TransNav International Journal on Marine Navigation and Safety of Sea Transportation
Volume 3
Number 3
September 2009

A Study of Marine Incidents Databases in the Baltic Sea Region

A. Mullai, E. Larsson & A. Norrman
Lund University, Lund, Sweden

ABSTRACT: A comprehensive risk analysis makes use of different datasets. Marine incidents data are essentially important datasets. The purpose of this study is to analyse marine incident databases in the BSR (Baltic Sea Region). The marine incident data in the region are inhabited by a wide rage of issues, such as limited data accessibility and availability, and the diversity in data quality, structure, amount, accuracy, degree of detail and languages. Preparing for the data analysis is a very cumbersome, labour intensive, time consuming and expensive process. Merging different datasets from different countries into a single dataset is a very difficult process, if not impossible for a complete data merging. The paper provides experiences on how to overcome some of these issues and proposes some suggestions for improvements in the future.

1 INTRODUCTION

1.1 *Background*

The Baltic Sea is the world's largest brackish body of water. It is designated as a PSSA (Particularly Sensitive Sea Area). More than 2000 large ships including large oil tankers are at any given time in the Baltic Sea (HELCOM 2005, Rytkönen et al. 2002). Maritime transport adversely affects different risk receptors in various forms and degrees of extents. Increasingly large amounts of different types of dangerous goods, including oil and oil products, gases and a wide range of chemicals, transported and handled in the BSR (Baltic Sea Region) (estimated between 300-1000 million tons per year) and accidents involving these goods are concerning issues for the countries in the region (TSE 2006). The most recent major oil spills that have occurred in the region are the cases of the m/v "Fu Shan Hai" (2003) (1200 tons of oil spilt) and the m/v "Baltic Carrier" (2001) (2700 tons of oil spilt). The costs of oil spills reported yearly and the worst-case scenarios in Öresund are respectively estimated $223,500 and between $150-300 million (Mullai & Paulsson 2002).

The DaGoRus project (Safe and Reliable Transport of Dangerous Goods in the Russian-EU Logistics Chain) is an European Union (EU)/Tacis project dealing with safe and reliable transport chains of dangerous goods. The project consists of a number of partners (including Lund University - LU, Sweden) and Working Packages (WP). It can be considered as continuation of the DaGoB project (Safe and Reliable Transport Chains of Dangerous Goods in the Baltic Sea Region) (INTERREG IIIB). The main objective of the project is to provide a risk analysis of dangerous goods transport in the BSR. The project is in many respects unique.

1.2 *Literature review*

An extensive literature review showed that a holistic view of the maritime risks in the BSR is limited, and they deserve a better understanding. Projects co-financed by the EU, including the BSR INTERREG Neighbourhood programme, have covered a wide range of issues concerning sustainable development in the region. Baltic Master (2005-2007) and OILECO (Integrating ecological values in the decision making process on oil) (2005-2007) are examples of the recent EU projects. None of these projects has particularly dealt with the risks of maritime transport of dangerous goods at a wider BSR context, including the Russian part. In addition, a few peer-reviewed papers have been confined to a limited number of risk issues, such as the m/v "Estonia" case (Soomer et al. 2001), marine pollution in coastal waters, oil spills detection and remote surveillance (Looström 1983) and monitoring by in-service aircraft (Von Viebahn & Gade 2000) and satellite (Kostianoy et al. 2005).

1.3 *Research questions and objective*

Marine incidents data are essentially important for the risk analysis. Time and financial resources are often limited for research projects, including the DaGoRus project. While the signing of the partnership agreement for the project was still pending, the relevant research questions for this particular study presented in this paper are: What is the current state-of-the-art marine incidents data in the region? Is it feasible to perform a comprehensive risks analysis for the entire region? The purpose of this study is to analyse marine incident databases in the BSR and propose suggestions for improvements in data accessibility, structure and quality.

1.4 *Materials, methods and paper outline*

After several months of communication with the responsible authorities of the BSR's countries, the following marine incident databases were acquired: 1) Danish Maritime Administration Database (DMA DB) (1997-2006; in Danish); 2) Finish Maritime Administration Database (FMA DB1 and DB2) – the database contains two datasets (1990-1996 and 1997-2007; in Finnish); 3) Swedish Maritime Administration Database (SMA DB) (1985-2007; in Swedish); and 4) Helsinki Commission Database (HELCOM DB) (1989-2006; in English). In section 3 of the paper, the main results and discussions including problems encountered during data collection are presented. The properties of the databases are described and compared. For the purpose of benchmarking with some of the best technology and practices in the field, in section 4, the USA's and world's largest incidents databases are described. Conclusions and suggestions for improvements are provided in section 5. Initially, in section 2, the concepts of risks and risk analysis are briefly described.

2 RISKS AND RISK ANALYSIS

2.1 *Maritime risks*

The risk is defined as the likelihood of consequences of undesirable events (Kaplan & Garrick 1981, HSE 1991). The terms "marine accident and incident" and "marine casualty" denote undesirable events in connection with ship operations (IMO 1996). The term "marine incident" is used to denote undesirable marine events, i.e. marine accidents, incidents and near missing events. The dangerous goods risks can be defined as the likelihood of consequences of hazardous release events (HSE 1991). Maritime transport risks are statistically verifiable technological and human activity risks. The maritime transport system and risks consists of many elements that are classified and defined by various coding systems.

2.2 *Risk analysis*

Contemporary risk management recognises the fact that the risk analysis, which is a rigorous scientific process facilitated by standardised frameworks and techniques, is prerequisite for the decision making process. The main purpose of every risk study is, to the best abilities of researchers and data and resources available, to provide decision makers with valid and reliable information and tools that would enable them to make informed decisions. The risk analysis varies from simple screening to major analysis that requires many years of efforts, substantial resources and a large team of experts using various risk analysis techniques and datasets. The main stages of the risk analysis are preparations for analysis, the analysis process and conclusions and recommendations. The first stage encompasses a wide range of activities, including identification, selection, compilation and preparation of the relevant datasets. Large amounts of diverse risk-related datasets are required, but the most important datasets are marine incident data.

3 MAIN RESULTS AND DISCUSSIONS

3.1 *Limited data accessibility and availability*

The Baltic Sea is an unique area in terms of sensitivity and diversity of countries surrounding the area. It is surrounded by nine different countries with different backgrounds, languages and practices, which may hamper data collection and merging and performance of a robust risk analysis.

Data accessibility may be an issue for the region. Our investigation suggests that marine incidents are recoded into databases in all BSR's countries. The websites of the relevant authorities in several countries of the BSR were reviewed. None of them had marine incident databases and other risk-related data available in electronic format for the public use. Requests for data acquisition were sent to all countries (Denmark, Estonia, Finland, Germany, Latvia, Lithuania, Poland and Sweden), except Russia, and HELCOM. Signing of the partnership agreement with Russian partners was still (2007-2008) pending. Further, data collection for the Russian part was the responsible of another project group. Contact information was obtained from the SMA and other sources. Requests were sent to the maritime administrations, coast guards, bureaus of maritime casualty investigation and maritime safety inspectorates. The mail delivery system confirmed that request messages were successfully delivered, received and displayed on the recipient's computer. We were able to receive four (see Section 1.4) marine incident databases. Two databases were primarily obtained as the result of our personal contacts with the relevant authorities. Two other databases were obtained after considerable communication and assistance from our personal contacts. Some respondents did not reply or were not interested in cooperation. Requests were sent several times to those who did not reply. In some countries, the authorities may be unwilling or uncomfortable in providing data, in particular to external parties. Some interrelated reasons were cited inconvenient database format, limited human and financial resources, and data confidentiality.

Inconvenient database format for preparing and sending data in electronic format was cited by several respondents as one of the main issues. In one country, the incident data recorded up until December 2007 were available only at a relatively old computer. According to the respondent from that country, the data were compressed in a way that was practically impossible converting the data into a modern program format, including Excel format. It was very difficult and time consuming to convert all data manually. One respondent from another country replied that their organisation did not work with the Excel program as database. They were still waiting (2008) for the EMSA (European Maritime Safety Administration) database for the statistical analysis of ship incidents. Another one stated that their database contains personal and other information that are not necessary for the risk analysis. Further, converting their entire dataset into a convenient data format was time consuming and impossible task for them.

All respondents stated that preparing and sending data in electronic format were time consuming and labour intensive processes. Due to workload and other inquiries and in combination with limited human and financial resources, they were unable to provide data at all or in due time. They were too busy to assist us as their daily work was high on their priority list. One respondent wrote that he will not send the entire database. But, if we needed simple extractions they would be able to assist us. In case of a large extraction requiring special adjustments, they had to charge us for that.

Data confidentiality might have been one of the main reasons why some authorities did not reply. One respondent stated that in his country marine incident data are confidential. The data are only available for the accident investigation in his country and in his country language. A risk analysis for the BSR as a whole based on exhaustive data may not be possible should all the countries share a similar policy. Two respondents made reference to annual accident reports in pdf-format published on their organisations' website for the public use. The review of numerous accident reports showed that they were comprehensive and well prepared. However, a number of issues are observed. The data are mainly analysed and presented in form of descriptive or summary statistics, such as frequency tables and charts. Application of advanced inference statistics and specific risk analysis methodology were lacking. Reports are prepared by or for the responsible authorities. The knowledge comes from different corners, from practitioners and scientific community alike. However, because of systematic and rigorous processes employed, it is widely accepted that the knowledge generated by the scientific community has a higher degree of confidence, validity and reliability than the other forms. The scientific literature on the maritime risks for the region is, however, very limited. Studies concerning dangerous goods risks are largely confined to oil spills in the territorial waters of individual countries or in certain areas of the BSR. Further, integrating information from different reports is an impossible task.

3.2 *Diverse and incompatible data*

In this section, the data properties are explored (see Tables 1-6 and Figs 1-3). The HELCOM DB contains marine incidents reported by the BSR's countries. This dataset may serve, to some degree, as a sample for studying and drawing conclusions for the maritime risks in the entire region. However, the dataset is a relatively small and biased sample. The review of other databases showed that incidents are selectively reported to the HELCOM. Thus, during the period 1989-2006, a total number of 906 incidents (50 incidents per year) has been reported, of which 123 (13.6%) and 82 (9.1%) are respectively pollution incidents and incidents with no information about pollution. These numbers are smaller than pollution incidents and marine incidents recorded in the BSR's databases (e.g. SMA DB - 5778 incidents reported during 1985-2007). During the period 2004-2006, the Swedish Coast Guard alone has observed on average 308 spills per year. In addition, the HELCOM DB contains 42 variables, where 17 variables (40%) represent ship properties and consequences (Table 6 and Fig. 1). The consequences are confined to the occurrence of pollution (yes/no), the amount and type of pollutants. The variable labels are not properly designed and partly or completely missing in some variables. For example, the "ship type details" variable contains some 126 items.

The risk estimation and presentation require exhaustive data. The results obtained from the risk estimation may serve as an empirical ground for establishing risk criteria in the region. The risk criteria may serve as benchmarking standard for measuring and comparing the maritime risks in the individual countries and the region as a whole. In Sweden and other countries in the region, these criteria are lacking. Further, the reliability and validity of risk estimation and presentation depends very much on the data quality, diversity and amount. Therefore, it is important to perform a comprehensive risk analysis based on all datasets available. The best alternative is to merge all datasets into a single dataset. A precondition in the data merging process is that all variables must be compatible, i.e. they have to share similar properties including variable type, number, label, and value and measurement level. Variables are not organized in any particular order in the databases. Based on the SMA DB, variables are labelled and organized in main categories (Table 1). A complete data merging, which is merging all databases including all cases and variables, is not possible. Merging parts of datasets may be possible, but considerable time, resources and expertise are required. This process includes translation, codification and de-codification and design and re-design of variables, conversion of data from Excel to SPSS data format, data merging and filtering. One case history is one A4 paper text (multiplied by ca 8000 incidents) and many variables are string or text format variables written in different languages. Text format variables contain very important information. The databases are mainly designed based on the DAMA coding system, which was originally agreed (1990) by Scandinavian countries for registration and analysis of marine incidents. The FMA DB and SMA DB share more in common than the two other databases. Deviations from the code and changes are also observed within the same database (e.g. FMA DB). The labels of many variables in the FMA DB are coded according to the DAMA coding system. These variables must be de-codified. The databases are specially designed databases, which may be inconvenient for converting data into advanced statistical program formats. The data were sent to LU in an Excel format. Data analyses and result presentations with this data format are limited. The present Excel data format of the databases is not readily convertible to the statistical program format. The data are organized on the "case" and "variable" bases. It is unclear whether the case histories are compiled on "event" or "ship" bases. In many cases, in a single incident two or more ships have been involved. The DMA DB is confined to a very limited number of events such as collision, grounding and fire. In one database, many cases were deleted. The SMA DB and FMA DB contain incidents that have occurred in the respective country territorial waters for all nationalities and ships flying the respective country flag outside the territorial waters. Therefore, variables should be designed for filtering or sampling purposes. In terms of data properties and structure, there are considerable discrepancies among the databases (Tables 2-6 and Figs 1-3). Only a very few variables are compatible. There are significant gaps in the number, category, type and measurement level of variables. The string (Str.) and nominal (Nom.) variables dominate (51-75%) all databases. The second largest numbers of variables are scale (Sc.) and ordinal (Ord.) variables (10-38%). More analysis methods are applicable to scale and ordinal variables than nominal variables. The variables for measuring risks of maritime transport of dangerous goods are very limited, if not lacking for certain databases.

Table 1. The main categories and examples of variables in the SMA DB (1985-2007).

Main category: examples of variables
Time: year, month, day, day of the week, time (hours)
Location: latitude and longitude, ports of departure and arrival, country, geographical areas, traffic area, fairway etc.
Ship: call sign, name, type, class society, nationality, built, size (dwt, brt, length), material etc.
Ship activity: ship activity (en-route, loading/discharging), activity onboard etc.
Exposure: crew, passengers, visitors, total numbers etc.
Cargo: description, type, amount, dangerous goods etc.
Event: type, event grading, description
Cause: categories codified, description
Other: pilot presence onboard
Environment: light, visibility, precipitation, sea, wind etc.
Consequence: human (fatality, injury, disappearance – crew, passengers, pilots, others, total), ship (damage – description, location, extent), environment (oil and other pollutants – type, amount)

Table 2. The categories, types and measurement levels of variables in the DMA DB (1997-2007).

Main category	Variable Type*				Measurement level**			
	Str.	Nrc.	Date	Other	Nom.	Ord.	Sc.	Other
Time			4		4			
Location	1			2				3
Ship	6	2			6		2	
Event	2				1	1		
Cause	2				2			
Other	1				1			
Consequences	1			1	1			1
Total	13	2	4	3	15	1	2	4

* Variable type: Str. (String variables whose values are not numeric and therefore are not used in calculations), Nrc. (Numeric variables). ** Variable measurement level: Nom. (Nominal variables whose values represent categories with no intrinsic ranking), Ord. (Ordinal variables whose values represent categories with some intrinsic ranking), Sc. (Scale variables whose values represent ordered categories with a metric) (SPSS 16.0 for Windows 2007)

Table 3. The main categories, types and measurement levels of variables in the FMA DB1 (1990-1996).

Main category	Variable Type*				Measurement level**			
	Str.	Nrc.	Date	Other	Nom.	Ord.	Sc.	Other
Time			5		5			
Location	4			2	4			2
Ship	11	2			11		2	
Ship activity	2				2			
Exposure		2					2	
Cargo	1	1			1		1	
Event	3				3			
Cause	4				4			
Other	1				1			

Environment	4				1	3		
Consequences	9	6		2	3	6	6	2
Response	1				1			
Total	40	11	5	4	36	9	11	4

*, ** See the foot note in Table 2

Table 4. The main categories, types and measurement levels of variables in the FMA DB2 (1997-2007).

Main category	Variable Type*				Measurement level**			
	Str.	Nrc.	Date	Other	Nom.	Ord.	Sc.	Other
Time			5		5			
Location	3			2	3			2
Ship	10	3	2		12		3	
Ship activity	2				2			
Exposure		2					2	
Cargo	1				1			
Event	3				3			
Cause	4				4			
Other	1				1			
Environment	4	1			1	3	1	
Consequence	11	19		4	11		19	4
Total	39	25	7	6	43	3	25	6

*, ** See the foot note in Table 2

Table 5. The main categories, types and measurement levels of variables in the SMA DB (1996-2007).

Main category	Variable Type*				Measurement level**			
	Str.	Nrc.	Date	Other	Nom.	Ord.	Sc.	Other
Time			3		3			
Location	10			2	10			2
Ship	13	3			13		3	
Ship activity	3				3			
Exposure		3					3	
Cargo	2				2			
Event	4				3	1		
Cause				1				1
Other	3				3			
Environment	5	1			3	2	1	
Consequence	8	26		1	5	3	26	1
Total	48	33	3	4	45	6	33	4

*, ** See the foot note in Table 2

Table 6. The main categories, types and measurement levels of variables in the HELCOM DB (1989-2006).

Main category	Variable Type*				Measurement level**			
	Str.	Nrc.	Date	Other	Nom.	Ord.	Sc.	Other
Time			4		4			
Location	1			2	1			2
Ship	10	7			10		7	
Cargo	1				1			
Event	2			1	2			1
Cause	4				4			
Other	3				3			
Consequences	6	1			5	1	1	
Total	27	8	4	3	30	1	8	3

*, ** See the foot note in Table 2

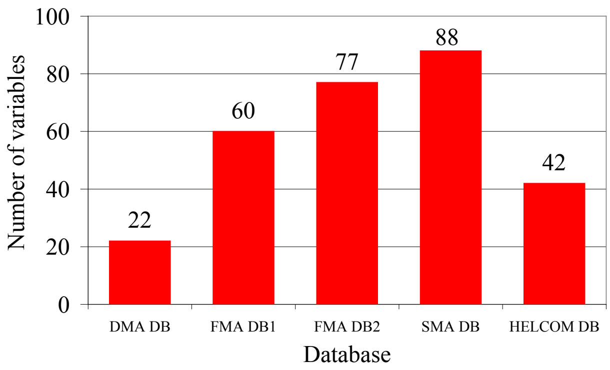

Figure 1. The number of variables in databases.

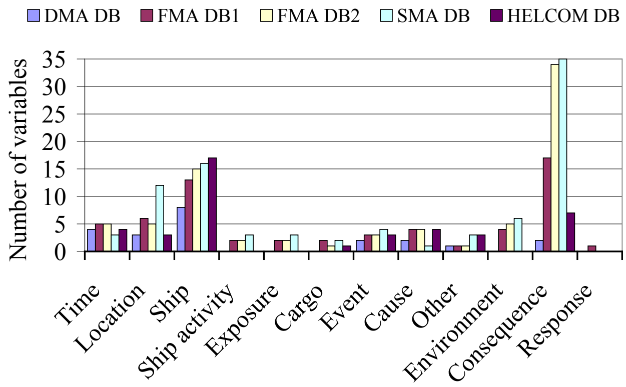

Figure 2. Comparison among the main categories of variables in databases.

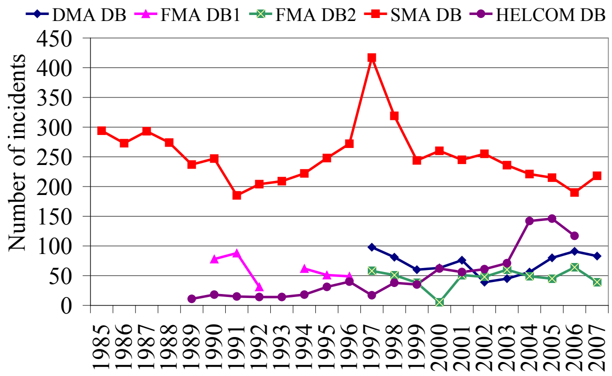

Figure 3. Comparison among periods and numbers of marine incidents recorded in the databases.

4 INCIDENT DATABASES IN THE USA

The study of many incident databases (see Mullai 2004) show that, in terms of the public accessibility and amount, diversity, accuracy, quality of dangerous goods risk-related data, the USA is one of the most advanced countries in the world. Many types of data are free of charge and available for public use in the Internet. The USA Freedom of Information Act (1974) requires all federal and national organisations to make data available in electronic form to the public. Hazardous Material Information System (HMIS) and National Response Center (NRC) databases are two of the USA's and the world's largest databases in the field. They are available to both scientific and practitioner communities. The HMIS database (ca 200,000 case histories organised in more than 180 variables) records all dangerous goods incidents occurring in all modes of transport. The NRC database (over a half million case histories organised in more than 230 variables) records all incidents involving all types of hazmat discharges into the environment anywhere in the USA and its territories. The data are reported by individuals and a wide range of organisations and agencies, and cover a wide range of systems of the USA's chemical supply chain. In contrast to the BSR's databases, both the USA's databases offer many advantages, including massive, diverse, high quality and very well organised data, no restriction and easy data access via the Internet, and very convenient data format. Our experience (see Mullai & Larsson 2008) shows that data preparation and analysis are significantly less time consuming, resource and labour intensive than working with the BSR's databases. The incidents recorded to all BSR's databases combined are only a small fraction of the HMIS and NRC databases.

5 CONCLUSIONS AND RECOMMENDATIONS

With reference to the research questions, merging all databases into a single dataset and performing a detailed risk analysis for the BSR may not be possible due to issues explored in this study. However, a risk analysis based on partly merged datasets is feasible. Some of the issues are partly attributed to different practices, priorities and languages. The marine environment and safety issues are gaining more attention in the region. In order to tackle some of the issues, we suggest the following solutions: (i) Enhance cooperation among maritime authorities and other parties in the region. Projects like the DaGoRus project and conferences like the TransNav 09 can contribute to cooperation. They can serve as forums where problems and solutions are identified and discussed, stakeholders meet and information is disseminated. (ii) Improve and harmonise marine incident databases in the BSR. Immediate changes cannot be expected in the near future as several databases are designed based on the established coding system. Significant changes may render many years (two decades) of data records incompatible. Therefore, the process should be well studied and performed in a stepwise manner. (iii) Marine incident data should be made publicly available in electronic format via the Internet, at least for the research purposes. The USA experience can serve as an inspirational example. (iv) Upgrade data compilation systems. (v) Improving the HELCOM DB is a good solution, which include reporting all marine incidents (accidental and deliberate events) occurring in the BSR, improving the quality of variables and ensuring a higher degree of data completeness. The maritime risks, including risks due to the large and increasing amounts of dangerous goods, deserve a better understanding and management. These can be achieved only by united efforts.

ACKNOWLEDGEMENT

We wish to thank the DaGoRus project office for financing this study, and Danish, Finnish, HELCOM and Swedish authorities for their assistance in data collection.

REFERENCES

HELCOM (Helsinki Commission) 2005. *Overview of the ships' traffic in the Baltic Sea*

HSE (Health and Safety Executive) 1991. Major Hazard Aspects of Transport of Dangerous Substances. *Report of Advisory Committee on Dangerous Substances, Health and Safety Commission*, HMSO, UK

IMO (International Maritime Organisation) 1996. Reports on Marine Casualties. Harmonized Reporting Procedures, Annex 3, Draft MSC/MEPC Circular, IMO FSI4/18

Kaplan S. and Garrick B.J. 1981. On the quantitative definitions of risks. *Journal of Risk Analysis*, Vol.1, No. 1, pp. 11-27

Kostianoy A.G., Litovchenko K.Ts., Lebedev SA Stanichny S.V., Soloviev D.M., Pichuzhkina O.E. 2005. Operational Satellite Monitoring of Oil Spill Pollution in the Southeastern Baltic Sea. *Oceans - Europe*, IEEE, pp. 182-183

Looström B. 1983. Swedish Remote Sensing Systems for Oil Spill Surveillance at Sea. *Oil and Petrochemical Pollution*, p. 235

Mullai A. 2004. A Risk Analysis Framework for Maritime Transport of Packaged Dangerous Goods (PDG), in Brindley, C (2004) Supply Chain Risk, *Ashgate Publishing Company*, UK, Chapter 9, pp130-159

Mullai A. and Larsson E. 2008. Hazardous Materials Incidents – Some Key Results of a Risk Analysis. *WMU Journal of Maritime Affairs*, pp. 65-108

Mullai A. and Paulsson U. 2002 Oil Spills in Öresund – Hazardous Events, Causes and Claims. *Report on the SUNDRISK Project, Lund University Centre for Risk Analysis and Management* (LUCRAM), Sweden, pp. 1-160

Rytkönen J., Siitonen L., Riipi T., Sassi J., and Sukselainen J. 2002. Statistical Analyses of the Baltic Maritime Traffic. Project financed by Finnish Environment Institute and Ministry of Traffic and Communications.

Soomer, H. Ranta, H. Penttila, A. 2001. Identification of victims from the M/S Estonia. *International Journal of Legal Medicine*, Vol. 114, Issue 4-5, pp. 259-262, Springer

TSE (Turku School of Economics, Finland) 2006. DaGoB Project (Safe and Reliable Transport Chains of Dangerous Goods in Baltic Sea Region): http://www.dagob.info/2006

Von Viebahn C. and Gade M. 2000 A proposal for improved oil spill monitoring of peripheral marine regions in the Baltic Sea by in-service aircraft. *Geoscience and Remote Sensing Symposium Proceedings*. IEEE international, Vol. 1, pp 342-344