Data Driven Multi-Hazard Risk Visualization

Ritesh Sharma[1], Ayme Tomson[2], Emilio Lobato[2], Marcelo Kallmann[1], and Lace Padilla[2]

[1] Electrical Engineering and Computer Science, [2] Cognitive and Information Sciences,
University of California Merced, California, United States

Abstract

This work presents an approach for visualizing aggregate spatial risk data for natural hazards in a way which is not restricted by fixed geographical boundaries and is intended to improve multi-risk awareness in at-risk populations. First, spatial proximity is analyzed to organize occurrences in clusters and the convex hull of each cluster is created in order to define our visualization regions. Then, each region is assigned a risk factor value which is visualized by selecting a color scheme specific to the data variation. The application of this technique is demonstrated using the state of California as a region of interest.

CCS Concepts

• ***Human-centered computing*** → ***Visualization techniques;***

1. Introduction and Related work

Earthquakes, wildfires, hurricanes, and floods are some of the natural hazards that occur every year leaving a trail of destruction impacting communities worldwide. Natural hazards are often categorized by the location in which they occur and the severity of their impact to that location. Due to the complexity of natural hazards, visualizing their occurrences with a unified risk index is difficult. The results are often confusing to both domain professionals and the general public. In this work we present a unified risk index that is visualized in proximity-based regions that allow users to see risk visualization outside of human-constructed boundaries. Our risk index is then classified into 5 categories ranging from low-risk to high-risk, based on recommendations from National Wildfire Coordinating Group (NWCG) [NWC20] and UPSeis [UPS20].

Across much of the literature on visualizing natural disaster risk, visualizations are often created using regions representing human-constructed boundaries, such as municipalities, counties, and states or countries [NCD20, OR13]. However, natural hazards occur without regard to anthropocentric regions - as such geographical boundaries in the context of natural hazards seems arbitrary. This work provides a more accurate representation of region by capturing events based on their Euclidean proximity. The regions are created by computing the convex hull of clustered data points which are the centers of natural hazard occurrences. In this way we transform point-based data into regions. The main focus of this work is to communicate to susceptible populations the aggregate risk without constraints related to geographical boundaries, and to provide an interface to locate a particular region of interest.

2. Multi-Hazard Risk Index

2.1. Data Acquisition

Our multi-risk data includes 17,894 wildfire and earthquake instances for the state of California. The earthquake data (12,452 events) was taken from United State Geological Survey (USGS) [USG20b] for events with magnitude greater than 2.5 over the past 10 years. The wildfire data (5,442 events) was downloaded from the Wildfire Wildland Fire Information website [WF20]. When combining these data sets, we have limited the latitudes and longitudes to 4 decimal places, in order to capture data at the street level [Wik20].

2.2. Regionalization

Regionalization is an important concept in spatial analysis of data. It divides large sets of data into a number of smaller spatially contiguous related regions maintaining the homogeneous nature of the data [Guo08]. Our choice of regionalization technique reflects our focus on improving multi-hazard risk awareness to the least informed audience - the general public. In this work we have created our regions of interest using the hierarchical algorithm *DIANA* (DIvisive ANAlysis) [RK90, ch. 6] available through the *cluster* library in R. To perform clustering, DIANA begins with one large cluster which is divided until each cluster contains a single observation. At each stage, the cluster with the largest diameter is selected and divided in two. The most extreme observation is first found within the cluster, based on dissimilarity, and a new cluster is created. The original cluster is then redistributed across the two clusters based again on dissimilarity.

We then compute the minimum enclosing polygon, or the convex hull, of each cluster. We have considered fire and earthquake occurrences in California for our multi-hazard risk visualization.

DOI: 10.2312/eurp.20201117

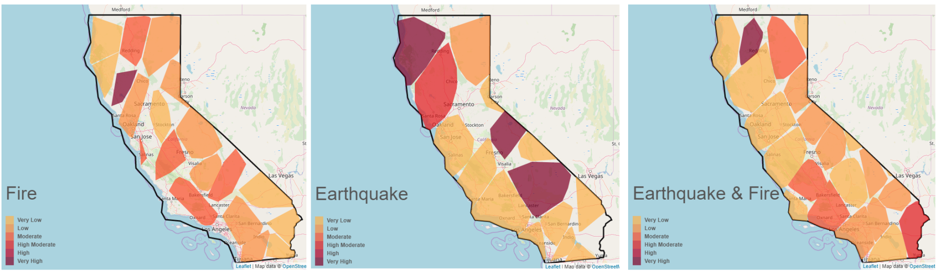

Figure 1: *Risk visualization with for fire (left), earthquake (middle), and aggregate risk (right).*

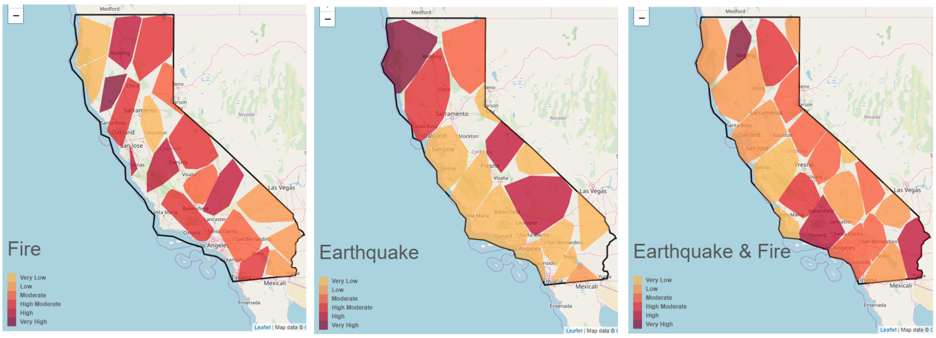

Figure 2: *Risk visualization with cluster-based color scheme for fire (left), earthquake (middle), and aggregate risk (right).*

Convex hulls that crossed the boundary of the state were clipped in order to achieve a result specific to the state.

2.3. Intensity Classification and Risk Index Creation

In order to appropriately aggregate the data, we have defined a relevant index for each natural disaster occurrence. The index represents the severity or intensity of occurrences. From the available data, earthquake magnitude was selected as our metric of earthquake intensity. Earthquake magnitude can be calculated via a variety of algorithmic approaches, such as short-period body wave, duration, moment w-phase, or multiple linear regression for prediction of lateral spread displacement. The USGS makes determinations for which approach is most appropriate for calculating the intensity of a given seismic event [USG20a]. As a result, our USGS Earthquake Catalogue site query included seismic events where magnitude was calculated using any of these approaches rather than only restricting our data to a single magnitude calculation. For the wildfire data, the total acreage burnt area was selected as our metric for wildfire intensity. The classification of each earthquake and wildfire occurrence can be seen in Table 1. The final risk index is

Earthquake Magnitude	Area burnt(Acres)	Intensity
2.5 – 5.4	<10	1
5.5 – 6.0	10 – 100	2
6.1 – 6.9	100 – 300	3
7.0 – 7.9	300 – 1000	4
8.0+	1000+	5

Table 1: *Intensity classification*

computed by taking the arithmetic mean of the intensity of a natural hazard for each cluster. The obtained risk index is normalized and then used to visualize the risk in a given region.

3. Results

To present our first results we have developed a web application to visualize our aggregate multi-hazard risk. Our visualization is implemented using Javascript, HTML5, Leaflet, OpenStreet Map and the D3 library.

Our implementation allows users to select risk visualization from wildfire data, earthquake data, or the aggregate data of both natural hazards. Figure 1 shows the obtained visualizations. Users can select the desired data set to visualize as well as a few different color schemes. It is also possible to inspect specific values of the original data points and to explore different combination of factors to visualize and evaluate the risk of natural hazards for a given location. Our default visualization scheme is based on the well-known choropleth map. This scheme works well in most cases when the data is reasonably uniformly distributed. However, it does not give reasonable results when the data is not uniform. For such cases, a cluster-based color scheme provides a better visualization utilizing the entire range of the selected color scheme range. Results with the cluster-based color scheme is shown in Figure 2. Our webpage is available at `https://sharmrit.github.io/Homepage/MultiHarzardVisualization`

3.1. Conclusions and Future Work

With a focus on communication to susceptible populations, this work presents a new approach to visualize natural multi-hazard risk with regions that are not confined to human-constructed boundaries. We also present a unified risk factor metric and an adaptive data-dependent color scheme selection. As future work we plan to evaluate the use of alpha shapes and non-convex hulls instead of the current convex hull, and we plan to conduct user studies in order to validate and understand the cognitive aspects of the main design choices in the system.

References

[Guo08] Guo D.: Regionalization with dynamically constrained agglomerative clustering and partitioning (redcap). *International Journal of Geographical Information Science 22*, 7 (2008), 801–823. 1

[NCD20] NCDP: US natural hazards index visualization, national center for disaster preparedness, 2020 (accessed April 2, 2020). 1

[NWC20] NWCG: Size class of fire. https://www.nwcg.gov/term/glossary/size-class-of-fire, 2020 (accessed April 2, 2020). 1

[OR13] Opach T., Rød J. K.: Cartographic visualization of vulnerability to natural hazards. *Cartographica: The International Journal for Geographic Information and Geovisualization 48*, 2 (2013), 113–125. 1

[RK90] Rousseeuw P. J., Kaufman L.: Finding groups in data. *Hoboken: Wiley Online Library* (1990). 1

[UPS20] UPSeis: Earthquake magnitude classes. http://www.geo.mtu.edu/UPSeis/magnitude.html, 2020 (accessed April 2, 2020). 1

[USG20a] USGS: Earthquake hazard. https://www.usgs.gov/natural-hazards/earthquake-hazards/, 2020 (accessed April 2, 2020). 2

[USG20b] USGS: United states geological survey: Earthquake catalog. https://earthquake.usgs.gov/earthquakes/search/, 2020 (accessed April 2, 2020). 1

[WF20] WF: Wildfire. https://wildfire.cr.usgs.gov/firehistory/data.html/, 2020 (accessed April 2, 2020). 1

[Wik20] Wikipedia: Decimal degree. https://en.wikipedia.org/wiki/Decimal_degrees/, 2020 (accessed April 2, 2020). 1