Goerlandt, Floris; Goite, Habtamnesh; Valdez Banda, Osiris A.; Höglund, Anders; Ahonen-Rainio, Paula; Lensu, Mikko

An analysis of wintertime navigational accidents in the Northern Baltic Sea

Published in:
Safety Science

DOI:
10.1016/j.ssci.2016.09.011

Published: 01/02/2017

Document Version
Publisher's PDF, also known as Version of record

Please cite the original version:

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.
An analysis of wintertime navigational accidents in the Northern Baltic Sea

Floris Goerlandt, a,⇑, Habtamnesh Goite, a,b, Osiris A. Valdez Banda, a, Anders Höglund, c, Paula Ahonen-Rainio, b, Mikko Lensu d

aAalto University, School of Engineering, Department of Mechanical Engineering, Marine Technology, Research Group on Maritime Risk and Safety, P.O. Box 12200, 00076 Aalto, Finland
bAalto University, School of Engineering, Department of Built Environment, Research Group of Geoinformatics, P.O. Box 15800, 00076 Aalto, Finland
cSwedish Meteorological and Hydrological Institute, Research Department, Norrköping 60176, Sweden
dFinnish Meteorological Institute, 00560 Helsinki, Finland

A R T I C L E   I N F O

Article history:
Received 3 April 2016
Received in revised form 5 July 2016
Accepted 20 September 2016
Available online 5 October 2016

Keywords:
Winter navigation operations
Accident data analysis
AIS data
Visual data mining
Maritime safety
Maritime risk

A B S T R A C T

Navigational accidents in wintertime conditions occur relatively frequently, yet there is little systematic knowledge available about the circumstances under which these occur. This paper presents an analysis of navigational shipping accidents in the Northern Baltic Sea area which occurred in the period 2007–2013. The analysis is based on an integration of various data sources, aiming to reconstruct the accident conditions based on the best available data sources. Apart from basic accident information from the original accident databases, data from the Automatic Identification System is used to obtain insight in the operation type during which the accident occurred, as well as into other dynamic aspects of the accident scenario. Finally, atmospheric and sea ice data is used to reconstruct the navigational conditions under which the accidents occurred. The analysis aims to provide qualitative insights in patterns and outlier cases in the accidental conditions. Correspondingly, visual data mining is selected as analysis approach, because of its utility in obtaining qualitative knowledge from data sources through a combination of visualization techniques and human interaction with the data. Special attention is given to the strength of evidence of the identified accident patterns. The results are primarily useful for improving risk analyses focusing on oil spill risks in winter conditions and for developing realistic training scenarios for oil spill response operations.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Maritime transportation is an important activity in the Northern Baltic Sea, with maritime trade of vital economic importance for several countries in the area. Harsh winter conditions result in sea environments with complex and dynamic ice. Together with winter darkness reigning over these northern areas, these condi-

Abbreviations: AIS, Automatic Identification System; DWT, deadweight; ECMWF, European Centre for Medium Range Forecasts; FMI, Finnish Meteorological Institute; FSWNS, Finnish-Swedish winter navigation system; GISIS, Global Integrated Shipping Information System; GT, Gross Tonnage; HFA3, Human Factors Analysis and Classification System; HIROMB, High Resolution Operational Model for the Baltic; IMO, International Maritime Organization; MESAN, Mesoscale Analysis System; MMSI, Maritime Mobile Service Identity; PCP, parallel coordinate plot; SMHI, Swedish Meteorological and Hydrological Institute; SoE, strength of evidence.

⇑ Corresponding author.
E-mail address: floris.goerlandt@aalto.fi (F. Goerlandt).
up-to-date information on ice conditions, operate in areas corresponding to the technical characteristics and receive timely and appropriate assistance by icebreakers (TraFi, 2010). For a concise description of these components, see Valdez Banda et al. (2015, 2016).

Understanding maritime accidents is an important aspect of improving maritime safety, and several authors have presented analyses to increase this understanding, e.g. Samuelides et al. (2008), Kum and Sahin (2015), Yip et al. (2015), Eliopoulou et al. (2016) and J. Zhang et al. (2016). Despite the importance of winter-time maritime transportation in the Baltic Sea, very little research has been dedicated to improving the understanding of wintertime accidents or their related risks in this area. Jalonen et al. (2005) presented a preliminary risk analysis of winter traffic, including a short description of selected number of winter accidents. Valdez Banda et al. (2015, 2016) performed a risk analysis of winter navigation in Finnish waters based on limited accident data and expert judgments. Their main finding is that while navigational accidents in ice typically lead to less serious consequences, especially ship collisions and groundings can lead to very serious consequences, particularly in relation to marine environmental pollution (Valdez Banda et al., 2016). The presence of the ice cover is an important reason for this, as it can be very challenging to retrieve oil spilled as a result from navigational accidents, as evidenced by the Runner-4 accident in the Gulf of Finland in 2006 (Wang et al., 2008). Therefore, a more elaborate risk management model, focusing on the probability of different oil spill sizes in collision accidents and the feasibility of different actions to reduce these risks is presented by Valdez Banda et al. (2015, 2016). Notwithstanding the above work, the understanding of wintertime accidents in the Baltic Sea area is very limited, both concerning the processes resulting in accident occurrence as well as the conditions under which these occur.

Therefore, this paper presents an analysis of wintertime navigational accidents. Given their comparatively high risks due to their potential for accidental large-scale oil spills, focus is on collisions and groundings. In particular, it is investigated under which conditions these accidents occur in Northern Baltic winter conditions. Such knowledge is useful for reducing uncertainty in risk analyses focusing on oil spill risks in winter conditions (Valdez Banda et al., 2016) and related response preparedness assessments (IMO, 2010), for increasing understanding and planning focused risk management actions (Valdez Banda et al., 2015), and for developing realistic and relevant training scenarios for oil spill response operations. The knowledge obtained on impact conditions in ship-ship collisions in ice conditions is useful for maritime waterway risk analysis, as also for open water conditions the related uncertainties are significant (Goerlandt et al., 2012).

The research questions addressed in this paper are as follows. What are the prevailing sea ice and meteorological conditions during the accidents? In which winter operation type has the accident occurred? For ship-ship collision accidents: what are the impact scenarios, i.e. under which angles, at what speeds and at which impact locations does the striking vessel impact the struck vessel? The remainder of this paper is organized as follows. In Section 2, the data sources used in the analysis are introduced and in Section 3, the analysis methods are described. Section 4 presents the results, while a discussion is made in Section 5. Section 6 concludes.

2. Data sources

In this section, the data sources used in the presented analysis are briefly described. The basic data source used is the North BAceD database, an integrated maritime accident database for the Baltic Sea area. As this database only contains consistent information about a limited number of basic characteristics of the accidents, additional data sources were added to this database to enable a reconstruction of the accident scenario based on the best available sources. In light of the research questions stated in the introduction, especially data from which the operation type and impact conditions can be obtained, and the prevailing environmental conditions, were needed. The utilized basic data sources are briefly introduced next. These were further processed into the database which is used for further analysis, as described in Section 3.

2.1. Maritime accident databases

Various challenges in using maritime accident databases have been reported. First, databases are known to suffer from various levels of underreporting, so that obtaining a “complete” image of the accidents in a given area cannot be achieved (Grabowski et al., 2009). Lützen (2002) addressed this issue in the context of accidents leading to water ingress. Second, different databases do not contain the same information and moreover apply different taxonomies of similar information types (Ladan and Hänninen, 2012). Third, several databases have changed their taxonomy over time, further complicating a consistent analysis over aggregated periods. Finally, databases suffer from missing and incorrect data, leading to gaps in the analysis and potentially to misleading results.

In order to obtain an as complete as possible picture of the accidents which occurred in the Northern Baltic Sea, the North BAceD database available from the Finnish Transportation Safety Agency is used. This is an integration of four maritime accident databases covering different geographical areas, all including the Baltic Sea. These concern the Lloyd’s Marine Intelligence Unit database, the HELCOM database, the DAMA database and the EMCIP database. Details about these, including a description of the included data fields, are provided by Ladan and Hänninen (2012). The method for integrating these is described by Mazaheri (2015).

The North BAceD database contains only very generic information about the accidents, so that e.g. detailed analyses of the progression of the accident over time, the systemic factors involved in the accident development or the crew actions and related human or organizational “errors” are not possible based on the available data. For knowledge acquisition about such factors, more in-depth analyses of specific accident investigation reports is needed. Hence, analysis of such factors is beyond the current scope.

Furthermore, while the accident databases distinguish different accident types, such as collision, grounding and sinking, none of the databases contain information about the winter navigation operation type during which the accident occurred. This information, however, is important for maritime risk analysis and management, because the oil spill risks associated with the different operation types differ significantly (Valdez Banda et al., 2016). The operation types can be identified using dynamic visualization of maritime traffic data, see e.g. Goerlandt et al. (2016). Details about the applied methodology to determine these are given in Sections 3.2 and 3.3.

Finally, the database does not contain consistent information about the environmental conditions. For instance, the EMCIP database does not contain this information, whereas in databases where the information fields are provided, different taxonomies and missing data made the recorded data very difficult to process in a meaningful way. For instance, the DAMA database applies standardized categories for wind direction, speed and sea state, but contains no information about sea ice conditions. The HELCOM database only distinguishes “ice-free” and “sea ice present” conditions. Having no other specific fields for environmental conditions,
use is sporadically made of the free text fields to provide a brief account of the sea state and visibility, often attributing these conditions as “causes” of the accident. However, the lack of uniformity of these free text fields render the descriptions given more or less useless in the context of the research questions stated in the introduction. For these reasons, it was decided to couple the accident cases with data from sea ice and weather data repositories, allowing uniform treatment of these environmental conditions.

2.2. Traffic data: the Automatic Identification System (AIS)

The 2002 IMO SOLAS Convention, Chapter V Regulation 19, mandates that most vessels over 300 GT on international voyages are to be equipped with a Class A type AIS transceiver. The data transmitted by this Automatic Identification System is known as AIS data. As an information exchange platform between vessels and shore organizations, AIS contains, amongst other, time-dependent data about the location, speed, course and navigational status of vessels.

The original purpose of AIS was collision avoidance but many other applications have since developed and proposed in the scientific literature. These include: ship surveillance, tracking and security (Ou and Zhu, 2008), collision avoidance and decision support (Mou et al., 2010), discovery of traffic patterns (Xiao et al., 2015), traffic simulation (Van Dorp and Merrick, 2011), ship routing development (Chen et al., 2015), near miss detection (W. Zhang et al., 2016), risk analysis (Qu et al., 2011), emission estimation (Jalkanen et al., 2014), ecological impact analysis (Merchant et al., 2012), and maritime spatial planning (Shermardine, 2015).

For the present study, full-rate AIS data from the period 01.11.2007 to 01.05.2013 was used, with data fields as shown in Table 1. The data was made available by the Finnish Meteorological Institute, based on an agreement with the Finnish Transport Agency regulating of access to historic AIS data for scientific research purposes. The data processing and visualization methods are presented in Sections 3.2 and 3.3.

2.3. Data related to sea ice and atmospheric conditions

The data for contextualising the accident cases in terms of sea ice and atmospheric conditions was obtained from different sources. Sea ice data was obtained from the High Resolution Operational Model for the Baltic (HIROMB) and the daily sea charts published by the Finnish Meteorological Institute (FMI). Cloud cover and wind conditions were taken from the Mesoscale Analysis System (MESAN), whereas visibility, precipitation and temperature were obtained from the PMP system. These data sources are briefly introduced below.

HIROMB (High Resolution Operational Model for the Baltic) is the operational ocean forecast model of the Swedish Meteorological and Hydrological Institute (SMHI) (Funkquist and Kleine, 2007; Wilhelmsson, 2002). It is a three dimensional baroclinic ocean model coupled with a Hibler-type sea ice model. It is used in several setups, those used in present paper is a 1 nautical mile resolution setup covering the Baltic Sea with boundary in Kattegat and two 3 nautical mile resolution setups with boundary in Skagerak and North Sea respectively. The model is run two times daily. Data assimilation of salinity, temperature and various ice parameters from in-situ measurements and satellite observations are done at the start of each run.

The gridded ice charts published by the Finnish Meteorological and Hydrological Institute (FMI) consist of data related to ice concentration and thickness, and ice type, and are used primarily for issuing ice reports as part of the Finnish-Swedish Winter Navigation System (TraFi, 2010). The charts are drawn daily by ice experts, making use of all available ice information (satellite images, icebreaker messages, observations), see Eriksson (2009). Charts are stored daily with a resolution of 1 nm.

MESAN (Mesoscale Analysis System) is a system for mesoscale analysis of selected meteorological variables (Häggmark et al., 2000). The analysis is performed with optimal interpolation of observations on a background field from the operational numerical weather forecast model HIRLAM (High Resolution Limited Area Model). The resolution is 0.1° (11 km) on a rotated latitude/longitude grid covering all of Scandinavia. The parameters in MESAN are typically those that are not directly available from other numerical weather prediction models or only available as diagnostic variables, e.g. wind at 10 m above sea level.

PMP (‘ProduktMonteringsParameter’) is a system at SMHI where the forecast meteorologist for each parameter selects a forecast from several models with possibility to edit the field. The result is the official forecast from SMHI. The input models vary over time and generally consists of the best local models and configurations (HIRLAM’s, AROME’s) and forecasts from European Centre for Medium Range Weather Forecasts (ECMWF). The resolution is 0.1° (11 km) on a rotated latitude/longitude grid with hourly output.

From the forecast products above, only data from the reference time immediately before the accident time were used.

2.4. Other data sources

Where the original North BACED database had missing parameters related to the vessel, use was made of two vessel information databases. IMO numbers (a unique identification number), ship type and vessel size were obtained from the Global Integrated Shipping Information System (GISIS) (IMO, 2016). MMSI numbers, needed to identify vessels in the AIS database as described in Section 2.2, were obtained from an online vessel tracking service provider (VesselFinder, 2016).

For categorizing the sea area where the accident occurred, use was made of the position as recorded in the North BACED database and/or as obtained from the videos made from AIS data as described in Section 3.2 and 3.3. This position was compared with sea charts, from which the accident position was classified either as port area, archipelagic waterway or open sea.

Finally, for determining the light conditions, which relates to the sun position relative to the horizon, use was made of the position and time of the accident. The location was obtained from the North BACED database and/or the videos made from the AIS data, used in several configurations for operational weather forecasting at the Swedish Meteorological and Hydrological Institute (SMHI).

Table 1

<table>
<thead>
<tr>
<th>AIS field</th>
<th>Unit</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMSI number</td>
<td>[–]</td>
<td>A 9-digit code uniquely identifying a vessel</td>
</tr>
<tr>
<td>Time stamp</td>
<td>[s]</td>
<td>Time at which the message is recorded, format: yyyy-mm-dd hh:mm:ss</td>
</tr>
<tr>
<td>Position</td>
<td>[–]</td>
<td>Longitude and latitude of transmitted message, in WGS-84 coordinate system</td>
</tr>
<tr>
<td>Ship type</td>
<td>[–]</td>
<td>A 2-digit code identifying the type of vessel, see USCG (2012)</td>
</tr>
<tr>
<td>Ship length and width</td>
<td>[m]</td>
<td>Dimensions from bow to stern and side to side, see USCG (2012)</td>
</tr>
<tr>
<td>Ship speed</td>
<td>[kn]</td>
<td>Speed over ground</td>
</tr>
<tr>
<td>Ship course</td>
<td>[°]</td>
<td>Course over ground, relative to true north. This is the direction in which the ship is moving, relative to true north</td>
</tr>
<tr>
<td>Ship heading</td>
<td>[°]</td>
<td>Direction in which the bow of the ship is pointing</td>
</tr>
</tbody>
</table>

1 HIRLAM is a hydrostatic numerical weather prediction model (Undén et al., 2002) used in several configurations for operational weather forecasting at the Swedish Meteorological and Hydrological Institute (SMHI).
2 AROME is a small scale numerical prediction model, developed by Meteo-France (Seyt et al., 2010).
3. Analysis methods

The overall process of constructing the integrated database and analysing the wintertime navigational accidents is shown in Fig. 1. The five steps, as well as the integrated database “North BAceD+” are described below.

3.1. Step 1: filtering and cleaning data

The original North BAceD accident database consists of data from four different databases, as outlined in Section 2.1. Of the records in these databases, only those occurring during winter months in the period November 2007 to May 2013 in the Northern Baltic Sea area are retained. The time period was selected because of the limitations of the AIS data, which was available only from this period. Winter months are defined here as November to May, because during these months, there may be ice present in the study area (Riska et al., 2007). The Northern Baltic Sea is defined here as the sea areas composed of the Gulf of Bothnia and the Gulf of Finland, as defined by HELCOM response, see Fig. 2. These areas were selected because they are totally ice-covered during normal winter conditions and because maritime accidents leading to large oil spills in these areas would require sub-regional cooperation in oil spill response between the relevant contracting parties of the Helsinki Convention (HELCOM, 2015). Thus, other potentially ice-covered areas in the Baltic Sea are outside the current scope. Within the study area, all accidents are considered, irrespective of whether or not there was ice present at the accident location at the accident time. Thus, the analysis concerns accidents during the winter season, rather than accidents in ice conditions.

Further filtering was performed based on the accident type. As the focus is on high-risk accidents from the viewpoint of large-scale accidental oil pollution, only collisions (including both ship-ship collisions and collisions with objects) and groundings were retained for further analysis, based on a previous risk analysis indicating that only these accident types have the potential for large-scale pollution (Valdez Banda et al., 2016).

Subsequently, the data was cleaned and missing values updated in the database. Based on the IMO number in the North BAceD database and information in other vessel databases, MMSI numbers and data related to ship type and main ship dimensions were added. This was done for about 30% of the cases. Cases with missing IMO numbers and missing coordinates, leaving no information to link accident conditions to AIS data, were not retained in the further analysis.

3.2. Step 2: constructing videos

The construction of the videos of the accident cases required the AIS data to be processed. A distinction was made between two

---

Footnote:

3 In this paper, twilight includes both civil and nautical twilight. Civil twilight occurs when the sun is less than 6° below the horizon, nautical twilight when the sun is less than 12° below the horizon.
cases, due to the incomplete records in the North BAceD database:

- **Type-I video**: ship identity (IMO and MMSI number) is known for the accident case. The video shows the vessels’ contours, speed vectors and dynamic data. This video type dynamically tracks the vessel involved in the accident, as outlined below for the example case in Video 1.
- **Type-II video**: ship identity (IMO and MMSI number) are unknown for the accident case, but the accident location is known. The video dynamically tracks the vessels in the vicinity of this location, showing the vessels’ contours, speed vectors and IMO numbers.

The steps to construct the videos are shown in the flowchart of Fig. 3, and are briefly outlined below.

**Step I.** All AIS data is grouped by ship (using the MMSI number) and chronologically sorted. This results in trajectories of each vessel over the considered time period. Trajectories are constructed for the entire calendar day on which, according to the North BAceD database, an accident occurred.

**Step ii.** The AIS data is resampled per ship to obtain the position, speed and course at equal time instances, using a 10 s time interval. This step is necessary because the transmission rate of AIS messages depends on the navigational status, the speed and turn of the vessel, see USCG (2012). However, to determine the impact conditions in collision accidents, simultaneous positions of the vessels are needed.

**Step iii.a.** If the identity (MMSI number) of the vessel(s) involved in the accident is unknown, the AIS data of the vessels in the vicinity of the given accident location is collected for each time step. An inspection domain with adjustable size is used, with default value of 2NM.

**Step iiib.** Using the data from Step iii.a., a Type-II video is created and the vessel(s) involved in the accident is/are identified.

**Step iv.** When the identity (MMSI number) of the vessel(s) involved in the accident is known, the AIS data of the vessels in the vicinity of the given vessel is collected for each time step. An inspection domain with adjustable size is used, with default value of 2NM.

**Step v.** Using the data from Step iv., a Type-I video is made for all accident cases. This video type is used for all analyses of the winter navigation operation type and, if applicable, the impact conditions.

Video 1 shows a Type-I visualization of the AIS data of a selected accident case, where following elements are shown.

**A. Dynamic data.** The time and location of the accident case indicates when and where it occurs. This is important information to link the accident with values for atmospheric and sea-ice conditions, as time and location are needed to search through the databases described in Section 2.3.

**B. Vessels involved in the accident.** Central in a 2 NM inspection domain delineated by concentric circles as shown in Video 1, the ships involved in the accident are shown by a contour at their instantaneous positions. A speed vector gives an indication of the projected positions of the vessels in a time window of 1 min, to facilitate identification of their dynamics.

**C. Vessel data.** Data of the vessel(s) involved in the accident is shown. Static data include the MMSI number, ship type and size (tonnage and main dimensions). Dynamic data includes the speed, course over ground and heading of the vessel(s), which are important parameters to describe the impact conditions.

Video 1 shows a bow-stern collision between two cargo vessels in a convoy in the Bothnian Bay. A third vessel ahead of the two vessels involved in the accident is seen to slow down around 06:00, upon which the centrally shown gradually slows down and comes to a near stand-still around 06:04. The following vessel also slows down but cannot avert a collision, which occurs at 06:05.

### 3.3. Step 3: analysing videos

For cases where the accident location is known, but the IMO number was missing, Type-II video was created and inspected. If an accident was apparent in the video, e.g. when two vessels come in contact, the IMO numbers were recorded and Type-I videos made, which were used for further analysis.

For each accident, the video was inspected and several parameters were recorded which are relevant in describing the accidental situation. First, the exact time of the accident was determined, as in the North BAceD database, for most accidents, only the date was provided. Second, the exact position of the accident was determined, as for many cases, the coordinates provided in the North BAceD database proved to be erroneous (often significantly). The authors can only speculate as for the reasons for these errors: they may occur e.g. due to poor recording standards, lack of procedures for quality assurance of data entries, or simply through typographical errors. Third, the winter navigation operation type is determined, distinguishing six categories based on Rosenblad (2007): independent navigation, towing, escorting, convoy, double convoy and cutting loose. One additional category (drifting) was identified based on the videos of the accident cases. These are defined in Table 2, indicating which characteristic patterns were looked for to determine the operation type from the videos. The authors have gained experience in identifying the operation types from the visualizations, based on earlier work where similar visualizations were applied, see Goerlandt et al. (2016). Fourth, for ship-ship collision

![Fig. 3. Sequence for constructing accident videos based on AIS data.](image-url)
Table 2
Winter navigation operations: definition and characteristics, Rosenblad (2007).

<table>
<thead>
<tr>
<th>Operation</th>
<th>Definition (D.) and characteristics (C.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent</td>
<td>D. The vessel navigates through the water or ice fields without icebreaker assistance</td>
</tr>
<tr>
<td>navigation</td>
<td>C. The video shows no icebreaker in the vicinity of the vessel</td>
</tr>
<tr>
<td>Towing</td>
<td>D. The assisted vessel is towed by an icebreaker as it cannot follow the icebreaker in an escort or convoy operation because of high ice pressure or because the channel has too much slush ice</td>
</tr>
<tr>
<td></td>
<td>C. The vessel appears closely behind the icebreaker, with same speed and course. Towing is preceded and followed by a period when the vessels are stationary for connecting/disconnecting the tow</td>
</tr>
<tr>
<td>Escort</td>
<td>D. The assisted vessel follows an icebreaker at a close distance, in a channel broken by the icebreaker</td>
</tr>
<tr>
<td></td>
<td>C. The icebreaker navigates with one vessel (disregarding a possibly towed vessel) following the trajectory of the icebreaker in close distance</td>
</tr>
<tr>
<td>Convoy</td>
<td>D. Two or more assisted vessels follow an icebreaker at close distance, in a channel broken by the icebreaker</td>
</tr>
<tr>
<td></td>
<td>C. The icebreaker navigates with two or more vessels (disregarding a possibly towed vessel) following the trajectory of the icebreaker in close distance</td>
</tr>
<tr>
<td>Double convoy</td>
<td>D. Two or more assisted vessels follow two icebreakers at a close distance, in a wide channel broken by the two icebreakers</td>
</tr>
<tr>
<td></td>
<td>C. Two icebreakers, the second ahead or behind but slightly abeam the first, assist a vessel which typically has a significantly larger width than the icebreakers</td>
</tr>
<tr>
<td>Cutting loose</td>
<td>D. A vessel which has become beset in ice, e.g. due to compressive ice, is assisted by an icebreaker which breaks the nearby ice, releasing the ice pressure</td>
</tr>
<tr>
<td></td>
<td>C. The assisted vessel is stationary and the icebreaker moves back and forth and/or around the vessel in close proximity, often accompanied with compressive ice conditions</td>
</tr>
<tr>
<td>Drifting</td>
<td>D. A vessel is moving, not due to its own propulsion system but due drifting currents or sea ice</td>
</tr>
<tr>
<td></td>
<td>C. A vessel moves at a low speed, where course and heading are not necessarily aligned</td>
</tr>
</tbody>
</table>

accidents, some parameters about the impact scenario were recorded. These are the speed of the vessel(s) at impact and for ship-ship collision accidents, the location of impact and the impact time. The sea area classification was determined using sea charts (see Section 2.1), the dataset was relatively small. In total, there were 45 accident cases with each 32 data attributes as shown in Table 2, where it is noted that the operation types ‘towing’ and ‘double convoy’ of Table 2 are not retained as no such accidents were identified in the database. No draught information was available in the data sources.

3.4. Step 4: integrating data

The information obtained from the video analysis was incorporated in the accident database. Furthermore, for each accident, the weather and sea ice data was extracted from the databases described in Section 2.3, based on the accident coordinates and time. The sea area classification was determined using sea charts as outlined in Section 2.4. The light conditions, i.e. whether there was daylight, night or twilight, are determined based on the accident coordinates and time, using an algorithm described by Seidelmann (1992). The integrated database applied for further analysis is referred to as North BAcE+D. The data fields in this database are summarized in Table 3, where it is noted that the operation types ‘towing’ and ‘double convoy’ of Table 2 are not retained as no such accidents were identified in the database. No draught information was available in the data sources.

3.5. Step 5: analysing data

After constructing the integrated database described in Section 3.4, it was analysed. Due to the relative infrequent occurrence of navigational accidents in the area and several uncertainties relating to the accuracy of the original North BAcE+D dataset (see Section 2.1), the dataset was relatively small. In total, there were 45 accident cases with each 32 data attributes as shown in Table 3. In cases where more accident data is available, it may be feasible to construct mathematical models to obtain insights, e.g. by learning Bayesian Networks (Hänninen and Kujala, 2014; J. Zhang et al., 2016) or by developing regression models (Yip, 2008).

In case of smaller datasets, a more feasible approach is to use visual data mining techniques to explore patterns and outlier cases. Visual data mining takes benefit of the high capacity of human visual perception. Humans are visual thinkers and efficiently identify patterns, trends, anomalies and outliers in what is seen. As such, visual representations extend and amplify our cognitive capacity, and support understanding, reasoning and decision making (Card et al., 1999; Keim et al., 2008). During the data mining process, data visualization assists in forming hypotheses as well as answering specific questions about possible patterns in the data. Large amounts of data can be presented compactly in a visual form, and even smaller data sets can reveal interesting features. Such techniques have been used in earlier work concerning maritime accidents, e.g. for investigating the relation between weather conditions and fishing vessel incidents (Wu et al., 2009) and for obtaining insights in recreational boating accident patterns (Sonninen and Goerlandt, 2015).

Different visualization techniques, implemented in software packages to enable user interaction with the data, combined with human cognition and insight in the nature of phenomenon described in the dataset, are used to reveal different aspects of the data. Conventional statistical graphics, such as bar chart, box plot and scatter plot, can be used for analysing distributions of individual variables or pair-wise correlations. However, multivariable data needs more advanced techniques that can reveal more complex patterns and relationships. These techniques, such as the parallel coordinate plot (PCP), are interactive, their effectiveness depending on the actions of and interpretations made by the user.

The parallel coordinate plot (Inselberg, 1997, 1985) organizes axes of multiple variables in parallel and thus enables access to a large number of variables concurrently. Each data object is presented as a multiline crossing the axes according to its values for the corresponding variables. PCP is a useful method for identifying relationships between groups of similar data objects across a range of variables. However, as the pattern depend on the order of axes, reordering them is an essential part of interaction and the knowledge construction process. In a static presentation, such as this paper, only snapshots of what a PCP analysis reveals can be presented. To more clearly distinguish certain features of interest in the data, different colors can be applied to data subsets. In the current analysis, use is made of the Orange software package (Demsar et al., 2013). For a description of how a PCP diagram is constructed, its mathematical basis and guidance on its interpretation, the reader is referred to Inselberg (2009).

4. Results

4.1. Patterns relating to sea ice conditions

Fig. 4 depicts a parallel coordinate plot (PCP) showing the patterns in the relations between operation types, sea area, sea ice
conditions, and accident types. First, it is noteworthy that no accidents are found to occur in the ‘double convoy’ and ‘towing’ operation types, defined in Table 2. It is seen that groundings and collisions with an object occur during independent navigation, whereas ship-ship collisions occur in all operation types but primarily in escort, convoy and cut loose operations.

Groundings occur in port areas and archipelagic waterways, typically in ice-free conditions or with thin ice in low concentration. Ship-ship collisions show a very different pattern, occurring in all sea areas but primarily in open sea and archipelagic waterways. Similar patterns are found for escort, convoy and cut loose operations. Accidents occur only in ice conditions with high ice concentration and ice thicknesses upwards of 20 cm. Ridging conditions vary significantly, but slight ridging with ridge densities of 1 ridge per km and ridge heights of about 1.5 m are predominant. Snow depths below 0.5 m are typical, but also here large variations occur.

One collision with an object occurred in ice free conditions in the open sea area, but this accident type is more commonly associated with port areas under typically rather moderate ice conditions. Most commonly, level ice conditions are found, which in port areas means there is fast ice, i.e. fastened to the coastline or to the sea floor.

4.2. Patterns relating to atmospheric conditions

Fig. 5 depicts a PCP showing the patterns in the relations between operation types, sea area, and various atmospheric conditions relating to precipitation, visibility and temperature. The plot shows that visibility conditions and light conditions show similar patterns for all accident types. Visibility conditions range from moderate to excellent, with very good and excellent visibility most frequent. More accidents occur during daylight conditions (21 cases) than during night (16 cases), which is somewhat surprising.

Table 3

Summary of data in the constructed North BAceD + database.

<table>
<thead>
<tr>
<th>Data field</th>
<th>Unit</th>
<th>Source</th>
<th>Categories/Additional notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>General accident characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td>[–]</td>
<td>NDB</td>
<td>Format: dd.mm.yyyy</td>
</tr>
<tr>
<td>Time</td>
<td>[–]</td>
<td>AIS-V</td>
<td>Format: hh:mm:ss</td>
</tr>
<tr>
<td>Location</td>
<td>[–]</td>
<td>NDB, AIS-V</td>
<td>WGS-84 coordinates</td>
</tr>
<tr>
<td>Sea area</td>
<td>[–]</td>
<td>AIS-V, SC</td>
<td>(Port area, Archipelagic waterway, Open sea)</td>
</tr>
<tr>
<td>Accident type</td>
<td>[–]</td>
<td>NDB, AIS-V</td>
<td>[Stranding/grounding/bottom contact, Collision with an object, Collision with other vessel]*</td>
</tr>
<tr>
<td>Operation type</td>
<td>[–]</td>
<td>AIS-V</td>
<td>[Independent navigation, Escort, Convoy, Cut loose, Drifting]</td>
</tr>
<tr>
<td>Vessel characteristics (in case of collision with other vessel: struck ship)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>[–]</td>
<td>NDB</td>
<td>(Work vessel, Tanker, Bulk carrier, Cargo vessel, RoPax, Icebreaker)</td>
</tr>
<tr>
<td>IMO number</td>
<td>[–]</td>
<td>NDB, VDB</td>
<td></td>
</tr>
<tr>
<td>MMSI number</td>
<td>[–]</td>
<td>VDB, AIS-V</td>
<td></td>
</tr>
<tr>
<td>Ship type</td>
<td>[–]</td>
<td>VDB, AIS-V</td>
<td></td>
</tr>
<tr>
<td>Gross Tonnage</td>
<td>[–]</td>
<td>VDB</td>
<td></td>
</tr>
<tr>
<td>Deadweight</td>
<td>[tonnes]</td>
<td>VDB</td>
<td></td>
</tr>
<tr>
<td>Length</td>
<td>[m]</td>
<td>VDB, AIS-V</td>
<td></td>
</tr>
<tr>
<td>Breadth</td>
<td>[m]</td>
<td>VDB, AIS-V</td>
<td></td>
</tr>
<tr>
<td>Vessel characteristics (only in case of collision with other vessel: striking ship)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>[–]</td>
<td>NDB</td>
<td>(Work vessel, Tanker, Bulk carrier, Cargo vessel, RoPax, Icebreaker)</td>
</tr>
<tr>
<td>IMO number</td>
<td>[–]</td>
<td>NDB, VDB</td>
<td></td>
</tr>
<tr>
<td>MMSI number</td>
<td>[–]</td>
<td>VDB, AIS-V</td>
<td></td>
</tr>
<tr>
<td>Ship type</td>
<td>[–]</td>
<td>VDB, AIS-V</td>
<td></td>
</tr>
<tr>
<td>Gross Tonnage</td>
<td>[–]</td>
<td>VDB</td>
<td></td>
</tr>
<tr>
<td>Deadweight</td>
<td>[tonnes]</td>
<td>VDB</td>
<td></td>
</tr>
<tr>
<td>Length</td>
<td>[m]</td>
<td>VDB, AIS-V</td>
<td></td>
</tr>
<tr>
<td>Breadth</td>
<td>[m]</td>
<td>VDB, AIS-V</td>
<td></td>
</tr>
<tr>
<td>Sea ice conditions</td>
<td></td>
<td>IC</td>
<td>(No ice, Level ice, Slightly ridged ice, Ridged ice, Heavily ridged ice)</td>
</tr>
<tr>
<td>Ice type</td>
<td>[–]</td>
<td>IC</td>
<td></td>
</tr>
<tr>
<td>Ice concentration</td>
<td>[%]</td>
<td>HIROMB</td>
<td>Fraction of area covered with sea ice relative to total area</td>
</tr>
<tr>
<td>Level ice thickness</td>
<td>[m]</td>
<td>HIROMB</td>
<td>Thickness derived from total mass of ice per unit area</td>
</tr>
<tr>
<td>Ice ridge density</td>
<td>[km⁻¹]</td>
<td>HIROMB</td>
<td>Number of ridges per unit distance</td>
</tr>
<tr>
<td>Ridge height</td>
<td>[m]</td>
<td>HIROMB</td>
<td>Average height of pressure ridge in the area (sail and keel)</td>
</tr>
<tr>
<td>Snow depth</td>
<td>[m]</td>
<td>HIROMB</td>
<td>Thickness of the snow layer</td>
</tr>
<tr>
<td>Atmospheric conditions</td>
<td></td>
<td>AA</td>
<td>(Daylight, Twilight, Night)</td>
</tr>
<tr>
<td>Light conditions</td>
<td>[–]</td>
<td>AA</td>
<td></td>
</tr>
<tr>
<td>Visibility</td>
<td>[nm]</td>
<td>PMP</td>
<td></td>
</tr>
<tr>
<td>Cloud cover</td>
<td>[%]</td>
<td>MESAN</td>
<td></td>
</tr>
<tr>
<td>Wind speed</td>
<td>[m/s]</td>
<td>MESAN</td>
<td></td>
</tr>
<tr>
<td>Wind gust</td>
<td>[m/s]</td>
<td>MESAN</td>
<td></td>
</tr>
<tr>
<td>Wind direction</td>
<td>[–]</td>
<td>MESAN</td>
<td>(N, NNE, NE, ENE, E, ESE, SE, SSE, S, SSW, SW, WSW, W, WNW, NW, NNW)</td>
</tr>
<tr>
<td>Precipitation type</td>
<td>[–]</td>
<td>PMP</td>
<td>(No precipitation, Drizzle, Rain, Snow and rain, Snow)</td>
</tr>
<tr>
<td>Precipitation amount</td>
<td>[g/m² h]</td>
<td>PMP</td>
<td></td>
</tr>
<tr>
<td>Temperature</td>
<td>[°C]</td>
<td>PMP</td>
<td></td>
</tr>
<tr>
<td>Impact conditions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impact speed vessel 1</td>
<td>[kn]</td>
<td>AIS-V</td>
<td>In case of ship-ship collision: struck vessel</td>
</tr>
<tr>
<td>Impact speed vessel 2</td>
<td>[kn]</td>
<td>AIS-V</td>
<td>Only in case of ship-ship collision: striking vessel</td>
</tr>
<tr>
<td>Impact angle</td>
<td>[°]</td>
<td>AIS-V</td>
<td>Angle between centre lines, measured from struck vessel’s stern</td>
</tr>
<tr>
<td>Impact location</td>
<td>[–]</td>
<td>AIS-V</td>
<td>Location on the struck vessel hull: {Bow, Side, Stern}</td>
</tr>
</tbody>
</table>


* Throughout this paper, the category ‘grounding/stranding/bottom contact’ is simply referred to as grounding, for reasons of brevity.
Fig. 4. Winter navigation accidents in relation to sea ice conditions.

Fig. 5. Winter navigation accidents in relation to atmospheric conditions: precipitation, visibility and temperature.
as traffic operations continue around the clock and because in the winter months, days are short at the latitudes where the Northern Baltic Sea is located. Comparatively many accidents occur during twilight conditions (8 cases) as well.

There is a large variation in the cloud cover under which the accidents occur, for all accident types. Nevertheless, it is apparent that the sky is typically almost completely covered with clouds. There is however a notably different pattern in the prevailing temperatures. For ship-ship collisions, temperatures range from ca. −20 °C to 5 °C, and most typically are around −5 °C. However, groundings occur at temperatures ranging from ca. −10 °C to 10 °C, with most typical values of ca. 5 °C. This is in line with the finding from Section 4.1 that groundings typically occur in ice-free conditions, whereas ship-ship collisions occur in sea ice conditions.

During all accident types, there typically is no precipitation (35 cases), with snow (7 cases) being the most frequently occurring precipitation type. Other types (drizzle, rain and snow and rain) occur only in one accident case each. If there is precipitation, the amount of snow typically is between 50 and 200 g/m² h. The one outlier case where a ship-ship collision occurred under heavy rain (670 g/m² h) was in a port area in night time under visibility conditions classified as ‘good’.

Fig. 6 depicts a PCP showing the patterns in the relations between operation types, sea area, and various atmospheric conditions relating to the wind conditions. The plot shows that there is no strong pattern for the relations between wind speed and wind gust and the different accident types. For groundings and ship-ship collisions, wind speeds range from 2 m/s (light breeze) to 13 m/s (strong breeze), with a median of ca. 5 m/s, i.e. a gentle breeze. The only significant difference is found for collisions with an object. These occur in port areas, which probably due to more sheltered conditions lead to lower wind speeds, with a median of ca. 4 m/s. The most typical prevailing wind directions during all accident types are north-western, northern, north-eastern and eastern. There is a quite stable relationship between wind speed and wind gust, in that higher wind speeds also lead to higher wind gusts.

One of the useful characteristics of visual data mining is the ability to ‘brush’ data, i.e. to interact with the dataset through intuitively selecting certain subsets of the data, which may reveal interesting features in the data. An example of an interesting finding resulting from brushing the data in the PCP focusing on winter accidents in relation to atmospheric conditions is given in Fig. 7. The same data features as in Fig. 5 are shown. In the upper PCP, only accident cases in cut loose operations are shown, whereas in the lower PCP only accident cases in escort and convoy operations are selected. Using this technique, it is seen that the patterns between these operation types are practically identical for cloud cover, precipitation amount and types, temperature and visibility conditions. However, whereas escort and convoy collisions primarily occur during night and twilight conditions, collisions in cut loose operations appear to occur more frequently in daylight.

4.3. Ship-ship collisions: impact conditions and vessel characteristics

Figs. 8 and 9 depicts a PCP showing the patterns in the relations between impact conditions, sea area and operation types, for ship-ship collision accident cases. Escort and convoy operations show largely similar patterns. The stern is typically the location where the struck vessel is impacted, and impact angles are close to 0°, indicating that the vessels are heading in the same direction. The impacted vessel typically has stopped or is proceeding at low speed, up to ca. 4 kn. The impacting vessel has a somewhat higher speed, up to ca. 7 kn. These characteristics are not surprising given the normal conditions under which these operations occur, as evident from the description in Table 2. Comparing the impact speeds

---

4 For an interpretation of the wind speeds: see MetOffice (2010).
with the normal operational speeds in escort and convoy operations, which are 10 kn and 9 kn based on an analysis by Goerlandt et al. (2016), it is found that these are significantly lower.

Ship collisions in cut loose operations show a different pattern. The impact location is most commonly the struck ship's bow or stern, but impacts on the ship side occur more frequently as well, whereas impact angles are typically oblique (stern-bow or bow-bow). The struck vessel always is stationary, whereas the striking vessel has a speed between ca. 1 kn and 7 kn.

The PCP of Fig. 9 also shows two outlier cases concerning a collision occurring in independent navigation and a collision between drifting vessels. The former is a collision in a port area, with the two vessels impacting at the struck vessel's side under an oblique angle at low impact speeds. The latter occurred in the open sea, with the vessels impacting each other's sides at very low speed with close to parallel headings.

Figs. 10 and 11 show PCPs clarifying the patterns between the sea areas, operation types and the characteristics of the vessels involved in ship-ship collision accidents. For escort operations, the struck vessel always is the assisting icebreaker, with striking vessel types tankers, bulk carriers or general cargo vessels. The striking vessels are typically rather small, with a deadweight of ca. 5000 tonnes and a length around 100 m. Accidents in escort operations typically occur in archipelagic waterways.
Convoy operations show a pattern which is in some sense quite similar to escorts. However, whereas escort accidents mainly occur in archipelagic waters, convoy collisions are more common in open sea areas. The struck vessel type here is typically a cargo vessel of ca. 5000 tonnes with a length of around 100 m. Striking vessels are larger than in escort operations, with a typical deadweight of 10,000 tonnes and a length of 140 m. Variations nonetheless are relatively large. In one accident was a tanker of 23,000 tonnes deadweight and 180 m struck by a large bulk carrier.

Cut loose operations show a different pattern. Here, the striking vessel is always the icebreaker assisting the beset vessel, and accidents occur both in open sea and archipelagic waterways. The icebreakers operating in the Northern Baltic Sea have a deadweight between 1400 and 6500 tonnes and a length around 110 m. Struck

Fig. 8. Winter navigation ship-ship collision accidents in relation to impact conditions: convoy and escort operations.

Fig. 9. Winter navigation ship-ship collision accidents in relation to impact conditions: cut loose, drifting and independent navigation.
vessels in these operations are most commonly cargo vessels with a deadweight between 4000 and 8000 tonnes and a length between 90 and 120 m. Two accidents involved a work vessel (pusher tugs) with a deadweight of ca. 450 tonnes and a length of ca. 40 m. One accident involved a Panamax tanker with deadweight ca. 64,000 tonnes and a length of 238 m.

Ship-ship collisions in independent navigation occur rarely but the records show only accidents in port areas. In one case, the stuck vessel...
Fig. 12. Winter navigation collisions with an object in relation to vessel characteristics and impact conditions.

Fig. 13. Winter navigation groundings in relation to vessel characteristics and impact conditions.
vessel was a cargo vessel or ca. 6000 tonnes deadweight and a
length of ca. 110 m, with a cargo vessel of 2500 tonnes deadweight
and a length of 80 m the striking vessel. In another, two harbour
tugs with a deadweight of ca. 50 tonnes and a length of ca. 20 m
collided. The ship-ship collision in a drifting situation occurred in
the open sea area between two RoPax vessels of 3500 deadweight
deadweight and 160 m length and 7800 deadweight and 120 m length,
respectively.

4.4. Collisions with object: impact conditions and vessel characteristics

Fig. 12 shows a PCP depicting data attributes of collisions with
an object, focusing on the sea area where these occurred, the
impact conditions and vessel types and sizes. It is seen that the
vessels involved typically are cargo vessels, but also a tanker and
a RoPax have been implicated. The impact speeds range from
1 kn to ca. 6.5 kn, and the ship sizes range from ca. 3000 to
6000 tonnes deadweight. Nevertheless, due to the relative infre-
frequent occurrence of this accident type (only 7 cases are found),
no strong patterns can be identified.

4.5. Grounding/stranding/bottom contact: impact conditions and vessel
characteristics

Fig. 13 shows a PCP depicting data attributes of grounding acci-
dents, focusing on the sea area where these occurred, the impact
conditions and vessel types and sizes. It is seen that groundings
in archipelagic waterways typically involve cargo vessels of a dead-
weight between 800 and 3000 tonnes and a length of around 80 m.
Nevertheless, also larger vessels of over 20,000 tonnes and ca.
170 m length have been involved. Impact speeds in archipelagic
waters range between ca. 4 and 12 kn.

A somewhat different pattern is found for groundings in port
areas. Here, the most frequently implicated ship types are cargo
vessels of a deadweight between 3000 and 5000 tonnes and a
length around 110 m. While impact speeds can be as high as
12 kn, these are mostly lower, often even below 2 kn, i.e. signif-
cantly lower than in groundings in archipelagic waters. One work
vessel (a dredger), a RoPax and two tankers have grounded in port
areas. Their sizes range between 300 and 8000 tonnes deadweight
deadweight and ca. 45–115 m length.

Comparing the speeds with the impact speeds in ship-ship col-
lissions as shown in Figs. 8 and 9, it is evident that these are signif-
cicantly higher in grounding accidents. This likely relates to the
finding from Fig. 4 that these operations typically occur in ice-
free conditions.

4.6. Summary statistics

While the focus of the analysis is the patterns occurring in win-
tertime accidents conditions, it is instrumental to also provide
some insights into the relative occurrence frequency of different
accident types. In Figs. 14–16, some aggregate analyses concerning
the number of accidents in different conditions, are shown. These
quantitative results provide some insights beyond the qualitative
findings from Section 4.5. However, it is stressed that for a proper
interpretation of these quantities, the multidimensionality of the
data must be considered. In particular, the identified patterns,
summarized in Table 6 of Section 5, should be considered alongside
the quantitative relative occurrence frequencies. It is also stressed
that the number of accidents only concerns those cases which were
identified both in the accident database and in the AIS data. Hence,
the numbers should be considered relative to one another, not as
absolutes.
Fig. 14 shows the number of accidents occurring, by accident type and by ship type. In the case of ship-ship collisions, the ship type is that of the struck vessel. It is seen that ship-ship collisions occurred 21 times, groundings 17 times and collisions with an object 7 times in the considered time period. Cargo vessels were implicated most frequently in all accident types, whereas icebreakers also often were the struck vessel. Other ship types (work vessels, tankers, bulk carriers and RoPax vessels), were occasionally involved in an accident.

From Fig. 15, it is seen that 17 accidents occurred in port areas, and the same number in archipelagic waterways. 11 accidents occurred in open sea. From Fig. 5, it is seen that most accidents in the open sea concern ship-ship collisions, whereas groundings occur in port areas and archipelagic waterways.

From Fig. 16, it is evident that most accidents (26 cases) occurred during independent navigation. Comparing this to Fig. 4, it is seen that most of these accidents are groundings. Cut loose operations led to 8 accident cases, whereas convoy and escort operations resulted in respectively 4 and 6 accidents. One accident occurred with the vessels drifting.

5. Discussion

The analysis using visual data mining techniques on integrated data sources has provided novel insights into the conditions under which wintertime navigational accidents in the Northern Baltic Sea occur. Jalonen et al. (2005) and Valdez Banda et al. (2015, 2016) focus on accident frequencies and vessel types involved in winter accidents in Finnish waters, relying both on accident data and expert judgments. Valdez Banda et al. (2015, 2016) also analyses the sea ice conditions to some extent, but do not relate this to the sea areas or the operation types under which these occur. The prevailing atmospheric conditions have not earlier been related to winter navigation accidents. Finally, whereas impact scenarios have been studied for ship-ship collisions in ice-free conditions, see e.g. Samuelides et al. (2008), Goerlandt et al. (2012) and Youssef et al. (2013), no similar work for accidents in ice conditions is known.

In scientific work, it is important to consider uncertainties in data and methods in relation to the conclusions of an inquiry (Douglas, 2009). As mentioned in the introduction, one significant use of the presented analysis is maritime transportation risk analysis. Also in risk analysis, the importance of assessing the uncertainty in the evidence base has been highlighted, see e.g. Flage et al. (2014), Gardoni and Murphy (2014) and Goerlandt and Reniers (2016).

For the current analysis, this means that it is important to make an informed judgment about the strength of evidence for the different identified patterns in the conditions under which the winter navigational accidents occurred. It is noted here that strength of evidence is an alternative way to communicate uncertainty in the background knowledge, see e.g. Flage et al. (2014).

The strength of evidence is qualitatively rated using a method suggested by Goerlandt and Reniers (2016). For data and models, different quality characteristics are considered and judged, from which an evidential strength is derived. These are shown in Table 4, where for conditions between strong and weak evidential characteristics, a ‘medium’ rating is applied.

These qualities are assessed in Table 5 for the data attributes used in the visual data mining, based on the data sources described in Section 2. A justification is given for the rating, to substantiate why a certain rating is given for a given data attribute. Contrary to the strength-of-evidence assessment method by Goerlandt and Reniers (2016), who focus on the evidence base for subjective probability assignments, the amount of data is here taken as a characteristic of the identified patterns through the use of PCPs, and is thus assessed separately.

In Table 6, the patterns identified through visual data mining using PCPs in Section 4 are summarized. Additionally, a rating is associated with the number of cases supporting the identified pattern, and the strength of evidence (SoE) of the attributes involved in the pattern, making use of the ratings of Table 5. Based on these ratings, an overall strength of evidence for the pattern is derived. This is done following a simple rule shown in Table 7. The rating for the strength of evidence of the pattern takes the rating of the strength of the pattern (related to the number of data points the identified pattern is based on) as a basis. This rating is moved down depending on the strength of evidence of the relevant attributes. Similar rating schemes are used in Gardoni and Murphy (2014) and Goerlandt and Reniers (2016).

The information in Table 6 is particularly useful for risk analysis purposes: the identified patterns can be used to define accident scenarios in a risk model, whereas the uncertainties about these scenarios can be appropriately accounted for through the strength of evidence assessment. It is noted that Table 6 only lists the main patterns found in the data. Outlier cases mentioned in Section 4 should be considered to have low strength of evidence.

In addition, it is important to note that the analysis performed in this paper has focused on identifying patterns between variables relevant to contextualize wintertime accidents, as obtained from accident databases. As stated in Section 1, the research questions

<table>
<thead>
<tr>
<th>Evidence characteristics</th>
<th>Strong evidential characteristics</th>
<th>Weak evidential characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>Low number of errors</td>
<td>High number of errors</td>
</tr>
<tr>
<td>Quality</td>
<td>High accuracy of recording</td>
<td>Low accuracy of recording</td>
</tr>
<tr>
<td></td>
<td>High reliability of data source</td>
<td>Low reliability of data source</td>
</tr>
<tr>
<td>Amount</td>
<td>Much relevant data available</td>
<td>Little data available</td>
</tr>
<tr>
<td>Models</td>
<td>Many different experimental tests performed</td>
<td>No or little experimental confirmation available</td>
</tr>
<tr>
<td>Empirical validation</td>
<td>Existing experimental tests agree well with model output</td>
<td>Existing experimental tests show large discrepancy with model output</td>
</tr>
<tr>
<td>Theoretical viability</td>
<td>Model expected to lead to good predictions</td>
<td>Model expected to lead to poor predictions</td>
</tr>
</tbody>
</table>
Table 5
Strength of evidence of data attributes.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Source(s)</th>
<th>Section</th>
<th>Data Model</th>
<th>Justification</th>
<th>SoE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accident type</td>
<td>NDB 2.1, 3.4</td>
<td>M-H</td>
<td>– –</td>
<td>Cross-checks between different accident databases underlying NDB and AIS-V. Accurate sea chart used to distinguish collision with object and groundings.</td>
<td>H</td>
</tr>
<tr>
<td></td>
<td>AIS-V 2.2, 3.3</td>
<td>H</td>
<td>– –</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SC 2.4</td>
<td>H</td>
<td>– –</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sea area</td>
<td>NDB 2.1, 3.4</td>
<td>M</td>
<td>– –</td>
<td>Accident locations, if erroneous in NDB, are corrected using AIS videos. Sea chart is used to determine the sea area.</td>
<td>H</td>
</tr>
<tr>
<td></td>
<td>AIS-V 2.2, 3.3</td>
<td>H</td>
<td>– –</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SC 2.4</td>
<td>H</td>
<td>– –</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operations type</td>
<td>AIS-V 2.2, 3.3</td>
<td>H</td>
<td>– –</td>
<td>AIS-V provide good insight in the operational features of Table 2.</td>
<td></td>
</tr>
<tr>
<td>Ice type</td>
<td>IC 2.3</td>
<td>L-M</td>
<td>– –</td>
<td>IC has good quality for operational purposes with good spatial resolution (1 nm), but temporal resolution (24 h) is low.</td>
<td>M</td>
</tr>
<tr>
<td>Ice concentration and thickness</td>
<td>HIROMB 2.3</td>
<td>M-H</td>
<td>M-H</td>
<td>Ice concentration and thickness are assimilated from Ice Charts. Validation shows ice concentration is underestimated in the ice charts.</td>
<td>M-H</td>
</tr>
<tr>
<td>Snow</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impact parameters</td>
<td>AIS-V 2.2, 3.4</td>
<td>M-H</td>
<td>– –</td>
<td>AIS-V provide good insight in ship dynamics. AIS data has good accuracy, but data gaps and interpolation may lower accuracy of data as shown in video.</td>
<td>M-H</td>
</tr>
<tr>
<td>Vessel data</td>
<td>NDB 2.1, 2.4</td>
<td>L-M</td>
<td>– –</td>
<td>Data in NDB is often erroneous, but cross-checks with AIS-V and VDB lead to accurate data.</td>
<td>H</td>
</tr>
<tr>
<td></td>
<td>AIS-V 2.2, 3.4</td>
<td>H</td>
<td>– –</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>VDB 2.1, 2.4</td>
<td>H</td>
<td>– –</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

focus on contextualizing wintertime accidents, to get insight into under which conditions these occur. Two caveats are worth stressing in this context.

First, while several patterns are quite clearly discernible, the existence of a pattern should not be confused with probabilistic statements that under the conditions given by the identified pattern, accidents are more likely to occur. This is because our data does not allow insights into the prevalence of these contextual conditions compared to the total set of conditions. For instance, while it is apparent in Fig. 4 that collisions with another vessel typically occur in ridged ice with a height of ca. 1.5 m, it is not possible to unambiguously state that accidents are more likely under these conditions than in other conditions. This is because the dependence between which operations occur in which ice conditions,
and the relative prevalence of ridge heights in ridged ice fields, is unknown. Investigating this would require a deeper understanding of the traffic patterns relative to the ice conditions, and the relative prevalence of different ice conditions. Analysis of such statistical relative probability, rather than the assertion of the existence of certain patterns, is beyond the scope of this paper. Further research is required to make more conclusive statements, and for instance methods for identifying the operation types for all wintertime traffic (i.e. not only accident cases) would need to be developed.

A second caveat is related to the above: correlation does not imply causation. Establishing that accidents typically occur in certain conditions, e.g. in ridged ice of a given thickness, does not mean that those conditions can be considered to have a causal connection to the accident occurrence. Accident causation is a complex issue, see e.g. Qureshi (2007), and the data does not provide any proof that the contextual conditions did have a causal effect to the occurrence of the accidents. The identified patterns should therefore primarily be understood as describing the historic accident cases. From this, hypotheses may be formulated as to why accidents occur under the given conditions, but the patterns themselves contain no causal explanatory power.

6. Conclusions

In this paper, wintertime navigational accidents in the Northern Baltic Sea in the period 2007–2013 are analysed, focusing on the sea ice, atmospheric, operational and impact conditions under which these occur. Various data sources were integrated to obtain an as complete and accurate as possible picture of the context of the accidents. Videos of AIS data were made to obtain insights in the winter navigation operation type and the impact conditions.

Visual data mining, enabling the identification of patterns in multidimensional datasets through interactive plotting tools, was applied to analyse the integrated dataset. Several patterns were identified, related to the sea areas and operation types in which different accident types occur. New insights were also obtained in the prevailing sea ice and atmospheric conditions in the different accident and operation types, and patterns in impact conditions were identified. Special attention is given to the strength of evidence, both concerning the strength of the identified patterns and the strength of evidence for the relevant data attributes. This is especially relevant as one envisaged use of the current analysis is for wintertime maritime transportation risk analysis, where the consideration of uncertainties in the background knowledge has gained recent attention.

Acknowledgements

The research presented in this paper has been conducted in the context of the “Strategic and Operational Risk Management for Wintertime Maritime Transportation System” (BONUS STORMWINDS) project. This has received funding from BONUS, the joint Baltic Sea research and development programme (Art 185), funded jointly from the European Union’s Seventh Programme for research, technological development and demonstration, and from Baltic Sea national funding institutions: the Academy of Finland and the Research Council for Environment, Agricultural Sciences and Spatial Planning (FORMAS). The financial support is acknowledged.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ssci.2016.09.011.


