A Visual Analytics Process for Maritime Response, Resource Allocation and Risk Assessment

Abish Malik†
Purdue University, USA
Yang Yang‡
Purdue University, USA
Ross Maciejewski§
Arizona State University, USA
Ben Maule ¶
United States Coast Guard
Yun Jang†
Sejong University, South Korea
Matthew White§
United States Coast Guard
Silvia Oliveros†
Purdue University, USA
David S. Ebert
Purdue University, USA

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 lost 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, resource allocation, 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 [18, 37] 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 [39]. Closing these stations implies a potential for longer response times that could potentially translate into the loss of lives and property.

To address these option evaluation 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
Figure 1: A screenshot of our risk assessment visual analytics system. Here, the user is visualizing all search and rescue (SAR) operations conducted by the U.S. Coast Guard in the Great Lakes region. The main viewing area (a) is a map view with the circles showing the locations of SAR incidents in the Great Lakes that occurred in July 2008. The right window (b) shows an interactive selection menu for distress types with SAR cases selected in blue. The top window (c) is a time-series view of the SAR incident report data showing a monthly distribution of SAR incidents over 5 fiscal years along with a legend of all District Nine maritime zones (d). The left window (e) shows a calendar view of the SAR incidents for the year 2008 with the month of July highlighted. An hourly view of the incidents responded by all stations (to all SAR incidents occurring in July 2008) is shown by the clock view display (f). Finally, the bottom-left window (g) shows the time slider with radio buttons that allow different temporal aggregation levels with the current time frame selected to July 2008.

- Parameterization of Coast Guard assets for use in risk and resource distribution analysis
- 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 scenarios 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., [36, 41]). From traditional scatterplots [9] and parallel coordinate plots [17] to tools like Theme River [16] and spiral graphs [8], 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 Worringer [24] 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 [14] 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 [30] 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 [13, 14]. 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 with investments and mutual funds, among other financial planning scenarios. Rudolph et al. [29] 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. [31] 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 correct conclusions, analysts must take these into account in their analysis. In this regard, Correa et al. [10] 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, 15]). As the needs of our end users are unique, this warrants developing a standalone 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. [36] use multiple coordinated views of documents to reveal connections between entities across different documents. Eick and Johnson [11] 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 [12] 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. [41] 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. [32] also present methods to explore multivariate trajectories with density maps and allow the exploration of anomalously behaving vessels. Lane et. al. [19] 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 [20, 27]. Researchers have also proposed several approaches to maritime domain awareness. For example, Roy and Davenport [28] 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.

In the maritime security domain, Orosz et al. [25] developed a decision making and planning tool that addresses the security needs based on the port resources. Their focus is on tactical and strategic tradeoffs that arise due to the different port resource reallocation strategies. They provide support to visualize the effects of the different day by day resource allocation strategies and also allow users to forecast the risks and impacts of expansion and changes made by the port security officials. Their definition of risk is based on three components: threat, vulnerability, and consequences. Their system provides users with the ability to model and visualize the risks in specified regions. The system responds mostly to outside threats and hypothetical scenarios whereas our system is focused on the historic spatiotemporal distribution of SAR incidents. We note that the authors provide an interesting research direction in the analysis of resource allocation strategies, and we plan on exploring these further in the future. Ulusu et al. [38] also model the security risks due to maritime traffic accidents. In their approach, they divide the waterways into regions and compute the risks based on potential accidents occurring within these regions. They utilize 3D graphs to visualize the regions and their corresponding risks.

There exist several geographic information systems (GIS) that allow users to visualize the datasets associated with coastal areas. There also exist several GIS systems that incorporate different decision making processes by visualizing geospatial data in their specified areas. Pelot and Plummer [26] focus on traffic modeling for both small and large commercial vessels, and use grid maps that color the regions based on the traffic density falling in the regions, thereby allowing the visualization of vulnerable spatial regions. Marven et al. [23] analyze Canadian Coast Guard Search and Rescue operations in the Pacific coast, and follow a Spatial and Temporal Analysis of Crime (STAC) model that is based on a clustering algorithm which identifies regions of high spatial incident densities. The authors explore nearest neighbor clustering and kernel density estimation (KDE) techniques in their work. Abu-Zeid et al. [5] also developed a geographical decision support system for planning aeronautical search and rescue missions that provides support for resource allocation decisions. Our custom built GIS system also builds upon the concept of providing analysts with the ability to visualize and analyze their multivariate spatiotemporal datasets in order for them to make informed decisions.

There has also been much work done to assess and mitigate risks to critical infrastructure and transportation in the maritime domain. Adler and Fuller [6] 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. [22] 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 [35]. 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 van Wijk and Selow [40] 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 (e)) 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-right window (Figure 1 (f)) shows a clock view that allows users to visualize the hourly distributions of SAR incidents. The bottom-left window (Figure 1 (g)) 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 center-right 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 (d)) 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 (e.g., boat, aircraft) 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 both 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 [18], 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:

\[
\text{Accuracy} = \frac{N - G}{N} \times 100
\]  

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 [7] 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 \( N \) [34].

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

Here, \( N \) is the total number of samples, \( d_{i,k} \) 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 [34] (equation 3) to reduce calculation time:

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

where the function \( 1_{\{|u| \leq 1\}} \) evaluates to 1 if the inequality is true and to 0 otherwise.
Figure 2: Average response time risk assessment when Auxiliary station Y is closed. The system automatically chooses the stations (shown in the upper-left window) that are optimally suited to respond to cases previously handled by station Y, along with a count of cases that each station absorbs. The station Y cases (black circles) to be handled by stations D and E are circled in red. The top graph shows the average response time distribution of these stations to respond to station Y’s cases. The decision makers are provided with accuracy estimates, and in this case, 93% of the incidents are displayed (7% may not have valid GIS coordinates).

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, calendar, and clock views to visualize time series SAR incident report data.

The line graph visualization allow 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 (g)). 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 [40]. 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 (e)), we shade each date entry based on the overall yearly trend. Users may interactively change the total 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.

Finally, the clock view visualization allows users to visualize the hourly distribution of the selected SAR incidents responded to by the Coast Guard stations. Our implementation of the clock view organizes the data in the form of a clock, which is a radial layout that is divided into slices to reflect the number of incidents that occur
during each hour of the clock. Each slice of the clock view is colored based on a sequential color scheme to reflect the corresponding number of incidents occurring during that hour. Furthermore, we allow users to visualize the hourly distribution of either all selected stations in a combined clock view display (Figure 1(f)), or of individual stations that are drawn as individual clock views over the stations on the map (Figure 3). We note that both these clock view representations are dynamically linked to the time slider (Figure 1(g)), and are dynamically updated as the user interacts with the system. Moreover, there also occur many SAR cases that do not have an associated hourly time component. In order to depict these inherent errors in the data, we utilize Equation 1 to display the percentage of cases with valid hourly components to the users.

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 Time = Distance / 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 reallocation of 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. 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
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 (e)) 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 4 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 4 (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 4 (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.

4.5 Dynamic assignment of station asset speeds

In order to assist analysts with the different resource allocation strategies, we note that not all stations are allocated resources uniformly. Stations that respond to a high number of incidents per year are typically assigned more assets and may be assigned assets with higher speeds and performance. This necessitates analysts to factor in the station asset allocation strategy in addition to the station case load when they determine the stations that would be viable to absorb the incidents of hypothetically closed station(s). In order to facilitate this process, our system allows users to dynamically assign asset speeds to different Coast Guard stations. The system then factors these asset speed assignments into the analysis process and displays the results to the users that is reflected in both the average response time analysis (section 4.2) and the risk profile heatmaps (section 4.4). An example of the user changing the asset speed for a station is shown in Figure 5, where the user has changed the asset speed changed to 30 nautical miles per hour for station J.

The resulting heatmap clearly shows that the regions that took more than 120 minutes to respond. We plan on incorporating contour lines showing the asset speed changed to 30 nautical miles per hour for station J.

4.5 Dynamic assignment of station asset speeds

In order to assist analysts with the different resource allocation strategies, we note that not all stations are allocated resources uniformly. Stations that respond to a high number of incidents per year are typically assigned more assets and may be assigned assets with higher speeds and performance. This necessitates analysts to factor in the station asset allocation strategy in addition to the station case load when they determine the stations that would be viable to absorb the incidents of hypothetically closed station(s). In order to facilitate this process, our system allows users to dynamically assign asset speeds to different Coast Guard stations. The system then factors these asset speed assignments into the analysis process and displays the results to the users that is reflected in both the average response time analysis (section 4.2) and the risk profile heatmaps (section 4.4). An example of the user changing the asset speed for a station is shown in Figure 5, where the user has changed the asset speed changed to 30 nautical miles per hour for station J.

The resulting heatmap clearly shows that the regions that took more than 120 minutes to respond. We plan on incorporating contour lines showing the asset speed changed to 30 nautical miles per hour for station J.

4.5 Dynamic assignment of station asset speeds

In order to assist analysts with the different resource allocation strategies, we note that not all stations are allocated resources uniformly. Stations that respond to a high number of incidents per year are typically assigned more assets and may be assigned assets with higher speeds and performance. This necessitates analysts to factor in the station asset allocation strategy in addition to the station case load when they determine the stations that would be viable to absorb the incidents of hypothetically closed station(s). In order to facilitate this process, our system allows users to dynamically assign asset speeds to different Coast Guard stations. The system then factors these asset speed assignments into the analysis process and displays the results to the users that is reflected in both the average response time analysis (section 4.2) and the risk profile heatmaps (section 4.4). An example of the user changing the asset speed for a station is shown in Figure 5, where the user has changed the asset speed changed to 30 nautical miles per hour for station J.

The resulting heatmap clearly shows that the regions that took more than 120 minutes to respond. We plan on incorporating contour lines showing the asset speed changed to 30 nautical miles per hour for station J.

4.5 Dynamic assignment of station asset speeds

In order to assist analysts with the different resource allocation strategies, we note that not all stations are allocated resources uniformly. Stations that respond to a high number of incidents per year are typically assigned more assets and may be assigned assets with higher speeds and performance. This necessitates analysts to factor in the station asset allocation strategy in addition to the station case load when they determine the stations that would be viable to absorb the incidents of hypothetically closed station(s). In order to facilitate this process, our system allows users to dynamically assign asset speeds to different Coast Guard stations. The system then factors these asset speed assignments into the analysis process and displays the results to the users that is reflected in both the average response time analysis (section 4.2) and the risk profile heatmaps (section 4.4). An example of the user changing the asset speed for a station is shown in Figure 5, where the user has changed the asset speed changed to 30 nautical miles per hour for station J.

The resulting heatmap clearly shows that the regions that took more than 120 minutes to respond. We plan on incorporating contour lines showing the asset speed changed to 30 nautical miles per hour for station J.

4.5 Dynamic assignment of station asset speeds

In order to assist analysts with the different resource allocation strategies, we note that not all stations are allocated resources uniformly. Stations that respond to a high number of incidents per year are typically assigned more assets and may be assigned assets with higher speeds and performance. This necessitates analysts to factor in the station asset allocation strategy in addition to the station case load when they determine the stations that would be viable to absorb the incidents of hypothetically closed station(s). In order to facilitate this process, our system allows users to dynamically assign asset speeds to different Coast Guard stations. The system then factors these asset speed assignments into the analysis process and displays the results to the users that is reflected in both the average response time analysis (section 4.2) and the risk profile heatmaps (section 4.4). An example of the user changing the asset speed for a station is shown in Figure 5, where the user has changed the asset speed changed to 30 nautical miles per hour for station J.

The resulting heatmap clearly shows that the regions that took more than 120 minutes to respond. We plan on incorporating contour lines showing the asset speed changed to 30 nautical miles per hour for station J.

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 scenario in which a group of analysts are trying to quantify and assess the risks and effects of closing multiple stations in the Great Lakes. An example that describes a complete analysis of a hypothetical single station closure scenario can be found in our earlier paper [21].

When analyzing the risks associated with station closures, analysts often have to consider several resource allocation scenarios and analyze all possible mitigation strategies in order to reach solutions that minimize the overall risk. This type of analysis requires them to focus on all aspects of possible strategies and extensively use their domain knowledge to reach to these solutions. Moreover, analysts are required to consider the different types of resources available at each station and the possible resource redistribution strategies once some stations are closed. Finding optimal solutions to these challenges often proves to be extremely laborious, and as such, requires advanced techniques to aid analysts in this process. We also note that a domain expert who can assess all these different possibilities is an integral component in this visual analytics process.

In order to assist analysts with these challenges, we utilize several techniques in our visual analytics framework. Here, we describe a hypothetical scenario in which a group of analysts are assessing the risks associated with closing two stations at the same time in the Great Lakes (stations M and N in Figure 6), and are analyzing different resource allocation strategies among the stations that absorb the closing stations’ incidents. We note that this type of analysis can be extended to the closures of more than two stations at the same time. In this analysis, we assume that stations M and N have two types of assets that must be redistributed after they are closed. These assets include two types of boats, the first set have typical boat speeds of 15 nautical miles per hour, and the second set are the hypothetical new generation boats with speeds of 30 nautical miles per hour. To simplify this scenario, we assume that there are equal numbers of 15 and 30 nautical miles per hour speed boats at stations M and N. We also assume that the analysts have a budget of purchasing a limited number of new generation speed boats and distributing them among the stations that absorb the work load of the closing stations.

In order to begin the analysis process, the analysts use the system and select the two stations for closure using the interactive menu. The analysts first use the system’s calendar view to determine the maximum number of cases that the two stations respond to in a sin-
Figure 7: Average response time analyses after stations M and N are closed and the resources are reallocated among stations F, G and H.

Figure 6: Spatial distribution of cases responded to by stations M and N in fiscal years 2007 to 2011. The incidents are represented on the map as circles and are color coded by their owner station.

gle day. They find that the maximum number of cases responded to by stations M and N combined are 9, occurring on the Independence Day weekend of the year 2007 on 7/8/2007. They now visualize the hourly distribution of these 9 incidents using the clock view in order to determine an optimal resource allocation strategy, and determine that only 7 of the 9 cases have an associated time component, and that these 7 incidents are uniformly spread out over a 24 hour period. This gives them an indication of a worst case daily resource distribution scenario when stations M and N are closed.

Restricting analysis to stations F, G and H

The analysts now perform an average response time risk analysis over a period of five fiscal years from 2007 to 2011. They first perform a visual analysis of the spatial incident distribution of the cases responded to by stations M and N and observe that the optimal stations to absorb the cases of these two closing stations are stations F, G and H (Figure 6). This is indeed confirmed when the analysts runs the average response time analysis with stations M and N closed, which reveals that the three stated stations optimally absorb 723 of the total 730 cases previously handled by stations M and N (assuming a uniform asset speed of 15 nautical miles per hour for all stations in the Great Lakes). They therefore restrict their analysis to these three stations and allow only stations F, G and H to absorb the cases previously handled by the closing stations M and N.

Now, in order to determine an optimal resource allocation strategy, the analysts perform a series of iterative steps in which they assign different types of resources to the three stations. They start their analysis by assigning equal asset speeds of 15 nautical miles per hour for stations F, G and H, and visualizing the average response time redistribution of cases previously handled by the two closing stations M and N. The resulting average response time distribution is shown in Figure 7 (a). As can been seen from the Figure, the analysts determine that the three absorbing stations yield a median average response time of around 91 minutes, with stations F, G and H absorbing 142, 407 and 181 cases respectively. The system also determines that 97% of the total cases fall within the tolerance limit of 120 minutes (shown as a percentage value below the graph). From a visual inspection of the cases responded to by stations M and N (Figure 6), the analysts note that the resulting case distribution among the three stations are valid since station G is located in between stations M and N. This allows them to formulate their initial hypothesis that station G may need to be assigned more assets than stations F and H combined after stations M and N are closed.
Allocating new generation boats to one station

The analysts now turn their attention to the new generation boat reallocation strategies to the absorbing stations once stations M and N are closed. They first consider assigning the available new generation boats (of stations M and N) to stations F, G, and H separately. The resulting average response time and case load distributions of the incidents previously handled by the closing stations M and N are shown in Figures 7 (b-d). The analysts note that assigning all available new generation boats to station G yield the best average response time distribution from among the three scenarios. This can be seen from Figure 7 (c), where station G absorbs 665 cases with a median response time of around 47 minutes. At first, they may be concerned that allowing station G to absorb 665 cases may be sub-optimal and would increase the work load of station G. But realizing that these 665 cases are distributed over a period of five fiscal years, with most incidents occurring during the summer months, they determine that this would constitute to an increase of only 1-2 additional cases a day, which may be acceptable. To confirm this, they use the system’s calendar view to visualize the daily temporal distribution of station G’s cases, and further using their domain knowledge conclude that assigning 1-2 cases a day to station G would indeed be acceptable. Furthermore, they note that the average response time distribution of the other two scenarios (Figures 7 (b, d)) look appealing as well, especially since the cases get distributed uniformly over the three absorbing stations, and that most of the cases fall within the tolerance level of 120 minutes. However, they note that the median average response time shifts to the right in both these cases than that of Figure 7 (c). They also note that station G yields the best performance in terms of average response time case distribution. This further increases their confidence in assigning the cases of the closing stations to station G.

Allocating new generation boats to multiple stations

The analysts now investigate the case in which they assign the available new generation boats (of stations M and N) to one of the three absorbing stations, and in addition, assign new boats to the other absorbing station utilizing the funds that they have in reserve in this analysis scenario. They investigate the three possible scenarios separately, the results of which are shown in Figures 7 (e-g) with each figure showing the distribution of cases when the new generation boats are assigned to stations F+G, G+H and H+F respectively. From the average response time distribution plots, they find that the best case scenario involves a combination of station G with one of the other two stations (e.g., Figures 7 (e, f)), with station G absorbing most of the cases from among the three stations. The third scenario involving assigning the new generation boats to stations F and H (Figure 7 (g)) also yield acceptable average response time case distributions with a relatively more uniform station case load distribution among the three stations than the other two scenarios. They note that this option may be opted for if this was more desirable for the Coast Guard.

Finally, in order to see the average response time impact of upgrading the rescue boats of all three stations to new generation boats with average speeds of 30 nautical miles per hour, the analysts assign all three stations with the new generation boats. The results of this action are shown in Figure 7 (h). The analysts observe that the case load redistribution of the three stations remains very similar to that of all stations having boats with average speeds of 15 nautical miles per hour (Figure 7 (a)). However, as expected, the average response time distribution goes down from a median time of about 91 minutes to 46 minutes. This allows the analysts to make a case to the Coast Guard decision makers for upgrading the assets at the stations and their impacts on potential station closures.

Visualizing station coverage risk profile heatmaps

Finally, the analysts decide to visualize the risk profile heatmaps for the three cases where the new generation boats are assigned to stations F, G and H respectively. This conforms to the average response time case distributions shown in Figures 7 (b-d). The risk profile heatmaps so obtained are shown in Figure 8, with the left image showing the assignment of the new generation boats to station F, the middle image to station G, and the right image to station H. The analysts note that assigning the new generation boats to only stations F or H introduces a potential high risk region in the Great Lakes, and that station G with the new generation boat allocation strategy pays off well in terms of station response coverage when stations M and N are closed. This further increases the confidence in their hypothesis of assigning the cases of the closed stations M and N to station G, and hence allows them to make an informed decision.

6 Domain expert feedback

Our system was informally assessed by two analysts at the U.S. Coast Guard’s Ninth District who are currently using the system to determine the potential risks in the maritime domain associated with the hypothetical allocation of Coast Guard resources, that include both planned day-to-day and contingency operations. In this section, we summarize the feedback that we received from them after conducting several informal interviews with them. The analysts emphasized the need of such systems in the maritime domain that allow them to quickly and easily process large datasets in order to derive actionable results. One of the analysts 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.

The analysts noted that the system had been leveraged throughout its inception and development to assist decision makers in making time-critical resource allocation decisions as well as support long-term alternatives analysis for force structure, coverage, and allocation. As an example, they noted that the system was used...
during the Coast Guard’s planning cycle for Hurricane Irene. Forecast models had Hurricane Irene striking the Eastern Seaboard of the United States over the Labor Day holiday weekend in 2011. Senior Coast Guard officials expected a need for additional resources, and desired to pull resources from the Ninth District to aid in the relief efforts. This system was leveraged, and output was provided to these senior level decision makers that nullified the request to move additional resources from the Ninth District. The system displayed a response need in the Great Lakes that is historically large over this holiday weekend. The potential increase in risk in the Great Lakes was deemed too high and the resources were pulled from elsewhere. The analysts said that to perform this analysis by leveraging Microsoft Excel would have taken hours where this system allowed analysis to be done in minutes. They further noted that the system was built for time critical analysis, but could be leveraged to assist in longer term planning and development of hypothetical alternatives analysis to find efficiencies in their current resource pool. They noted that these efforts have been noticed by other Coast Guard District Commanders as well, and that the system was currently being extended for the Fifth Coast Guard District that encompasses the Mid-Atlantic U.S. States.

We note again that the discussion presented so far in this section is a result of our informal discussions with our end-users to determine the merits and usability of the system. The system has undergone several iterative refinements after receiving feedback from a group of end-users throughout the development process. Although the initial feedback received so far by our end-users has been positive, we plan on conducting a formal user-evaluation of our system in the future to better understand and evaluate the impacts and limitations of our system in the decision making abilities of our end-users. We plan on developing several evaluation and performance metric schemes that will consider the different aspects of our visual analytics system (e.g., visual data cognition, system interactivity and utility from a user perspective, validation of data and results) [33]. We also intend to perform qualitative assessments of the system to assess the merits of the system in the decision making abilities of our end-users (by, for example, providing them with questionnaires after user-training and another after they use the system). Finally, we note that designing an analysis based evaluation to evaluate the likelihood that our end-users have achieved their objectives and minimized the risks after station closures and resource allocations is especially challenging due to the lack of solutions that would act as a control. Thus, domain knowledge and expertise becomes critical in these types of decision making processes.

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 spatial-temporal correlation algorithms that correlate different datasets (e.g., weather, water temperature) with the response dataset in order to provide analysts with insights into the operations of the Coast Guard stations. Furthermore, we plan on quantifying the impacts of reallocating different Coast Guard resources among stations in order to determine optimal mitigative strategies in case of station closures. We also plan on implementing different visualization techniques that compare the trends of different variables to assist analysts in the resource reallocation process. Finally, 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, Zichang Liu and Dr. Brian Fisher for their feedback.

10 FUNDING

This work is supported by the U.S. Department of Homeland Security’s VACCINE Center under Award Number 2009-ST-061-CI0002.

11 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
