A Visual Analytics Process for Maritime Resource Allocation and Risk Assessment

Abish Malik *, Ross Maciejewski†, Member, IEEE, Ben Maule †, David S. Ebert †, Fellow, IEEE

*Purdue University Visualization and Analytics Center (PURVAC) † United States Coast Guard

ABSTRACT
In this paper, we present our collaborative work with the U.S. Coast Guard’s Ninth District and Atlantic Area Commands where we developed a visual analytics system to analyze historic response operations and assess the potential risks in the maritime environment associated with the hypothetical allocation of Coast Guard resources. The system includes linked views and interactive displays that enable the analysis of trends, patterns and anomalies among the U.S. Coast Guard search and rescue (SAR) operations and their associated sorties. Our system allows users to determine the potential change in risks associated with closing certain stations in terms of response time, potential lives and property loss and provides optimal direction as to the nearest available station. We provide maritime risk assessment tools that allow analysts to explore Coast Guard coverage for SAR operations and identify regions of high risk. The system also enables a thorough assessment of all SAR operations conducted by each Coast Guard station in the Great Lakes region. Our system demonstrates the effectiveness of visual analytics in analyzing risk within the maritime domain and is currently being used by analysts at the Coast Guard Atlantic Area.

Keywords: Visual analytics, risk assessment, Coast Guard

1 INTRODUCTION
As modern datasets increase in size and complexity, it becomes increasingly difficult for analysts and decision makers to extract actionable information for making effective decisions. In order to better facilitate the exploration of such datasets, tool sets are required that allow users to interact with their data and assist them in their analysis. Furthermore, such datasets can be utilized to explore the consequences and risks associated with making decisions, thereby providing insights to analysts and aiding them in making informed decisions.

Besides the sheer volume and complexity of such datasets, analysts must also deal with data quality issues, including uncertain, incomplete and contradictory data. Moreover, analysts are often faced with different decisions and are required to weigh all possible consequences of these decisions using such datasets in order to arrive at a solution that minimizes the associated risks within a given time constraint. Using traditional methods of sifting through sheets of data to explore potential risks can be highly inefficient and difficult due to the nature and size of these datasets. Therefore, advanced tools are required that enable a more timely exploration and analysis. Our work focuses on the use of visual analytics [17, 31] in the realm of risk assessment and analysis and demonstrates the effectiveness of visual analytics in this domain. The work described in this paper is based on the application of visual analytics to analyze historic response operations and assess the potential risks in the maritime environment based on notational station closures. Our work was done in collaboration with the U.S. Coast Guard’s Ninth District and Atlantic Area Commands that are responsible for all Coast Guard operations in the five U.S. Great Lakes. In particular, we focused on the Auxiliary stations that are staffed by Coast Guard volunteers and civilians. These Auxiliary stations assist their parent stations in their operations and usually operate on a seasonal basis using a small fleet of boats for conducting their operations. However, the number of Auxiliary personnel that volunteer their time at these stations has decreased over recent years. This has required Coast Guard analysts to develop possible courses of action and analyze the risks and benefits with each option. Several options include seasonal or weekend only staffing of these units, or at worst, closure. Closure, however, may involve increased risks to the boating public and a complete analysis of the risks associated with closing an Auxiliary station needs to be evaluated. The results of this type of analysis would assist the decision makers in determining the optimal course of action.

In particular, the analysts are interested in determining the spatial and temporal distribution of response cases and their associated sorties (a boat or an aircraft deployed to respond to an incident) for all SAR operations conducted in the Great Lakes and how closing certain Auxiliary stations affects the workload of the stations that absorb these cases. Coast Guard policy mandates the launch of a sortie within 30 minutes and have an asset (boat or aircraft) on scene within two hours of receiving a distress call [32]. Closing these stations implies a potential for longer response times that could potentially translate into the loss of lives and property.

To address these challenges, we developed a visual analytics system that supports decision making and risk assessment and allows an interactive analysis of trends, patterns and anomalies among the U.S. Coast Guard’s Ninth District operations and their associated sorties. Our system, shown in Figure 1, allows enhanced exploration of multivariate spatiotemporal datasets. We have incorporated enhanced tools that enable maritime risk assessment and analysis. Our system includes linked spatiotemporal views for multivariate data exploration and analysis and allows users to determine the potential increase or decrease in risks associated with closing one or more Coast Guard stations. The system enables a thorough assessment of all operations conducted by each station. In addition, the system provides analysts with the tools to determine which Coast Guard stations are more optimally suited to assume control of the operations of the closed station(s) by comparing the distances from available stations to all SAR cases previously handled by the closed station(s). Our system features include the following:

- Risk profile visualizations and interactive risk assessment tools for exploring the impact of closing Coast Guard stations
- Optimization algorithms that assist with the interactive exploration of case load distribution in resource allocation
- Linked filters combined with spatial and temporal views for interactive risk analysis/exploration

Our work focuses on providing analysts with interactive visual analytics tools that equip them to deal with risk assessment scenar-
ios associated with closing Coast Guard stations. We emphasize that although our risk assessment toolkit and the examples given in this paper have been based in the maritime domain, these techniques apply equally as well to other domains (e.g., criminal offense analysis, syndromic surveillance).

2 RELATED WORK

In recent years, there has been a rapid growth in the development of new visual analytics tools and techniques for advanced data analysis and exploration (e.g., [30, 34]). From traditional scatterplots [8] and parallel coordinate plots [16] to tools like Theme River [15] and spiral graphs [7], these systems incorporate different forms of visualizations to provide enhanced analytical tools to users. Although these tools allow users to explore their data and assist them in their decision making process, researchers have only recently started to employ visual analytics techniques for risk-assessment and decision-making domain that allow users to perform a thorough analysis of risks associated with different decisions.

Migut and Worring [21] propose an interactive approach to risk assessment where they demonstrate a risk assessment framework that integrates interactive visual exploration with machine learning techniques to support the risk assessment and decision making process. They use a series of 2D visualizations including scatterplots and mosaic plots to visualize numerical and ordinal attributes of the datasets. While the authors demonstrate the effectiveness of using visual analytics in the field of risk assessment, their work is mainly focused on building classification models that users may interactively use to classify their data entities and visualize the effects on their classification. Gandhi and Lee [13] also apply visual analytics techniques to the realm of requirements-driven risk assessment. Specifically, they use cohesive bar and arc graphs to illustrate the risks due to the cascading effects of non-compliance with Certification and Accreditation requirements for the U.S. Department of Defense. Sanusi and Mustafa [25] introduce a framework to develop a visualization tool that may be used for risk assessment in the software development domain. Their proposed framework allows users to identify the components of the software system that are likely to have a high fault rate. Direct visualizations of risk use tools like bar graphs and confidence interval charts to visualize measures of risk and are usually constructed using spreadsheet programs like Microsoft Excel [12, 13]. Although widely used, these techniques fail to work for our purposes primarily due to the nature of the risk analysis that is required. The Coast Guard SAR dataset is spatiotemporal in nature and the exploration of risk requires domain knowledge that is difficult to incorporate algorithmically.

With respect to the temporal nature of risk assessment, researchers have also developed different visualization systems that allow users to explore risks associated with financial decisions related to investments and mutual funds, among other financial planning scenarios. Rudolph et al. [24] propose a personal finance
decision making visual analytics tool that allows users to analyze both short-term and long-term risks associated with making investment decisions. Savikhin et al. [26] also demonstrate the benefits of applying visual analytics techniques to aid users in their economic decision making and, by extension, to general decision making tasks. Both of the previous examples only explore temporal datasets. In this work, we apply visual analytics techniques to explore risks using multivariate spatio-temporal datasets that guide analysts in making complex decisions.

As is the case with most multivariate datasets, data tends to be inherently unreliable, incomplete and contradictory. In order to reach to correct conclusions, analysts must take these into account in their analysis. In this regard, Correa et. al. [9] describe a framework that supports uncertainty and reliability issues in the different stages of the visual analytics process. They argue that with an explicit representation of uncertainty, analysts can make informed decisions based on the levels of confidence of the data. Our system factors data reliability issues in the risk assessment process and provides confidence levels at all stages of risk assessment that, in turn, enable analysts to better understand the underlying nature of the data and guides them in making effective decisions.

There also exist many geospatial and temporal analytical systems that provide users with the ability to explore their spatiotemporal datasets in order to find patterns and provide an overview of the data in a visual analytics platform (e.g., [1, 2, 3, 14]). As the needs of our end users are unique, this warrants developing a stand alone system to address the challenges faced by the Coast Guard analysts. We plan on further examining these robust geo-temporal analysis tools and the degree to which they can be extended to meet the Coast Guard requirements that have been identified in this paper.

There has also been much work done in visualizing large datasets using interactive cross-filtered and linked views that allow users to explore their datasets. Stasko et. al. [30] use multiple coordinated views of documents to reveal connections between entities across different documents. Eick and Johnson [10] utilize multiple linked views to visualize abstract, non-geometric datasets in order to reduce visual clutter and provide users with insights into their datasets. Eick and Wills [11] also demonstrate the effectiveness of linking and interaction techniques in the visualization of large networks. Our system utilizes these practices and allows users to interactively explore their multi-dimensional and multi-attribute datasets using a series of multi-coordinated linked views.

Researchers have also explored different methods to address the challenges posed to maritime security and safety. Willems et. al. [34] introduce a novel geographic visualization that supports coastal surveillance systems and decision making analysts in gaining insights into vessel movements. They utilize density estimated heatmaps to reveal finer details and anomalies in vessel movements. Scheepens et. al. [27] also present methods to explore multivariate trajectories with density maps and allow the exploration of anomalously behaving vessels. Lane et. al. [18] present techniques that allow analysts to discover potential risks and threats to maritime safety by analyzing the behavior of vessel movements and determining the probability that they are anomalous. Some other models for anomaly detection in sea traffic can be found in [19, 22]. Researchers have also proposed several approaches to maritime domain awareness. For example, Roy and Davenport [23] present a knowledge based categorization of maritime anomalies built on a taxonomy of maritime situational facts involved in maritime anomaly detection. We observe that these methods and models may help in risk analysis and understanding the impact of weather and varying speeds of Coast Guard vessels in the Great Lakes to identify high risk regions.

There has also been much work done to assess and mitigate risks to critical infrastructure and transportation in the maritime domain. Adler and Fuller [5] provide dynamic scenario- and simulation-based risk management models to assess risks to critical maritime infrastructure and strategies implemented for mitigating these risks. Mansouri et. al. [20] also propose a risk management-based decision analysis framework that enables decision makers to identify, analyze, and prioritize risks involved in maritime infrastructure and transportation systems. Their framework is based on risk analysis and management methodologies that allows understanding uncertainty and enables analysts to devise strategies to identify the vulnerabilities of the system. Furthermore, work has been done to quantify risks in the maritime transportation domain, a summary of which can be found in [29]. While these methods facilitate maritime infrastructure risk analysis, our work is focused on assessing maritime risks from multivariate spatiotemporal SAR data sets. In this paper, we present a visual analytics approach to maritime risk assessment and provide examples that demonstrate the advantages of applying visual analytics in this domain.

3 Visual Analytics Risk Assessment Environment

Our visual analytics system provides enhanced risk assessment and analytical tools to analysts and has been built to operate for SAR incident report data. Our system has been implemented in a custom Windows-based geographical information system that allows drawing on an OpenStreetMap map [4], using Visual C++, MySQL and OpenGL. The system displays geo-referenced data on a map and allows users to temporally scroll through their data. We provide linked windows that facilitate user interaction between the spatial and temporal domains of the data. We also provide advanced filtering techniques that allow users to interactively explore through data. In addition, we have adapted the calendar view presented by vanWijk and Selow [33] and extended it to explore seasonal and cyclical trends of SAR operations and also as means to filter data to support advanced analysis.

Figure 1 presents a screenshot of our system. The main viewing window (Figure 1 (a)) shows the map view where the user can explore the spatial distribution of all cases handled by the Coast Guard. We utilize density estimated heatmaps (Section 3.2) to quickly identify hotspots. Users may draw a bounding box over incident points on the map that generates a summary of all incidents enclosed by the box. We also provide tape measure tools that allow users to measure the distance between two points on a map. The top-most window (Figure 1 (c)) shows the time-series view of the data where multiple lines graphs can be overlaid for comparison and analysis. Users may visualize time-series plots by department, distress type and Coast Guard Captain of the Port (COTP) zone to explore summer cyclical patterns. The left-most window (Figure 1 (d)) shows the calendar view of the selected Coast Guard cases. The total number of columns on the calendar may be changed as desired to reveal seasonal trends and patterns. The bottom window (Figure 1 (e)) shows the time-slider widget that is used to temporally scroll through the data while dynamically updating all other linked windows. The radio buttons beneath the time slider provide several temporal aggregation methods for the data. The right-most window (Figure 1 (b)) shows the distress type menu where all SAR cases (highlighted in blue) have been selected for visualization. Users may select multiple distress types using this menu, dynamically updating all linked views. We use similar menus to filter cases by other data fields. Users may also interactively search the menu using the search box provided on top of the menu. Finally, the top-right window (Figure 1 (f)) shows an interactive legend of the different Coast Guard District Nine maritime zones. This legend allows users to click on any of the zones that highlights all cases falling in the zone by filling the circles on the map with a solid color and dimming out the other cases being displayed on the map.

A key feature of our system is the interactive distress, station and COTP zone filtering component. Users interactively generate combinations of filters that are applied to the data being visualized.
through the use of menus (like the one shown in Figure 1 (b)) and edit controls. The choices of filters applied affects both the geospatial viewing region and all temporal plots.

3.1 Coast Guard SAR data

The SAR data is collected by all U.S. Coast Guard stations and stored in a central repository. When the Coast Guard is called into action, a response case is generated, usually by the maritime zone that has authority in that region that receives the distress call (referred to by the Coast Guard as the Search Mission Coordinator or SMC). Upon receiving the call, this authority will determine if resources will be applied, including which unit will provide the resource, the resource type and number. Therefore, a response case may generate zero, one, or many sorties to respond to an incident. While analyzing risks associated with the various mitigation options, including station closure, analysts are interested in analyzing the spatiotemporal distribution of all the response cases and their associated sorties.

The SAR data consists of two main components: (1) response cases and (2) response sorties. Each entry in the response case and sortie dataset contains information that provides details of the incidents (e.g., number of lives saved, lost, assisted) and contains the geographic location of the distress.

Uncertainty in decision making

As is the case with most large datasets, anomalies and missing data introduce errors and uncertainty. The SAR data is no exception. We find that many SAR cases do not have an associated geographic location, or have a wrong geographic location associated with them. These inherent errors in data affect the spatial probability estimates and introduce a certain amount of uncertainty in the decisions that must be considered for an effective risk analysis and assessment. As noted in [17], visual analytics methods help people make informed decisions only if they are made aware of data quality problems. In this regard, we incorporate uncertainty and confidence levels associated with the SAR dataset in our visualizations by displaying the accuracy of the results at each step of the risk assessment process. This is shown as a percentage that shows the total cases with reliable data that can be used in the decision making process (Figure 2). This percentage is calculated by using the following formula:

\[
\text{Accuracy} = \frac{N - G}{N} \times 100 \quad (1)
\]

Here, \(N\) is the total number of cases and \(G\) is the number of cases with unreliable values (e.g., unknown geographic coordinates, swapped negative signs). When such errors are not obvious, the data is assumed to be correct and is displayed to the analyst on the map. The analyst can further report errors in the data and contribute to the data cleaning process.

3.2 Geospatial displays

Our system provides analysts with the ability to plot incidents as points on the map and as density estimated heatmaps (Figure 1 (a)). In addition, we provide users with the option of coloring each incident circle with a color on a sequential color scale [6] that represents its data value. For example, users may choose to visualize the average response time to respond to an incident for all SAR cases on the map and identify cases with higher response times. Furthermore, to explore the spatial distribution of the SAR cases and quickly identify hotspots, we employ a modified variable kernel density estimation technique (Equation 2) that scales the parameter of estimation by allowing the kernel scale to vary based upon the distance from the point \(X_i\) to the \(k\)th nearest neighbor \(x\) in the set comprising of \(N\) [28].

\[
\hat{f}(x) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\min(h,d_{ik})} K \left( \frac{x - X_i}{\min(h,d_{ik})} \right) \quad (2)
\]

Here, \(N\) is the total number of samples, \(d_{ik}\) is the distance from the \(i\)th sample to the \(k\)th nearest neighbor and \(h\) is the maximum allowed kernel width. We choose the maximum kernel width based on asset speed and travel time. Furthermore, we use the Epanechnikov kernel [28] (equation 3) to reduce calculation time:

\[
K(u) = \frac{3}{4} (1 - u^2) 1_{|u| \leq 1} \quad (3)
\]

where the function \(1_{|u| \leq 1}\) evaluates to 1 if the inequality is true and to zero otherwise.

3.3 Time series displays

Along with the graphical interface, our system provides a variety of visualization features for both spatial and temporal views. For temporal views, we provide line and stacked bar graphs and calendar views to visualize time series SAR incident report data.

The line graph visualization allows users to overlay multiple graphs for easy comparison and to visualize trends. Both line graph and stacked bar graph visualizations are supported and can be interchanged using the radio buttons provided. Users may choose to visualize SAR cases handled by individual stations or maritime zones, or visualize them by distress types. The data is plotted based on a temporal aggregation level that the user selects on the time-slider widget (Figure 1 (e)). In Figure 1 (c), we show the line graph display of all SAR cases aggregated by month. We can easily observe peaks in the number of SAR cases in the summer months for all maritime zones in the Great Lakes region.

The calendar view visualization was first developed by van Wijk and Selow [33]. This visualization provides a means to allow the visualization of data over time, laid in the format of a calendar. In our implementation (Figure 1 (d)), we shade each date entry based on the overall yearly trend. Users may interactively change the temporal number of columns of the calendar thereby changing the cycle length of the calendar view, enabling users to explore both seasonal and cyclical trends of their datasets. The system also draws histograms for each row and column. This allows analysts to visualize weekday and weekly trends of SAR incidents and further assists them in determining an effective resource allocation scheme. Furthermore, we have modified our calendar view to support an interactive database querying method for easily acquiring summary statistics from the SAR database.

4 Risk Assessment process

In this section, we describe the different methods and techniques that we apply in the Coast Guard risk assessment process.

4.1 Problem description

To bound the problem, the Coast Guard analysts provided a series of questions for use in their analysis. These questions are briefly summarized below.

1. Assuming a maximum transit speed of 15 nautical miles per hour, how many cases occur per year in which a parent station could not have a surface asset on scene within two hours?
2. For each Auxiliary station, what are the types (by percentage) of SAR response cases occurring per year?
3. For each Auxiliary station, what is the temporal (by hour, month and day of week) distribution of the response case load?
4. What is the average annual case load that would be absorbed by each parent station in the absence of the Auxiliary station and what percentage increase would this represent to the parent station’s annual case load?

5. Based on the historical data for all cases (SAR and others), what is the expected annual response case demand broken down by response type (i.e., Person in Water, Vessel Flooding, etc.)?

6. Assess the potential risks associated with closing certain Auxiliary stations in terms of additional case load absorbed, lives potentially lost, and other available factors.

Our visual analytics system was developed to assist the Coast Guard analysts in answering these questions and to model the potential risks of closing one or more Auxiliary stations. Furthermore, we allow analysts to explore the effects of closing multiple stations and provide a summary of stations that are most optimal to absorb the work load of the closed stations. Analysts may restrict the stations that absorb the work load of the closed stations to determine the stations that prove most effective, thereby informing optimal operational execution for the station that is nearest to respond to the distress case.

We perform our analysis under the assumption that the path between a station and a distress location is a straight line. While this assumption presents a best-case scenario to the analyst, discussions with our Coast Guard partners indicated this was an acceptable approximation as using channel and waterway information would result in a large computational overhead. With this assumption in place, if a station absorbing an Auxiliary station’s cases increases the maritime risks in the region (e.g., if the average response time exceeds the two hour time limit for most SAR incidents), then closing the Auxiliary station could prove to be dangerous for the maritime and public safety of the region. This straight line approximation provides details on the best case scenario.

4.2 Average response time for SAR incidents

As stated before, a Coast Guard policy mandates the rescue resource to be on scene within two hours of a distress (e.g., disabled vessel, person in water). Given the cold water temperatures in the Great Lakes, even in the summer, increase in response time can potentially impact the success of a case. Therefore, given the option of closing a station, the analysts desire to know the nearest available resource to respond and calculate the time to respond to the scene. A typical Coast Guard vessel travels at a speed of 15 nautical miles
per hour. After an Auxiliary station is closed, the parent station should still be able to reach most of the cases handled by the Auxiliary station within the two hour limit. In this section, we describe how our system can be used to determine the average response time for cases if a parent station (or any combination of stations) absorbs an Auxiliary station’s cases.

In order to generate the average response time for the station(s) that absorb the work load of the closed station, we sift through all the incidents that the closed station handled and find the closest station (excluding the closed station) for each incident by comparing the distance between all stations and the incident. This distance between the closest station and incidents may also be visualized separately to reveal more details. Once the closest station is found, we obtain the time for an asset to reach the incident location using the distance formula \( \text{Time} = \frac{\text{Distance}}{\text{Speed}} \). Users may also change the speed of the asset, changing the results dynamically.

We provide users with several filtering options while performing average response-time analysis. Users may choose to analyze the average response-time temporal distribution of incidents by applying any possible filters on distress type, department or maritime zone. Users may also analyze the distribution of only the non-SAR cases. Moreover, users may choose to close several stations all at once and model the resulting effects. They may also specify which stations absorb the cases of the closed stations and thus determine the stations best suited for closing and the optimal methods for re-allocating available resources. We also note that our system can be easily modified to incorporate other risk metrics including, for example, normalizing SAR cases by the underlying population density, correlating SAR incidents with other parameters, etc.

Figure 2 shows the output generated when the analyst opts to close Auxiliary station Y, shown in Figure 2. In this example, we examine all cases responded to by station Y between January 2004 and September 2010. The system automatically suggests the stations that should absorb Auxiliary station Y’s cases along with the total number of cases that each station absorbs. We find that stations C (the parent station of Y), D and E absorb Auxiliary station Y’s cases, with each absorbing 84, 2 and 1 cases, respectively. The analyst may instead select a specific station to absorb station Y’s cases and analyze the results generated. In Figure 2, the map view shows all cases that each of the four stations responds to during this time period (shown as circles, with each case color coded by its station). We have also highlighted the two cases that station D and the one case that station E responds to in Figure 2. It may be noted that the one case absorbed by station E appears to be out of place (possibly due to a human error in entering the geographic coordinates for that particular case). The top-right bar graph shows the count of all SAR cases handled by station Y during this time period versus the average response time (in minutes) taken by the resulting stations to reach these cases, assuming a transit speed of 15 nautical miles per hour. From this time-series plot, we observe that all cases responded to by the Auxiliary station would fall well within the tolerance level of 120 minutes when the suggested stations take over. The system also determines the accuracy of the results dynamically by determining the number of cases that have no associated geographic coordinates. We find that 93% of the cases responded to by station Y in the time range January 2004 and September 2010 have an associated geographic coordinate (as seen from the accuracy percentage in Figure 2-top-right). Data integrity is a necessary parameter to report to the analysts and decision makers. Thus, the user is made aware of these uncertainties at every step of the risk assessment process.

4.3 Temporal distribution of response case load

One important aspect of risk assessment is analyzing the work load and distribution of response cases of the stations being analyzed over different temporal ranges. This becomes necessary to determine the feasibility of a station to be closed and to determine how the available resources may be reallocated (e.g., what times of day and what months would the stations need to have more personnel deployed). Analysts also use their domain experience and expertise to determine whether a particular station can absorb a closing station’s cases. In particular, the Coast Guard officials were interested in understanding the hourly, daily and monthly trends of SAR cases occurring in the Great Lakes.

Using traditional methods of sifting through SAR datasets turns out to be highly inefficient for determining the temporal distribution of the SAR cases and, as such, advanced database querying tools are necessary to facilitate this process. To this end, we adapt the calendar view for querying the SAR database. We provide three different interaction methods within the calendar view widget (Figure 1 (d)) to obtain a detailed summary of response cases occurring over the selected date-range. Users can select date ranges by simply clicking on the start and end dates that selects all the dates between the two clicked dates. Users may also select one or more columns...
of the calendar to generate the summary statistics. This allows them to query the database and acquire the summary of events occurring, for example, only on a particular week day. Finally, users may select any combination of individual dates and obtain the summary of all selected response cases on those dates. These querying methods allow analysts to easily determine the temporal patterns of response cases over any date range. The system provides summary statistics of SAR incidents for all stations and includes the total number of lives saved, assisted, affected, total property damaged and saved and the count of all cases occurring over the selected date range. Users may select any date, row, column, or combinations thereof in the calendar view using the mouse to access the summary statistics. Furthermore, the system also allows users to visualize the hourly and monthly distribution of cases for any time period after all filters are applied.

4.4 Risk profile

Our system also provides users with the ability to interactively generate risk profiles that can be used to identify regions with little SAR coverage by the Coast Guard stations in the Great Lakes. Figure 3 illustrates the risk profile heatmaps that present an overview of the Coast Guard SAR coverage in the Great Lakes. Selected filter settings affect the visual output, and in this case, we are looking exclusively at small boat station coverage. When areas of low coverage exist, resources with additional capability (e.g., aircraft) are often provided to ensure coverage of all areas. Figure 3 (Left) shows the time (in minutes) that the Coast Guard stations would take to respond to a SAR incident in the Great Lakes, assuming a transit speed of 15 nautical miles per hour. This profile is generated assuming that the station closest to a location responds to an incident in the Great Lakes. The regions in the Great Lakes that take the longest time for the Coast Guard to respond to a SAR case can be clearly seen in this figure. Users may interactively close stations, filter on a different resource type (e.g., boat, aircraft), or change the asset speed, updating the risk profile interactively. This further enables the analysts to visualize the increase or decrease in risk when a station is closed. Moreover, analysts can set the lower threshold of the color scale to 120 minutes (or any arbitrary time), thereby allowing them to easily identify regions that may take more than 120 minutes to respond. We plan on incorporating contour lines into our system to demarcate the regions that may take more than the set threshold response time.

Figure 3 (Right) provides another risk profile visualization that allows officials to identify regions with low Coast Guard coverage for SAR operations in the Great Lakes. Regions with high SAR coverage by the Coast Guard stations are shown by darker colors. This further allows analysts to identify stations where resources may be reallocated without increasing maritime risk.

5 Exploring risk using spatiotemporal linked views

While examining which Auxiliary stations are most suitable to close, analysts need to weigh all options and analyze the potential increase or decrease in associated risks. They must also consider
the increase in workload of the stations that absorb the closed station's cases to effectively determine the optimal response of available resources. In this section, we describe a typical scenario where an analyst is trying to determine the risks associated with closing an Auxiliary station in one of the sectors in the Great Lakes and the stations that absorb the work load of this closed station.

Suppose the analyst chooses to close Auxiliary station X whose parent station is A. Since the parent station of X is station A, the analysts' first inclination may be to assign all cases to station A after station X is closed. The analyst first uses the system's calendar view and finds that the maximum number of cases that station X responds to in one day is 7 in the peak boating season. He also visualizes the hourly distribution of cases for station X and determines that the incidents are spread out during the day. Next, the analyst uses our risk assessment system to perform an average response time risk analysis over a time range of January 2006 to September 2010 and selects Auxiliary station X to be closed. Once station X is closed, the system automatically generates the result seen in Figure 4. As seen in the figure, the system determines stations A and B to be the optimal stations to respond to the cases handled by Auxiliary station X. In this figure, we see the spatial distribution of all SAR cases that the three stations responded to during this time range (seen as circles that are color coded by station). We observe that station B absorbs 144 additional SAR cases as opposed to parent station A which only absorbs 15 cases. Moreover, with this case load distribution, we find that 154 out of the total 159 cases are responded to within the 120 minute limit (as seen from the time series plot in Figure 4-top-right). These results suggest that station B is better suited to absorb most of the cases of Auxiliary station X.

In order to get a better picture, the analyst now restricts the stations that respond to the cases handled by station X to first, its parent station A, and then to station B and analyzes the average response time distribution of cases for each of the two stations separately. The results of this step are shown in Figure 5, with the left graph showing the average response time taken by station A, and the right graph corresponding to station B. As the analyst compares the two graphs, he realizes that if only station B is allowed to absorb station X, there are 14 cases that take more than 2 hours for the Coast Guard to arrive on scene. On the other hand, if station A is allowed to absorb station X, 16 cases take longer than 2 hours. However, as we can clearly see from Figure 5, station A takes a longer time to respond to most cases than station B, with the median time of station A being 110 minutes, and that of station B being 92 minutes. The analyst also notes that station B takes between 264-271 minutes to respond to about nine cases, whereas station A takes 275 minutes to respond to one case. To get a better understanding of why this may be happening, the analyst explores the spatial distribution of station X's cases on the map and discovers that some cases of station X get mapped to an inland lake (which requires trailering the boat to the scene). In order to confirm that these cases do not occur as a result of errors in the database, he draws a bounding box over these incidents on the map and obtains a summary of these incidents. The summary confirms that these incidents do indeed occur in that particular lake. As station A is closer to these cases, the analyst concludes that station A would be a better candidate to absorb these cases as opposed to station B. But for the rest of the cases, the results clearly suggest that station B would be a better candidate to absorb station X's cases, and that station A would increase the maritime risks if allowed to absorb station X's cases alone. Thus, this analysis confirms that a combination of these two stations yields the best results. With these results at hand, the analyst may also recommend using an aircraft to respond to cases that take more than 120 minutes. Or, as our preceding analysis showed, stations A and B may absorb the work load of station X together, with station B receiving a higher share of resources than station A. We also note that in the future, our system could be modified to perform a real time analysis of SAR cases and could then be used to assign each case to the correct station in real time.

The analyst now uses the risk profile tools to observe the increase in risks when station X is closed. This is shown in Figure 6, with the left figure showing the average response time risk profile when Auxiliary station X is functional whereas the right figure shows the risk profile when station X is closed. The analyst explores the new average response time on the map when station X is closed and determines the potential increase in maritime risks in the region. The
The analyst thus identifies the regions in station X’s area of operation that take a greater time to respond. The analyst also visualizes the monthly and hourly distributions of all SAR cases responded to by the closed Auxiliary station X between January 2006 and September 2010 (shown in Figure 7, where the left graph shows the monthly distribution and the right graph shows the hourly distribution). We see that all activity occurs in the summer months with a peak occurring in July and the station responds to most cases mainly during evening hours. The analyst also visualizes the temporal activity of stations A and B, and determines the potential work load increase for both stations. This helps the analyst determine how the available resources must be reallocated if Auxiliary station X is to be closed. Furthermore, the analyst chooses to visualize the case distribution of Auxiliary station X between January 2006 and September 2010 using the interactive calendar view widget. This generates a summary dialog box that provides the details of the SAR cases responded to by station X and includes details including the total number of lives assisted, saved and lost. This further helps the analyst understand the risks associated with closing this station by providing an overview of the cases occurring in this region. With all these results at hand, the analyst uses his domain knowledge to make an informed decision.

6 Domain expert feedback

Our system was assessed by an analyst at the U.S. Coast Guard’s Ninth District who is currently using the system to determine the potential risks in the maritime domain associated with the hypothetical allocation of Coast Guard resources. The analyst emphasized the need of such systems in the maritime domain that allow users to quickly and easily process large datasets in order to derive actionable results. The analyst noted that processing the desired queries took him a fraction of the time when using our system as compared to using other software (e.g., Microsoft Excel) that he had been previously using in his analysis. He was impressed by the fact that the system is intuitive to use and requires little user training. He observed that the system’s ability to process large datasets allows him to quickly filter the data into manageable subsets while providing interactive spatiotemporal displays that further aid him (and ultimately the senior level decision makers) in making a decision using the best information available.

7 Conclusions and future work

Our current work demonstrates the benefits of visual analytics in analyzing risk and historic resource allocation in the maritime domain. Our visual analytics system provides analysts with a suite of tools for analyzing risks and consequences of taking major decisions that translate into important measures including potential lives and property lost. Our results show how our system can be used as an effective risk assessment tool when examining various mitigation strategies to a known or emergent problem.

Before this system was developed, Coast Guard officials explored possible mitigation strategies, including the implementation of seasonal or weekend only Auxiliary duty stations, but the sheer volume of data and information inhibited the efficient processing of the data. However, using our system, the decision makers were quickly made aware that most response cases happened on Mondays/Tuesdays at some of the units. This further asserts the benefits of the use of visual analytics in the maritime domain.

In addition to performing risk analysis on the Coast Guard SAR cases, our system can also be used to conduct a thorough review of the operations (i.e. non-distress cases) conducted by different Coast Guard stations. Users may choose to visualize different datasets and analyze how each station performs in terms of factors including average response times, average distance to target, lives saved, lives assisted, lives affected, etc. Hence, the officials may analyze the efficiency of each Coast Guard station and identify problem areas that may require further attention.

Future work includes deploying our system to assist in the analysis and optimization of all operations conducted by the U.S. Coast Guard Ninth District and expanding the use of our system to other Coast Guard districts. We plan on implementing algorithms that factor the geography of the coast line in the risk assessment process in order to get accurate response times by the Coast Guard assets. We also plan on employing prediction algorithms in the temporal domain as well as spatiotemporal correlation algorithms that correlate different datasets (e.g., weather, water temperature) with the response dataset to provide insights into the operation of the Coast Guard stations. Furthermore, we plan on incorporating additional risk metrics to provide insights into different risk scenarios.

8 Appendix

In this section, we briefly provide some domain specific terms and definitions:

- **Coast Guard Auxiliary**: Volunteers that support the Coast Guard.
- **Coast Guard Ninth District**: The area of Coast Guard operations that encompasses the Great Lakes.
- **Atlantic Area Command**: The area of Coast Guard operations East of the Rocky Mountains.
- **Captain of the Port (COTP) Zone**: Further division of Coast Guard operations within a Coast Guard District.
- **Unit or Station**: The operational execution arm of the Coast Guard. For example, the small boat station provides the boat and personnel to execute the assigned mission.
- **Coast Guard asset**: A boat or an aircraft reserved to perform Coast Guard operations.
- **Coast Guard sortie**: An asset that responds to an incident.

9 Acknowledgments

The authors would like to thank Capt. Eric Vogelbacher, Steffen Koch and Zichang Liu for their feedback. This work is supported by the U.S. Department of Homeland Security’s VACCINE Center under Award Number 2009-ST-061-CX0002.

10 Disclaimer

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security or the U.S. Coast Guard.
REFERENCES


