# Injuries at Sea: a Geo-spacial Analysis of Marine Accidents

V. Enerstvedt[1], H. Jia[1]
Department of Business and Management Science, Norwegian School of Economics, Bergen, Norway
(Vegard.Enerstvedt@nhh.no)

*Abstract* – **Marine safety is of paramount importance. Accidents on board a vessel can have a great impact on, or even end, the life of a sailor. This is not only a significant personal tragedy, but it can also have an impact on the safety of the entire crew and the operations of the vessel. In this paper, we use data from the Norwegian Maritime Authority to investigate the locations of accidents. We can observe that a significant number of accidents occurred off the coast of Western and Northern Norway. We further investigate how accidents occur in clusters and conclude that both fatal and non-fatal accidents are clustered. Lastly, we investigate the optimal locations for search and rescue helicopters to be stationed in order to minimize the expected distance between accidents and airports. Our results suggest that additional helicopters should be stationed in Northern Norway.**

*Keywords* – **Safety, accidents, data analysis.**

## I. INTRODUCTION

The ocean gives, and the ocean takes away. The sea provides us with access to vast natural resources, such as petroleum and fish, and has served as a transportation route for millennia. Transporting vast amounts of cargo every day. However, the sea is also a perilous place. Many have lost their lives at sea. This paper aims to better understand the geographic distribution of accidents. We utilize data obtained from the Norwegian Maritime Authority (NMA). The data is publicly available and contains information on over 27 000 accidents since 1981.

Marine safety research is an important field of study. Works on this topic often focus on regulatory or cultural aspects, as seen in [1], and [2]. There has been research conducted on marine safety using the NMA data, for example [3]–[5]. These papers, on the other hand, focus on aspects such as vessel characteristics and other operational factors. To the best of our knowledge, this is the first research to study the geographic distribution of NMA accidents. We create geospatial plots and heatmaps and conduct a cluster analysis on the geographical distribution of fatal and non-fatal accidents. Our hypothesis of a correlation between areas of high activity, such as Western and Northern Norway, and the frequency of accidents is confirmed by the clustering behavior of both fatal and non-fatal accidents. Lastly, we investigate the optimal distribution of search and rescue (SAR) helicopters. Having the optimal base structure is crucial in the event of an accident, as every minute counts and it can save lives.

This research fills in the gap in the literature concerning the geographic distribution of marine accidents and its association with SAR resources by analyzing a unique dataset spanning over 40 years.

## II. METHODOLOGY

### *A. Data*

This paper is based on historical accident statistics from the Norwegian Maritime Authority (NMA). This is a publicly available register of accidents that have occurred on vessels sailing under the Norwegian flag or within Norwegian territorial waters (NTW) [6]. The dataset covers the accidents that occurred in the period 1981-2021. The publicly available data only goes back to 1981. This means that, for example, the Alexander Kielland (a major drilling rig disaster killing over 100 in 1980) accident is not included in the data. This highlights a potential issue where accidents with low probability, but high casualty may be underrepresented in the data.

The dataset contains 33 variables, which include vessel details, such as name, International Maritime Organization (IMO) number, callsign, length, tonnage, year of built, and flag. Additionally, it provides information about the accident, including the geographic area, coordinates, whether the accident resulted in fatalities, the vessel's state of operations, and the type of accident. Note that accidents with multiple injuries are recorded as one entry per injured person.

There are some issues with the data. out of the total record of 27,226 accidents, only 4141 accidents have recorded latitude and longitude. Moreover, accidents are reported manually. This means that human error may be present in the data. A particularly problematic issue is the mix-up of latitude and longitude. This is where the "geographic area" variable comes in handy. If the reported location does not match the coordinates, we switch them around.

Reporting accidents is regulated by law [7]. However, assessing the reporting practices can be difficult. Hence, we do not know how reporting practices vary among vessels, operators, and even over time. Vessels sailing under non-Norwegian flags outside NTW are not subject to the reporting regime at all and are therefore excluded from the analysis. This leaves 3575 accidents.

We chose to focus on the vessels within the Norwegian Economic Zone (NEZ) because of their natural relevance to search and rescue (SAR) operations. Therefore, we only examined accidents that occurred within the NEZ

*Fig 1 Location of injuries on Norwegian vessels in NEZ (1981-2021)*

geographic area. Note that accidents on vessels at berth are also included in the data. Because of the application to SAR operations, we only consider vessels that were classified as "underway". After excluding all accidents on non-Norwegian vessels, outside of NEZ, vessels not underway, or those with unknown positions, 2422 accidents remain.

### B. Heatmap

The heatmap enables us to visualize the locations where most accidents have occurred. We use kernel estimation methods to create a heatmap. This is a technique for empirically estimating the intensity function. In general, the method works by estimating the following:

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - X_i}{h})$$

Where, $X_1, X_2, \ldots, X_n$ are the coordinates for accident i, h is the bandwidth and K is some kernel function, [8]. In this case, the Gaussian kernel is a suitable option. To balance our interest in accuracy with the dangers of overfitting this sparse dataset, we employ Silverman's "rule-of-thumb" [9] using standard deviations of 10,000 and 50,000.

### C. Cluster Analysis

Cluster analysis is important for understanding the statistical patterns of accidents. Clustering refers to the phenomenon of accidents occurring in close proximity to each other. If we can detect clusters of accidents through statistical analysis, it would be beneficial to concentrate SAR resources in the vicinity of the cluster's center. This approach would minimize the distance that rescuers need to travel during an emergency. On the other hand, if accidents are not clustered, then the most effective allocation of resources is to distribute them as evenly as possible.

Observations within one cluster have shorter distances to their neighbors than those in other clusters. Therefore, the methodology used in this clustering analysis is based on analyzing the distance between each accident and its geographically closest neighboring accident. The results are presented by summary statistics and plots, which include histogram graph, boxplot, and the empirical cumulative distribution function (CDF). This allows us to investigate various aspects of the clustering measures such as mean, standard deviation, and the Clark-Evans index.

The Clark-Evans index is a metric developed by Phillip J. Clark and Francis C. Evans. The original motivation was to study the aggregation of populations of plants or animals to gain insights into the behavior of the species [10]. The Clark-Evans index calculates the ratio between the average distance of all points to their nearest neighbor and the theoretical average distance between a point and its closest neighbor if the points are randomly distributed. Hence,

$$R = \frac{\overline{r_A}}{\overline{r_E}},$$

Where, $\overline{r_E}$ is the theoretical average distance between a point and its closest neighbor given that they are randomly distributed. In fact, $\overline{r_E} = \frac{1}{2\sqrt{\rho}}$, where $\rho$ depends on the proposed random distribution. $\overline{r_A} = \frac{1}{n}\sum_i^n r_i$, where $r_i$ is the distance between point $i$ and its closest neighbor. It can be shown that R varies between 0 (maximum aggregation) and 2.1491 (maximum dispersion).

Moreover, the Clark-Evans index can be utilized to determine if there is statistically significant clustering. The test statistic in this case is:

$$c = \frac{\overline{r_A} - \overline{r_E}}{\sigma_{\overline{r_E}}}$$

Where, $\sigma_{\overline{r_E}}$ is the standard error of the average distance to the nearest neighbor if the points are randomly distributed. The critical values are determined using the standard normal distribution. There is also a two-population version of this test [10].

### D. Optimal distribution of SAR resources

A wide range of SAR resources are available in case of an emergency, both at sea and in the air. For example, private ships, fishing vessels, naval vessels, transport helicopters, and dedicated SAR helicopters. To limit the scope of this paper, we focus on the dedicated SAR helicopters operated by the Royal Norwegian Air Force (RNoAF). SAR is one of the most crucial tasks performed by the RNoAF during peacetime. This SAR operation is organized as part of the 330 Squadron and operates out of 6 airports: Rygge in Moss, Sola, Florø, Ørland, Bodø, and Banak in Lakselv. The 330 Squadron has been operating the Westland WS-61 Sea King since 1973. However, they are to be retired due to their age. The RNoAF is currently replacing its Sea King with the new AW101 SAR Queen [11].

The final section of the paper aims to investigate the optimal distribution of air bases for SAR helicopters. We take a naïve approach and consider all possible combinations of public airports in Norway. Then we will use brute force to determine the average distance between the accidents and the closest airport. The combination with the lowest average distance is the preferred one.

## III. RESULTS

### E. *Heatmap*

Using the method described in Section II. B, we can determine the intensity of accidents, including fatal and non-fatal accidents. Fig 2 shows the distribution of non-fatal accidents, and Fig 3 shows the distribution of fatal accidents.

The heatmap provides a graphic interpretation of where accidents occur. It might be tempting to interpret this as indicating the most dangerous zone. This does not necessarily have to be the case. A seemingly safe area can still experience many accidents due to high traffic volume. More traffic naturally leads to more accidents. If an area is

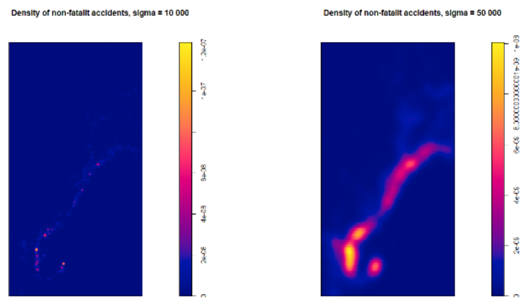


*Fig 2 Density estimates of non-fatal accidents on Norwegian flagged vessels in the NEZ*

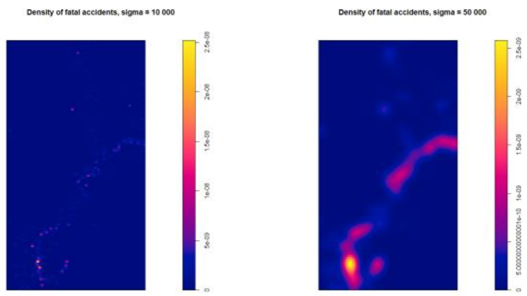


*Fig 3 Density estimates of fatal accidents on Norwegian flagged vessels in the NEZ*

known to be safe, the number of vessels sailing through it increases, and as a result, the number of accidents also increases. Therefore, one should interpret the heatmap only as an indication of where accidents occur, and not as a measure of which areas are the most hazardous.

We can observe that accidents are dispersed throughout the coast, but there is a greater concentration of accidents outside the coasts of Northern and Western Norway. As well as a small area outside Telemark. There are fewer accidents off the coast of Trøndelag and Central Norway.

The area outside Western Norway has a significant amount of petroleum activity. The high petroleum activity can account of some of the concentration of accidents [12]. Moreover, the Stadt area is infamously perilous to navigate.

There are many accidents in the area outside Troms and Finnmark. This makes sense because this area has a lot of fishing activity. The Barents Sea is typically frequented by larger oceangoing vessels, whereas coastal fishing also holds significant importance [13]. This could help explain the intensity of accidents in this area.

A vast amount of the goods that flow in and out of Norway are transported through ports in Eastern Norway. Ports in cities such as Moss, Fredrikstad, Porsgrun, and Oslo account for over 60% of the cargo coming from foreign ports [14]. This means that there is a large amount of cargo shipping in the Skagerrak and the Oslofjord. This activity might explain the high number of accidents in this area.

There are differences between fatal and non-fatal accidents. We can observe that fatal accidents appear to be more concentrated. In particular, around the southern part of the North Sea. A possible explanation could be that petroleum-related operations are inherently more dangerous, or the sheer volume of activity increases the risk of accidents. Interestingly, the concentration of fatal accidents outside the coast of Western Norway appears to be further south than the concentration of non-fatal accidents. There also appear to be fewer fatal accidents off the coast of Trøndelag, Helgeland, and in the inner Oslofjord. The reasons for this phenomenon are unclear, but possible explanations include the ability to conduct SAR, the inherent safety of sailing in the area, or a lower volume of maritime activity.

Note that the accidents appear to be clustered, the fatal accidents exhibiting a higher degree of clustering than the non-fatal ones.

### F. *Cluster Analysis*

TABLE I
EXPLORATORY STATISTICS ON NEAREST NEIGHBOUR

| | Fatal accidents | Non-fatal accidents |
|---|---|---|
| Mean | 8.39 nm | 4.37 nm |
| Median | 1.95 nm | 1.64 nm |
| Standard deviation | 16.27 nm | 8.53nm |

TABLE II
CLARK-EVANS INDEX FOR FATAL AND NON-FATAL ACCIDENTS

| | | Clark-Evans Index |
|---|---|---|
| Fatal: | | |
| | *Aggregate* | $0.447^{***}$ |
| | *Separate* | $0.283^{***}$ |
| Non-fatal: | | |
| | *Aggregate* | $0.409^{***}$ |
| | *Separate* | $0.354^{***}$ |

*Note: *** indicates statistically significant at 99% confidence level*

Given that the NMA database records every casualty as a distinct entry, multiple casualties or fatalities involved in the same accident are recorded as individual entries in the database. Therefore, we test for clustering both when considering the separate accidents and when aggregating multi-casualty accidents.

Despite the expected outliers as observed in the histogram and boxplots, most accidents appear to be clustered. The clustering feature is confirmed by the Clark-Evans test results. The Clark-Evans index value reflects the degree of clustering of the observations. The smaller index value suggests that the data points in the cluster are more congregated. The Clark-Evans index is 0.354 for non-fatal accidents and 0.283 for fatal accidents. If we consider incidents involving multi-casualties as a single accident. The Clark-Evans index is 0.409 for non-fatal accidents and 0.447 for fatal accidents. This is far below the rule-of-thumb that $R < 1$ indicates clustering. The p-value for all four hypotheses tests is close to zero [10].

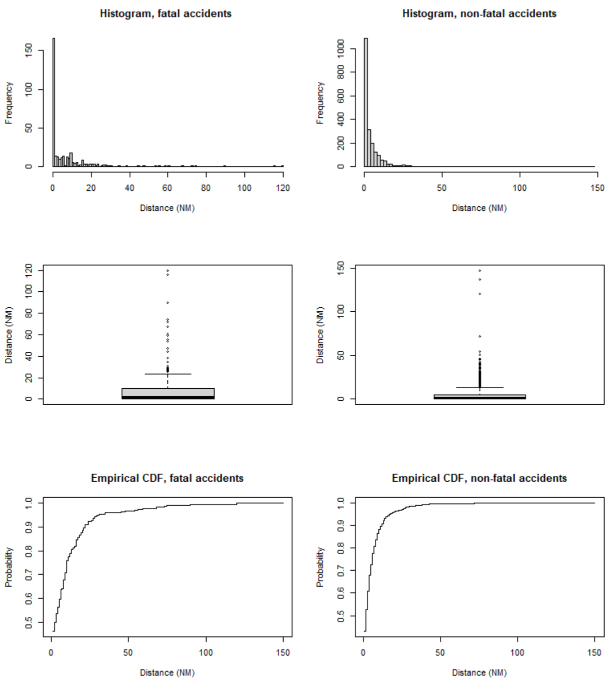


*Fig 4 Exploratory plots for fatal (L) and non-fatal (R ) accidents. Top row: histogram. Middle row: boxplot. Bottom row: Empirical cumulative distribution function.*

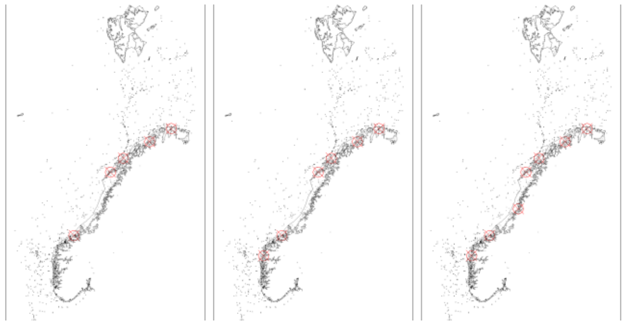

*Fig 5 The optimal distribution of airports given 5, 6, or 7 airports.*

It can be concluded that all types of personal accidents on Norwegian-flagged vessels in the NEZ are clustered.

## G. *Optimal distribution of SAR helicopters*

In this final section, we attempt to provide a recommended distribution of SAR helicopters such that the average distance to accidents is minimized. Since accidents tend to occur in clusters, we can utilize the locations of previous accidents to estimate the likely travel distances to future accidents. Note that each observation entry records a single casualty. Therefore, the severity of any given accident is reflected by the number of observations at the specific time and location.

We use a brute-force approach to identify all possible combinations of public airports in Norway. Then we calculate the average distance between accidents and the closest airport. Lastly, we identify the combination that has the smallest average distance.

Fig 5 and Table 3 present the optimal geographic distribution of SAR helicopters, for n = 5, 6, and 7 locations. Our results suggest that it is more important to have good coverage of Northern Norway. This is partly due to the nature of accidents in this region. They tend to be further from land. This means that the difference between the distances from the closest and second closest airports can be significant. On the other hand, accidents occurring outside of Western Norway tend to be closer to the coast. Hence, the additional distance of flying from Florø compared to Stavanger is not significant.

TABLE III
CURRENT AND OPTIMAL DISTRIBUTION OF SAR HELICOPTERS

| Current distribution | N = 5 | N = 6 | N=7 |
|---|---|---|---|
| Lakselv | Andøya | Andøya | Andøya |
| Bodø | Hasvik | Florø | Brønnøysund |
| Ørland | Kristiansund | Hasvik | Florø |
| Florø | Leknes | Kristiansund | Hasvik |
| Stavanger | Mehamn | Leknes | Kristiansund |
| Moss | | Mehamn | Leknes |
| | | | Mehamn |

# IV. DISCUSSION

This paper is based on self-reported marine accident records. Even though reporting is regulated and mandated by law, practices can vary. In particular, we have reason to believe that reporting has improved over the last couple of years [6]. The absence of geographic locations for most of the accidents (only 15% of accidents have coordinates) poses a problem. It is reasonable to assume that reporting will improve with the advancement of technology, and as a result, the quality of the reports will also improve.

Consider the density plots. It appears that most injuries occur near the coast. However, outside of Northern and Western Norway, many injuries occur in more remote areas. This is most likely due to the amount of traffic in this area, and the type of activity, which includes petroleum and fishing. An interesting extension of this would be to determine which areas are the most dangerous.

One issue that arises when using the Clark-Evans test is how to handle accidents that occur near the edge. Accidents that occur closer to the edge naturally have fewer neighboring accidents compared to those that occur in the center of the area of study. The closest neighbor might be outside the area of study. There are different approaches to addressing this issue, but given the robustness of our results, this should not be a challenge.

As previously mentioned, we use a naïve brute force algorithm to determine the optimal distribution. There are certainly more efficient and supplicated ways of solving this problem. Moreover, despite our recommendations, other factors also influence the choice of organization. For example, the logistics of the site, weather conditions, availability of skilled labor, and other factors must be considered.

## V. CONCLUSION

In this paper, we investigate the geographic distribution of accidents on Norwegian flagged vessels in the NEZ in the period from 1981 to 2021. We have observed that the majority of accidents occur near the coastline, although certain regions experience a higher frequency of accidents than others. These areas, with a higher frequency of accidents, are located off the coasts of Western and Northern Norway. These are areas with significant petroleum and fishing activities. We have observed that there is a distinction between fatal and non-fatal accidents. Fatal accidents are more concentrated than non-fatal accidents. The southern part of the North Sea experiences a particularly high number of fatal accidents. However, we also observe a higher number of fatal accidents occurring outside the coast of Northern Norway and in Skagerrak.

We investigated the level of clusteredness for both fatal and non-fatal accidents. To do this, we studied the distance between an accident and its nearest neighbor. Most fatal accidents seem to occur in the same location as their closest neighbor. These correspond to accidents resulting in multiple fatalities. We calculated the Clark-Evans index for fatal and non-fatal accidents, both when treating multi-casualty accidents as separate events and as singular accidents. We conclude that all accidents are heavily clustered.

Lastly, we attempted to determine the optimal distribution of SAR helicopters. We conclude that increasing the number of station helicopters in Northern Norway would improve the average distance from the base to an accident. In particular, Andøya should be considered.

Future research could employ more data sources and analytic methods. Using data from the vessel tracking Automated Information System (AIS) could help estimate the probability of accidents happening. Further investigations on the trends and periodicity of accidents would be beneficial for multiple stakeholders. Lastly, an optimization method could be utilized to find the optimal distribution of helicopters.

## REFERENCES


[1] T.-O. Nævestad, "Safety culture, working conditions and personal injuries in Norwegian maritime transport," *Mar. Policy*, vol. 84, pp. 251–262, Oct. 2017, doi: 10.1016/j.marpol.2017.07.019.

[2] K. V. Størkersen, S. Antonsen, and T. Kongsvik, "One size fits all? Safety management regulation of ship accidents and personal injuries," *J. Risk Res.*, vol. 20, no. 9, pp. 1154–1172, Sep. 2017, doi: 10.1080/13669877.2016.1147487.

[3] R. J. Bye and A. L. Aalberg, "Maritime navigation accidents and risk indicators: An exploratory statistical analysis using AIS data and accident reports," *Reliab. Eng. Syst. Saf.*, vol. 176, pp. 174–186, Aug. 2018, doi: 10.1016/j.ress.2018.03.033.

[4] E. McGuinness, H. L. Aasjord, I. B. Utne, and I. M. Holmen, "Injuries in the commercial fishing fleet of Norway 2000–2011," *Saf. Sci.*, vol. 57, pp. 82–99, Aug. 2013, doi: 10.1016/j.ssci.2013.01.008.

[5] P. Stornes, "Risk influencing factors in maritime accidents : An exploratory statistical analysis of the Norwegian Maritime Authority incident database," *124*, Nov. 2015, Accessed: May 16, 2023. [Online]. Available: https://samforsk.brage.unit.no/samforsk-xmlui/handle/11250/2360509

[6] "Ulykkesstatistikk - Sjøfartsdirektoratet." https://www.sdir.no/sjofart/ulykker-risiko-og-sikkerhet/ulykkesstatistikk/ (accessed Dec. 06, 2022).

[7] "Forskrift om melde- og rapporteringsplikt ved sjøulykker og andre hendelser til sjøs - Lovdata." https://lovdata.no/dokument/SF/forskrift/2008-06-27-744 (accessed Dec. 19, 2022).

[8] T. J. Hastie, *The elements of statistical learning : data mining, inference, and prediction*, 2nd ed. in Springer series in statistics. New York: Springer Science + Business Media, 2009.

[9] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. CRC Press, 1986.

[10] P. J. Clark and F. C. Evans, "Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations," *Ecology*, vol. 35, no. 4, pp. 445–453, 1954, doi: 10.2307/1931034.

[11] "Luftforsvaret," *Forsvaret*. https://www.forsvaret.no/om-forsvaret/organisasjon/luftforsvaret (accessed May 18, 2023).

[12] Directorate of Fisheries, "NPD FactMaps Desktop." https://factmaps.npd.no/factmaps/3_0/ (accessed May 16, 2023).

[13] "Fiskeri." https://portal.fiskeridir.no/portal/apps/webappviewer/index.html?id=ea6c536f760548fe9f56e6edcc4825d8 (accessed May 16, 2023).

[14] "03648: Havnestatistikk. Gods, etter havn, containertype og innenriks-/utenriksfart 2003K1 - 2022K4. Statistikkbanken," *SSB*. https://www.ssb.no/statbank/table/03648 (accessed May 15, 2023).