Quantitative Visualization about Differences between Scientists concerned Nature Disasters and Historic Events

Shi Shen, Changxiu Cheng*, Kai Su, Jing Yang, Shanli Yang
Key Laboratory of Environmental Change and Natural Disaster
State Key Laboratory of Earth Surface Processes and Resource Ecology
Academy of Disaster Reduction and Emergency Management, Beijing Normal University
Beijing, 100875, China
Email: chengcx@bnu.edu.cn

***Abstract*—How to process massive historic natural disasters events is a great challenge to recognize patterns. And more and more scientific research data provides a new source of nature disasters. In this study, the biclustering method is used to categorize the scientists concerned natural disasters and historic events. Cartograms, one kind of transformed maps, are created to highlight numbers of publications and events in a country. Reaction index (RI) is introduced to evaluate the difference between scientists concerned nature disasters and historic events. The results show that biclustering is a useful method to categorize data with high volumes and dimensions. Cartograms could represent conceptual patterns that are difficult to be displayed in regular maps. Analysis indicates that earthquakes and landslides attract relatively more concerns from scientists in the north hemisphere; floods are more focused by scientists in the south hemisphere. Although droughts are not significant in the cartogram of historic events, they obtain attentions from scientists of inland as well. The distribution of RIs shows that more scientists need to put more efforts in dealing with natural disasters, especially in Indonesia and Philippines.**

Keywords—natural disasters; spatial patterns; biclustering; cartograms; reaction index

I. INTRODUCTION

Spatial patterns about natural disasters are the distributions of occurrence or appearance of natural disasters on the Earth. As a kind of prior knowledge, they play an important role in the research about natural hazards, warnings, and preventions [1]. With the rich of statistical data, scientists often summarized some patterns according to historical events. Guan et al. [2] analyzed the trends and distributions of five main meteorological disasters in China based on the historical and official disaster event records during 1950-2013. Karremann et al. [3] clustered and discussed the Potential losses associated with winter storms in Germany during the recent decades. Peña et al [4] concluded seven patterns about strong winds by working with strong wind events data in Catalonia.

On the other hand, research of natural disasters themselves also reflects natural disasters which scientists concern. Thus, the research about natural disasters provides a new avenue to indicate conceptual patterns about natural disasters from a scientists' view. Analyzing academic data of natural disasters (i.e., published scientific literature about natural disasters) and disaster events provides a quantitative approach for evaluating differences between scientists concerned natural disasters and historic natural disaster events.

However, due to the development of data store technology and data sharing policy, more and more historic natural disasters events data is available. Meanwhile, disasters associated with natural hazards have been gaining increasing scientific interest in the last decades due to their significant socio-economic and environment impacts [5]. Hence, how to process these massive data has become a great challenge for experts to recognize patterns.

This paper focuses on spatial patterns of scientists concerned natural disasters and historic events on a global scale from massive data through biclustering and cartograms methods. Additionally, a reaction index (RI) is designed evaluating the difference between scientists' attentions and historic events. The rest of this paper is organized as follows: Section II describes characters of the data and sources used in this study. In section III, biclustering and cartograms are introduced to cluster natural disaster categories and illustrate spatial distributions. Results and discussions are shown in section IV. Main conclusions are given in section V.

II. DATA AND SOURCES

Historic natural disasters events are derived from the EM-DAT [6] database. The EM-DAT database includes all significant disasters from 1900 until present. These events resulted in ten or more people died, or affected at least 100 people, or made governments declared a state of emergency, or called for international help. The historic natural disaster events dataset（i.e., historic events dataset） is retrieved by national searching in the EM-DAT database from 1900 to 2015. It results in more than 23000 records, which has a coverage of almost all the global land areas except Greenland and the Antarctica. These two territories are excluded due to lacking of human activities. Scientific research datasets are derived by topic searching the Web of Science (WoS) database with a period of from Jan 1900 to June 2015.

TABLE I. THE SOURCES OF SCIENTIFIC RESEARCH DATASET

Classification	Description	Content
Part I	List of journals whose name contain disaster or hazard	Disasters; Natural Hazards; Disaster Advances; Natural Hazards Review; Journal of Hazardous Materials; Natural Hazards and Earth System Sciences; Disaster Prevention and Management; Geomatics Natural Hazards Risk; International Journal of Disaster Risk Science; Risk Management Journal of Risk Crisis and Disaster. Environmental Hazards-Human and Policy Dimensions; Remote Sensing: Inversion Problems and Natural Hazards; Fire and Polymers Iv Materials and Concepts for Hazard Prevention; Japca The International Journal of Air Pollution Control and Hazardous Waste Management; Journal of Environmental Science and Health Part A-Toxic/Hazardous Substances & Environmental Engineering; Journal Of Environmental Science And Health Part A Environmental Science And Engineering Toxic And Hazardous Substance Control;
Part II	Two famous journals	Nature; Science.
Part III	WOS categories	Ecology; Forestry; Geology; Limnology; Geography; Soil Science; Remote Sensing; Water Resources; Geography Physical; Environmental Studies; Engineering Geological; Environmental Sciences; Geochemistry Geophysics; Engineering Environmental; Geosciences Multidisciplinary; Agriculture Multidisciplinary; Meteorology Atmospheric Sciences;

Table I lists sources of research dataset datasets including 17 journals whose names contain “Disaster”, “Disasters”, “Hazard” or “Hazards”. Nature and Science are included because these two journals have worldwide influence and authorities. The third source is the 17 relevant disciplines associating disasters and hazards. Finally, 51,202 scientific research records are derived. Records associating natural disasters are manually identified 26,312 (51.4% of all) through searching titles or keywords of a record. The rest of scientific research datasets are chemical disasters and none-natural disasters.

III. METHOD

In this study, aforementioned two datasets are taken advantage to visualizing natural disaster patterns. An integrated methodology is shown in *Fig. 1*. First of all, biclustering method is used to categorize natural disasters and countries. Biclustering refers to a class of clustering algorithms that cluster data matrix on columns and rows dimensions simultaneously [7]. This method was identical in various disciplines with different names, e.g. co-clustering, subspace clustering, bidimensional clustering, etc. Chen and Church applied the term of biclustering into analyzing gene expression data for the first time [8]. Comparing with one-dimensional clustering method, this approach is capable of detecting clusters of columns and rows without potential perceptual bias.

The first step of biclustering is to build data matrices. To compare the difference of scientists concerned disasters and historic events, two matrices are constructed separately based on the two datasets. Information about the matrix carrying the scientists concerned disasters is derived from scientific research data. A natural disaster research article is analogous to a natural disaster record. So, countries, where scientists come from, represent places where natural disasters attract attentions. This information is extracted from the address of corresponding author or first author if the corresponding author

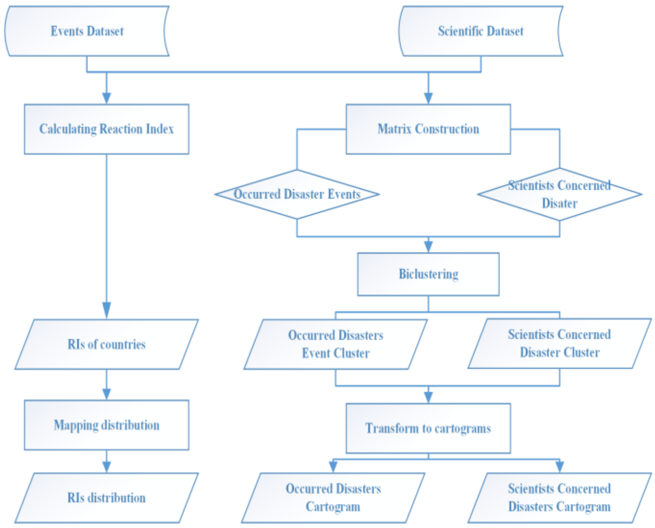

Fig. 1. A diagram of quantitatively visualizing scientists concerned natural disasters and historic events.

is not available. Scientists concerned natural disasters types are concluded from the keywords from articles.

The rows of the matrix storing the natural disaster events represent the countries where disaster events occurred, according to a three-letter ISO (International Organization for Standardization) country code. The columns indicate natural disasters events types which are identified through “disastertype” and “subdisastertype” variables in event metadata. The matrix of natural disaster events has the same structure with the matrix of scientists concerned disasters that contain rows with countries and columns with natural disaster types

In order to highlight variables, such as numbers of publications and frequency of disasters, cartograms are used. A cartogram is a special thematic map in which land area or distance is substituted for a certain variable, whereas the topological structure of transformed geographic entities keeps the same. In this study, we fellow Gastner and Newman's method [9] to create density-equalizing cartograms. Two cartograms will be generated using this method with relative geographic positions among countries remain.

The first cartogram that transforms countries' areas adapting to their published articles shows efforts and attentions of scientists from different countries. Larger size indicates more attentions and efforts. Meanwhile, countries are colored according to results of scientific data using the biclustering method. This cartogram indicates spatial patterns of scientists concerned natural disasters.

In the second cartogram, the territory of a country represents the number of the countries' recorded disaster events instead of areas. The size of a country's area is proportional to their recorded disasters events. The second cartogram shows spatial patterns of historic natural disaster events.

Additionally, to quantitatively evaluate the difference between scientific research and historic events, we introduce a reaction index (RI) that indicates scientists' publications from one country comparing to its suffered disasters events. The RI is shown as follows:

$$RI_i = \left(P_i / TP\right) - \left(E_i / TE\right) \tag{1}$$

In the Equation (1), the RI_i indicates reaction index of the country i; P_I represents the number of research articles of the country i; TP is the total number of articles associating natural disasters; E_i indicates the frequency of natural disasters occurred in the country i; TE is the sum of natural disaster events from 1900 to 2015. The value of RI indicates the level of scientists' attentions to natural disasters. The larger the RI is the more attentions response to natural disasters. The positive or negative of RI shows that the attentions from scientists are adaptive to natural disasters or not. Classification of absolute values of RIs is natural breaks classification that seeks to minimize each class's average deviation from the class meanwhile maximizing each class's deviation from the means of the other groups.

IV. Results and Discussion

Relying on the natural disasters events from EM-DAT and scientific research data from WoS, natural disasters types are summarized. Table II lists five disaster groups: atmospheric disasters, geophysical disasters, meteorological disasters, biological disasters and others.

The resultant matrix of scientific research contains 135 countries and 27 natural disaster types. And the resultant matrix of natural disaster events contains 136 rows and 19 columns. Three thematic clusters of scientists concerned natural disasters are illustrated in *Fig. 2A*. Besides, five categories of historic natural disaster events are summarized as well (shown in *Fig. 2B*).

The distribution of scientists concerned natural disaster displays a great difference from the distribution of historic disaster events. In general, distribution of scientists concerned natural disasters shows a more unbalanced pattern. North America, Europe, South Asia and East Asia are the four great active groups of scientists concerning about natural disasters due to the high volume of scientific research. In the southern hemisphere, only Australia and New Zealand show significant research effort in defending natural disasters. Besides, South America, Africa, and Russia are a relative lack of efforts and attentions comparing to their physical areas

TABLE II. TYPES OF NATURAL DISASTER

Category	Natural Disasters Events Types (Abbreviation)	Scientists Concerns Natural disasters (Abbreviation)
Atmospheric	Tornado (STO); Violent Wind(SVW); Tropical cyclone(STC); Extratropical Cyclone(SEC); Local Storm(SLS); Extreme Heat Temperature(EH); Extreme Cold Temperature(EC);	Extreme Heat Temperature(EH); Extreme Cold Temperature(EC); Climate Change(CC); Sandstorm(SD); Storm(ST)
Geophysical	Earthquake(EQ); Volcano(VO); Landslide(LS);	Earthquake(EQ); Volcano(VO); Landslide(LS); Soil Erosion(SE); Wind Erosion(WE); Collapse(CL);Rocky Desertification(RD); Salination(SL); Subsidence(SB);
Hydrological	Coastal Flood(FCF); Riverine Flood(FRF); Flash Flood(FFF); Drought(DR); Ocean Disaters(OD);	Lake Hazard(LH); Flood(FL); Drought(DR); Red Tide(RT); Sea Ice(SI); Extreme Wave(EW); Tsunami(TS)
Biophysical	Wild Fire(WF); Famine(FA); Insect Infestation(II); Epidemic(EP);	Wild Fire(WF); Insect Infestation(II); Epidemic(EP); Weed(WD); Rodent(RD);
Others		Impact(IM)

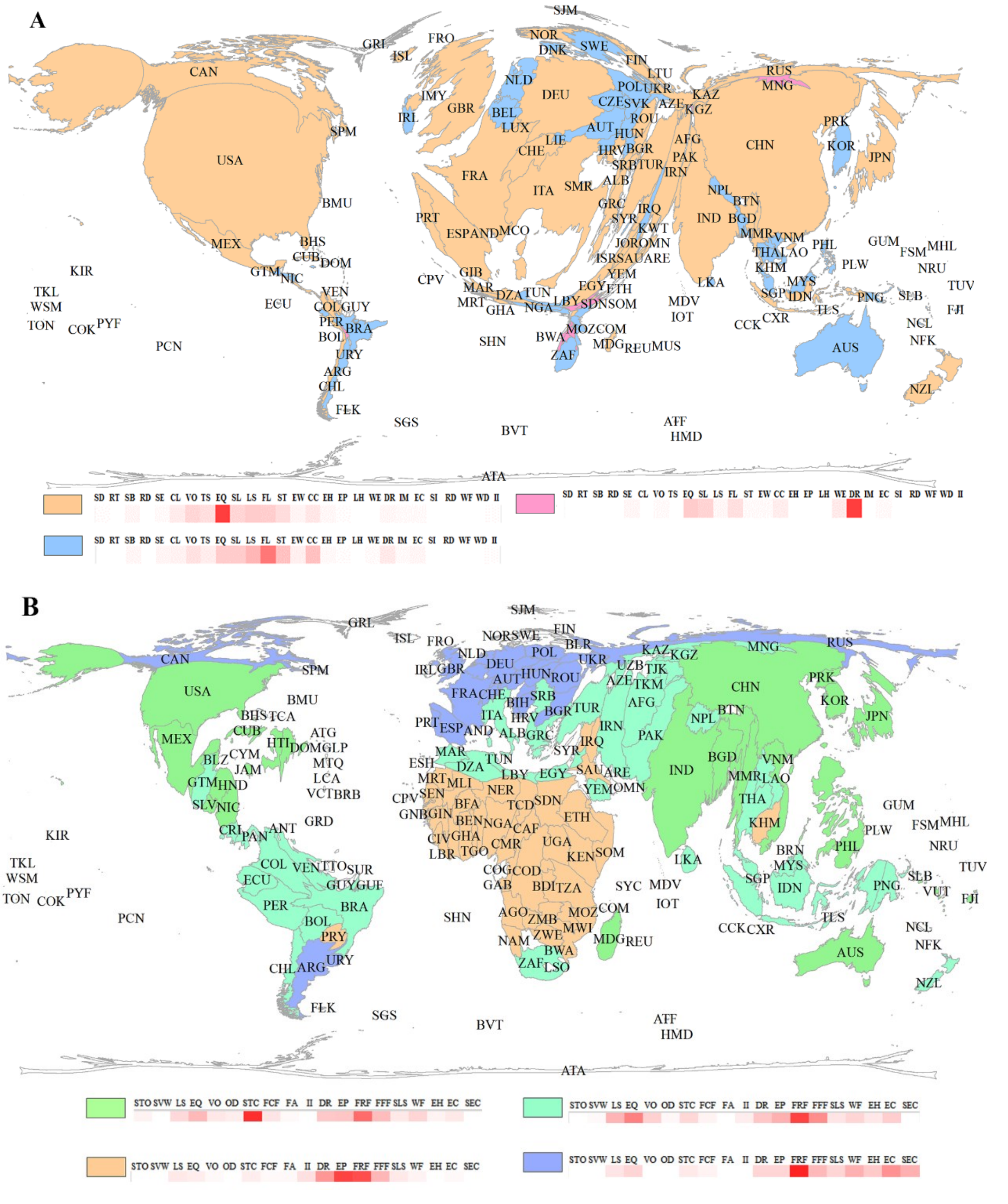

Fig. 2. Spatial patterns of natural disasters from 1900 to 2015. (1A indicates scientists concerned natural disasters; 1B shows historic natural disasters events; Blanks indicate none data)

According to *Fig. 2A*, yellow countries whose scientific societies pay more attentions on the research of earthquakes and landslides, occupy most part of the earth. This reveals that most scientists have strong interests and concerns to crippling natural disasters (e.g., earthquakes, landslides). Actually, in these yellow countries, the U.S., India, China, South Korea and Japan (grass green countries in *Fig. 2B*) suffer more frequent storms. Canada, UK, Germany, France Spain etc. (purple countries in *Fig. 2B*) suffer more floods in reality. West Asia and most parts of Africa face frequent floods and droughts. In addition, the earthquake is the second frequent natural disasters in Italy, North Africa and Middle Asia (jade green in *Fig. 2B*).

Floods are the second natural disasters that draw attentions from human societies. In *Fig. 2A*, blue area indicates where floods are got attentions from scientists. Netherland, Belgium, most Central Europe, Ecuador, Uruguay, Argentina, South Africa, Nigeria, and Southeast Asia not only suffer frequent floods but also focus on this disaster. On the contrary, Australia, and South Korea are recorded suffering more storm events than floods. The third natural disaster natural disaster is the drought. This disaster especially raises concern from societies of Mongolia, Bolivia, former Sudan, Mozambique, Zimbabwe these inland countries suffering drought events in the physical world as well.

Fig. 3 indicates a spatial distribution of RIs. The absolute values of reaction indexes substitute territories of countries and are proportional to the darkness of the color. The green region shows that these countries have non-negative RIs. In specifically, the U.S., Canada, West Europe, Northeast Asia, Australia and New Zealand are with positive RIs. It indicates, in these countries, scientists put more concerns and attentions to natural disasters than their suffered disaster events. Oppositely, red region indicates countries with negative RIs. Overall, the red region is much larger than the green. Especially, Indonesia and Philippines are two countries with high negative RIs, which represent scientists in these two countries need more attentions on natural disasters

V. CONCLUSIONS

The matrix of historic events, with 137 rows and 27 columns, is clustered into four categories. The matrix of scientists concerned disaster, with 136 rows and 19 columns, is summarized three thematic categories. This confirms biclustering is a useful method for clustering high dimensional and volume data.

A cartogram is a powerful tool to illustrate nontraditional variables for experts and policymakers. In *Fig. 2A*, territories of countries deform greatly since their volume of natural disaster research. Experts can easily realize the great difference of scientists' concern among countries. In a similar way, *Fig. 2B* display physical spatial patterns of historic events distinctly and significantly, that will help researchers and policymaker understand natural disasters distribution more conveniently.

In addition, spatial patterns of scientists concerned natural disasters and historic events show that earthquakes and landslides are the focus of most human societies of the world, especially in the northern hemisphere. Floods and droughts are another two types getting more attention from scientists. The proportion of reactions to floods in the south hemisphere is much larger than in the north.

Finally, the distribution of reaction indexes indicates that most part of the world needs more scientists' concerns about natural disasters than the suffered disasters, especially in Central and South America, the Caribbean, East Africa, and Southeast Asia. Indonesia and Philippines are two countries with a great gap between scientists' concern and historic suffered disasters events.

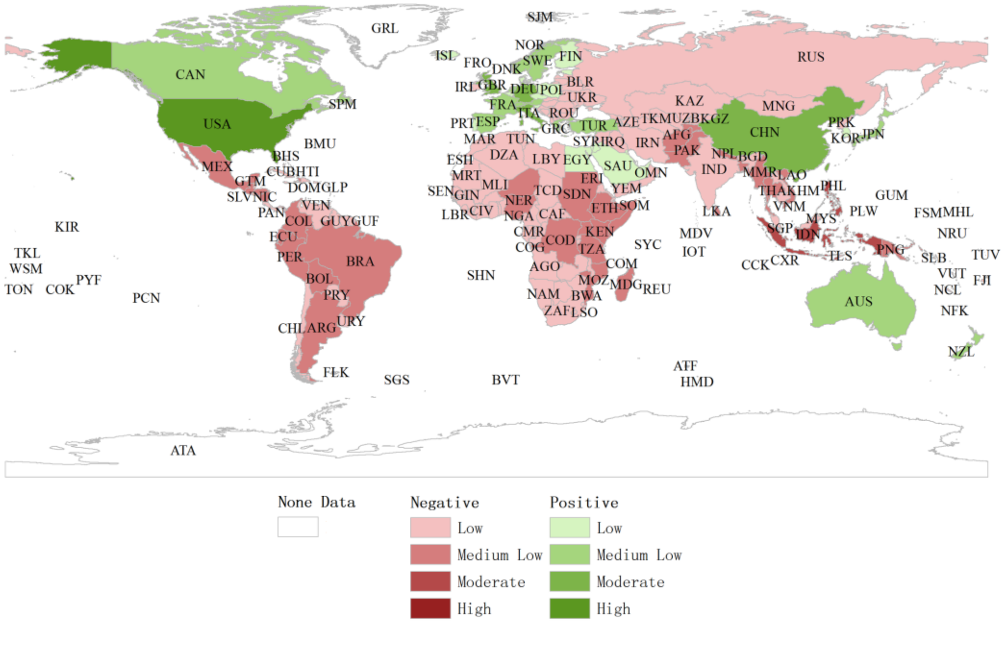

Fig.3.A Spatial distribution pattern of reaction indexes in a country level

Acknowledgment

This work was supported by the National Natural Science Foundation of China [Grant Numbers: 41222009, 41271405] and International Partnership Program of Chinese Academy of Sciences [Grant Number: 131551KYSB20160002].

References

[1] E. O'Neill, M. Brennan, F. Brereton, and H. Shahumyan, "Exploring a spatial statistical approach to quantify flood risk perception using cognitive maps," Nat. Hazards, vol. 76, no. 3, pp. 1573–1601, 2015.

[2] Y. Guan, F. Zheng, P. Zhang, and C. Qin, "Spatial and temporal changes of meteorological disasters in China during 1950--2013," Nat. Hazards, vol. 75, no. 3, pp. 2607–2623, 2015.

[3] M. K. Karremann, J. G. Pinto, P. J. von Bomhard, and M. Klawa, "On the clustering of winter storm loss events over Germany," Nat. Hazards Earth Syst. Sci. Discuss., vol. 2, no. 3, pp. 1913–1943, 2014.

[4] J. C. Peña, M. Aran, J. Cunillera, and J. Amaro, "Atmospheric circulation patterns associated with strong wind events in Catalonia," Nat. Hazards Earth Syst. Sci., vol. 11, no. 1, pp. 145–155, Jan. 2011.

[5] K. Papagiannaki, K. Lagouvardos, and V. Kotroni, "A database of high-impact weather events in Greece: a descriptive impact analysis for the period 2001–2011," Nat. Hazards Earth Syst. Sci., vol. 13, no. 3, pp. 727–736, Mar. 2013.

[6] P. H. D. Guha-Sapir, R. Below, "EM-DAT: International Disaster Database." [Online]. Available: http://www.emdat.be/.

[7] S. C. Madeira and A. L. Oliveira, "Biclustering algorithms for biological data analysis: a survey.," IEEE/ACM Trans. Comput. Biol. Bioinform., vol. 1, no. 1, pp. 24–45, 2004.

[8] Y. Cheng and George m Church, "Biclustering of Expression Data," Intell. Syst. Mol. Biol., vol. 8, p. 93, 2000.

[9] M. T. Gastner, M. E. J. Newman, and M. F. Goodchild, "Diffusion-Based Method for Producing Density-Equalizing Maps," Proc. Natl. Acad. Sci. U. S. A., vol. 101, no. 20, pp. 7499–7504, 2004.