

Table 1 Structure and examples of entries in the table titled Military-Unit Deployment Locations

No.	Technical identificator	Human settlement name/human settlement names used in the historic retrospective	Location latitude	Location longitude
1.	12345HE	City of Orenburg/City of Chkalov	53 6′1 S	24 12′2 W
2.	54321A	City of Lepel, Vitebsk region, BSSR	33 6′1 S	28 12′2 W
…	…	…	…	…

Table 2 Structure and entries examples in the table titled Military-Unit Listing

No.	Military-unit name	Deployment	Deployment start date	Deployment end date
1.	44 cavalry regiment, 11 cavalry division, Turkestan military region	City of Orenbrg	June, 1919	May, 1925
		City of Lepel, Vitebsk region, BSSR	June, 1925	June, 1941
		City of Teheran, Iran	July, 1941	May, 1947
2.				
	…	…	…	…

Table 3 Structure and some examples of the entries in the table titled Individual's Geo-Chronological Track

Name: Ivakin, Nikolaj Petrovich		Date of birth: May, 1895	
Additional data …			
No.	Military unit name	Enrollment date	
1.	Krasnodar courses of the Workers' and Peasants' Red Army (WPRA) commanders	Month	Year
2.			
	…		

When identifying the listing of military units for their gradation, a military unit, such as Regiment, is taken. In the absence of the regiment membership (affiliation), the user should decide on the regiment equal or matching categories: military school, separate artillery battalion within a division, sub-division, air regiment, district hospital (at healing), and other.

The redeployment dates (change of deployment locations) are considered to have an accuracy of one month. Time gaps are not meant to exist in a chronology of the military unit redeployments. A time identifier, such as Up To Now, is permitted.

Military unit renaming, i.e., its reorganization, involves entering other units into the composition, and this is considered as the first unit's extinction followed by the emergence of a new unit bearing a new name.

Filling out the field Deployment in the military units listing is made through the picking units up from the list proposed by the available menu as determined by the table titled Deployment Locations of Military Units. Thus, if the military unit, over some significant time period, was deployed in a human settlement not listed in the table titled Deployment Locations of Military Units, the above-mentioned human settlement would be preliminary specified by the appropriate entry in the table.

3. The existence of structured data in the above-mentioned tables allows for the tabular formation of personal service records called the Individual's Geo-Chronological Track. The record format is given in Table 3.

The individual's (service person's) date of birth is filled out with an accuracy of one month. The field "Military-unit name" is filled out through the picking up from the list proposed by the available menu as determined by the table titled Military Unit Listing. In the field, the date of enrollment in the military unit has an accuracy that equals one month. At that, it is considered that a service record has a continuous character, and the new enrollment date is also a date matching the end date of the previous enrollment.

Based on the common data in the tables titled Military Unit Deployment Locations and Military Unit Listing, the user forms a statistically significant database out of individual service records titled Individual's Geo-Chronological Track.

Visual interpretation of such a database on the platform of GIS underlay exactly represents the essence of the software-tool operation as follows:

4. Initiation of construction and generalization of the individuals' tracks on the geographic map is performed by the separate request (i.e., clicking the appropriate virtual button). Result of the functionality realization for the software analysis tool of geo-chronological tracking is a geographic map with a mapped graph that generalizes the individuals' geo-chronological tracks, the service records of which are filed in the database. The nodes of such a graph are the military units' deployment locations, and the arcs are directed lines characterized by:

    - Thickness: Thickness indicates the number of enrollments-redeployments over a considered time period having the main gradation in accordance with the principle: 1 to ≥ 100 enrollments.
    - Color: Color indicates the middle age of people enrolled in the given directions during the analyzed time period with a smooth gradation in accordance with the principle of smooth change of the color gamut from warm colors to the cool ones where the light ellipse = 18 years and younger and the dark ellipse ≥ 55 years.
    - Arc-cut color: Arc-cut color indicates the generalizations of additional data preset by the user in the field of "Additional data" entries in the table titled Individual's Geo-Chronological Track, which is depicted by an example shown in Fig. 2.

Accounting for the fact that settings, as a rule, bear a bilateral character, visually the adjacent arcs between two nodes of a graph generalizing geo-chronological tracks are convex in regard to the shortest line connecting such nodes. The user has the possibility to receive numeric parameters, each of which determine a characteristic for each arc of the graph that generalizes the individuals' geo-chronological tracks in the emerging modal window by initializing the arc with a cursor.

6 Conclusion

Implementation of the methodical apparatus of geo-chronological tracking—in combination with modern intellectualization technologies [5], mathematical–statistical modeling aids applicable to the humanities [6–8], and information integration and fusion in GIS [9]—allows to assure a new quality of certain research related to humanitarian knowledge that incorporates history, ethnography, anthropology, and other disciplines.

Summing up the new capabilities demonstrated by the geo-chronological tracking as a result of combining the information-fusion methods and geo-information technologies, we can formulate quite a few directions of further development of this research-methodical apparatus and its applicability to historic research as well as of its software-data ware improvement as follows:

- determination of additional attributes (properties) that specify spatial displacements of historic individuals that can be reliably revealed, generalized, and visualized as parameters of the generalizing graph;
- study of the specifics in representation, by geo-chronologic tracking means, of sparse sampling of individual tracks for large time-intervals (>100 years) at the maximum possible geographic theater;
- analysis of the electronic-map scale's impact and the positioning accuracy for the individuals locations on geo-chronological trackings' effectiveness and representable appearance;
- development of ways and technological procedures aimed at implementation of software tools that realize geo-chronological tracking in integrated GIS-environments oriented to solving specific historic, ethnographic, anthropological, and other research problems;
- application of the mathematical apparatus of statistical estimation and simulation in order to reveal and increase the confidence level of the information received due to the use of geo-chronological tracking; and
- development of auxiliary- and service-information infrastructure of geo-chronological tracking such as test and service databases, specialized digital sets of electronic maps with a binding to a definite time interval, etc.

The approach that proposes the implementation of geo-chronological tracking research-methodical apparatus as a special purpose GIS-analysis tool for historic research, as proposed in this paper, will obviously benefit from the further development. However, the clarity of its realization and its high effectiveness allow to arrive at a tentative conclusion on its broad research and development applicability.

Acknowledgements This research was supported through the RFBS Project No. 16-07-00127.

References

1. Popovich VV et al (2013) Intelligent geographic information systems for the marine environment monitoring. In: Jusupov RM, Popovich VV (eds). Nauka, St. Petersburg (in Russian)
2. Ivakin YA (2015) Application of GIS technologies in historic and ethnographic research. In: Proceedings International Workshop Information Fusion and Geographic Information Systems IF&GIS'2015, Grenoble, France, 18–20 May 2015. pp 149–160
3. Ivakin YA, Ivakin VY (2013) New features of historical research using GIS technology of the information integration. Istoricheskaja informatika [Hist Inform] 4(6):62–71 (in Russian)
4. In world of science. Sci Am, 11 Nov 2015. pp 36–40 (in Russian)
5. Gavrilova TA, Muromcev DI (2008) Intelligent technologies in management: tools and systems, 2nd ed. Vysshaja shkola menedzhmenta, St. Petersburg (in Russian)
6. History & Mathematics: Political Demography & Global Ageing (2015) Yearbook. In: Goldstone JA, Grinin LE, Korotaev AV (eds). Uchitel Publishing House, Volgograd
7. Borodkin L (2015) Spatial analysis of peasants' migrations in Russia/USSR in the First Quarter of the 20th century. In: Proceedings of the 7th International Workshop "Information Fusion and Geographic Information Systems: Deep Virtualization for Mobile GIS (IF&GIS'2015)". Springer International Publishing, Switzerland, pp 27–40
8. Henke S (2015) Smarter Software Solutions 2015. Artificial Intelligence/History. http://www.stottlerhenke.com/ai_general/history.htm
9. Thill J-C (2011) Is spatial really that special? A tale of spaces. In: Proceedings International Workshop Information Fusion and Geographic Information Systems: Towards the Digital Ocean, Brest, France, 10–11 May 2011. pp 3–12