Safety Science 49 (2011) 192–197

Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/ssci

An application of data mining tools for the study of shipping safety in restricted waters

Dimitrios X. Kokotos [a,*], Dionisios S. Linardatos [b]

[a] Department of Maritime Studies, University of Piraeus, 40 Karaoli & Dimitriou, Piraeus 185 32, Greece

[b] Department of Informatics and Telecommunications, University of Athens, Panepistimiopolis-TYPA Buldings, Ilissia 15784, Athens, Greece

ARTICLE INFO

Article history:
Received 18 February 2010
Received in revised form 26 July 2010
Accepted 30 July 2010

Keywords:
Shipping safety
Data mining
ISM-Code
Human error

ABSTRACT

The effectiveness of the enforcement of the ISM-Code and the examination of its role in the distribution of causes of shipping accidents between human and non-human error was studied. All accidents involving Greek-flagged ships from 1995 to 2006, a time-scale which spans over the pre- and post-ISM period in navigational regions of restricted waters, were analyzed.

The accident data was processed through a classification tree analysis which enabled the classification of various accident factors. The analysis revealed that although the human error maintained its position as the dominant factor in shipping accidents, there is also substantial evidence in support of the ISM-Code effective control over shipping accidents during the post-ISM period. The implementation of the ISM-Code led to an overall reduction of human-induced accidents in total. Furthermore, in terms of location, the ISM-Code improved the human-induced accident record within restricted waters.

Conclusively, the ISM-Code constitutes an effective policy measure for shipping safety. The results of the classification tree analysis reported in the present work can be used by decision makers in companies and international organizations to build knowledge-based expert systems and augment their information in the field of safety policy and management.

1. Introduction

The International Safety Management Code (ISM-Code) was enforced by the International Maritime Organisation (IMO) in 1998 in order to provide an international standard for the safe management and operation of ships and for the prevention of pollution.

Accidents in shipping and marine environmental pollution are attributed to various factors. The UK Marine Accident Investigation Branch (MAIB) states that "*one factor still dominates the majority of maritime accidents: human error*" (MAIB, 2000). In recognition of this statement, the ISM-Code aims at the promotion of safety in shipping by ensuring all onboard activities must adhere to an operating standard of "who does what and when", according to the Safety Management System (SMS) which had to be maintained by all shipping companies after the implementation of the ISM-Code. To this extent, an analysis of shipping accidents through the application of classification trees structured on "human" or "non-human" split as the primary cause is a valuable tool in accessing the effectiveness of the ISM-Code.

A lot of effort from researchers world-wide has been devoted in estimating dependency of shipping accidents from various factors (Celik et al., 2010; Grech et al., 2008; Tzannatos, 2005; Tzannatos, 2002) as well as in assessing the effectiveness of the measure of the enforcement of the ISM-Code.

Previous research in this area (Psaraftis et al., 1998) provides an 'in-depth' analysis of the human element as a factor in Greek shipping accidents during the 1984–1994 period and in conclusion it comments that "*It is impossible to tell for each of the cases reviewed that the accident would not have occurred if ISM were in place for the ship in question. It is also early to assess the impact of ISM on the safety of the ships on which the Code has been implemented. This will take years to ascertain, and the analysis to do so will not be trivial. However, the very fact that ISM certification implies that all procedures related to the operation of the ship would at least be established, monitored and controlled, means that the risk of a situation getting out of hand would be minimised.*"

Though comparable data on the effectiveness of the ISM-Code are not as straightforward as they should be, firstly due to the presence of inconsistencies in reporting and administration between different Memoranda of Understanding (MoUs), as well as amongst the signatory nations of each MoU on the other hand due to the influence of subjective issues, such as the attitude of the crew, the ease of inspection, the inspector's mood results have shown that the ISM system is working (Paris MoU CIC, 2007).

However, the ISM-Code assessment should not only focus on providing answers if the ISM system has overall improved safety at sea, but also on providing answers to several detailed questions.

* Corresponding author. Tel./fax: +30 210 414 2526.
E-mail address: dkokotos@unipi.gr (D.X. Kokotos).

0925-7535/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ssci.2010.07.015

Thus, the ISM-Code assessment procedure should reveal knowledge of 'why people make mistakes', or 'in which cases of accidents ISM system has improved safety the most' or 'in which cases (if any) ISM has not improved safety' and so on. The extraction of this type of information may lead to proposals of appropriate remedial action.

Tools used for the ISM-Code assessment include mainly classical statistical analysis techniques (Giziakis et al., 1996) applied in the records of vessel detentions and deficiency notices produced within the framework of inspections performed by the relevant authorities. These techniques produce a thorough shipping accidents analysis giving the statistics of human influence upon shipping accidents.

The study of shipping accidents with data mining techniques (Kokotos and Smirlis, 2005) may reveal information not already extracted by researchers who had used the classical statistical analysis techniques.

A powerful data mining function is the classification tree, applied in order to produce optimal classification rules. This algorithm is trained to classify the cases of a dataset, as the dataset of shipping accidents used in the current work, to certain categories of a target (dependent) variable using the information of several predictive (independent or explanatory) variables. Hence, the resulting tree diagrams relate the categories of the target variable to its predictors (Goodman, 1979).

Classification trees are easily understood by both experts and non-experts. The classification trees may be utilized in the context of a potential decision support system and a risk management information system that will record evaluate and process data for ship accidents.

The current work aims at mining hidden information and estimating the dependency of the source of accidents (human or non-human) upon the year of accident, the location of accident, the vessel type, size and age. The mined information will be used to assess the role of the ISM-Code in the distribution of causes of shipping accidents between human and non-human error. Finally, the mined information will measure the effectiveness of the enforcement of the ISM-Code. The data mining was performed in accidents of Greek-flagged ships world-wide that cover both the pre-ISM-Code and post-ISM-Code periods. With respect to the choice of flag administration, it is assumed that Greek shipping, by virtue of its size and diversity, constitutes a valuable reference for the analysis of accidents.

2. Methodology

2.1. The accidents dataset and the variables

The current work utilises the information included in official investigation reports of the Hellenic Coast Guard (HCG) referring to accidents of Greek-flagged ships world-wide over the period of years 1995–2006. The time span under consideration was determined by the scope of the work and the availability of data. In this respect, widening the analysis to accidents prior to 1995 inevitably strengthens the influence of non-human (such as technical) shipping safety related measures against the ability to capture the influence of the ISM-Code as implemented in 1998 for the control of the human element in shipping accidents. More specifically, in the end of 1994, the enforcement of SOLAS[1] amendments relating to the stability and fire protection of passenger vessels and amendments to the IBC[2] and the IGC[3] Codes are considered very important technical measures in shipping safety. Accidents happened after 2006 have not been included in our dataset as some official reports regarding the causes of these accidents had not been finalised at the time of our data processing. Prerequisites for the inclusion of an accident in HCG's official investigation reports is that at least one of the following situations were encountered in the accident:

- Total or partial loss of a ship.
- Ship is taken-over by insurers.
- Permanent or temporary abandonment of ship by the crew.
- Cargo loss or failure (more than 25%).
- Prolonged loss of ship command due to serious failure.
- Loss of life or serious injury to a crew member or passenger.

Because the primary source of information was on textual form, a thorough process of data entry, editing and validation was carried out. This information was transformed and coded in order to produce the dataset.

Furthermore, since ships under 500 grt have been exempted of ISM-Code compliance, all accidents involving vessels under 500 grt were discarded from the dataset. On the basis of these criteria and after the application of data cleaning (Kokotos, 2003) for eliminating cases with identical accident information, with missing values and unreliable data, a dataset consisted of 268 shipping accidents (the accidents dataset) appeared.

For every accident, the accidents dataset includes the information depicted in Table 1.

Table 1
List of dependent and independent variables.

Variable label	Measure
Source (or cause) of accident	Dependent
Year of accident	Predictor
Location (of accident)	Predictor
Vessel tonnage (grt)	Predictor
Vessel type	Predictor
Vessel age	Predictor

Analytically, we have:

1. *Source (or cause) of accident.* It is a dichotomous variable accepting values of "human" or "non-human" with 57.1% and 42.9% of observations, respectively. These values represent the conclusion of the reports based upon the formal investigation of the shipping accidents, with regard to whether a human entity (Ship's Master or 1st Engineer, Pilot, Offshore Personnel etc.) was ultimately responsible for the accident or otherwise (Random Event, Act of God, Unidentified Source). In our analysis it is used as the target variable.
2. *Year of accident.* It is a dichotomous variable accepting values of "pre-ISM" ("until 1998") or "post-ISM" ("after 1999"), with 44.4% and 55.6% of accidents, respectively. In our analysis it is the main explanatory variable intended to measure the pattern of effects before and after the ISM-Code enforcement.
3. *Location (of accident).* It is a dichotomous variable accepting values of "restricted waters" or "open sea". It is related to the position of the vessel at the time of the accident, namely restricted waters (ports, canals, straits, anchorages, coastal waters etc.) and open sea, with 58.6% and 41.4% of observations, respectively.
4. *Vessel tonnage.* The vessel tonnage ranges between 488 and 132,590 grt, with standard deviation of 17,296 and mean value of 13,040 grt.
5. *Vessel type.* It is a categorical variable accepting the values of "general cargo" or "bulker" or "container" or "tanker" or "cruise

[1] International Convention for the Safety of Life at Sea (SOLAS), 1974.
[2] International Code for the Construction and Equipment of Ships Carrying Dangerous Chemicals in Bulk.
[3] International Code for the Construction and Equipment of Ships Carrying Liquefied Gases in Bulk.