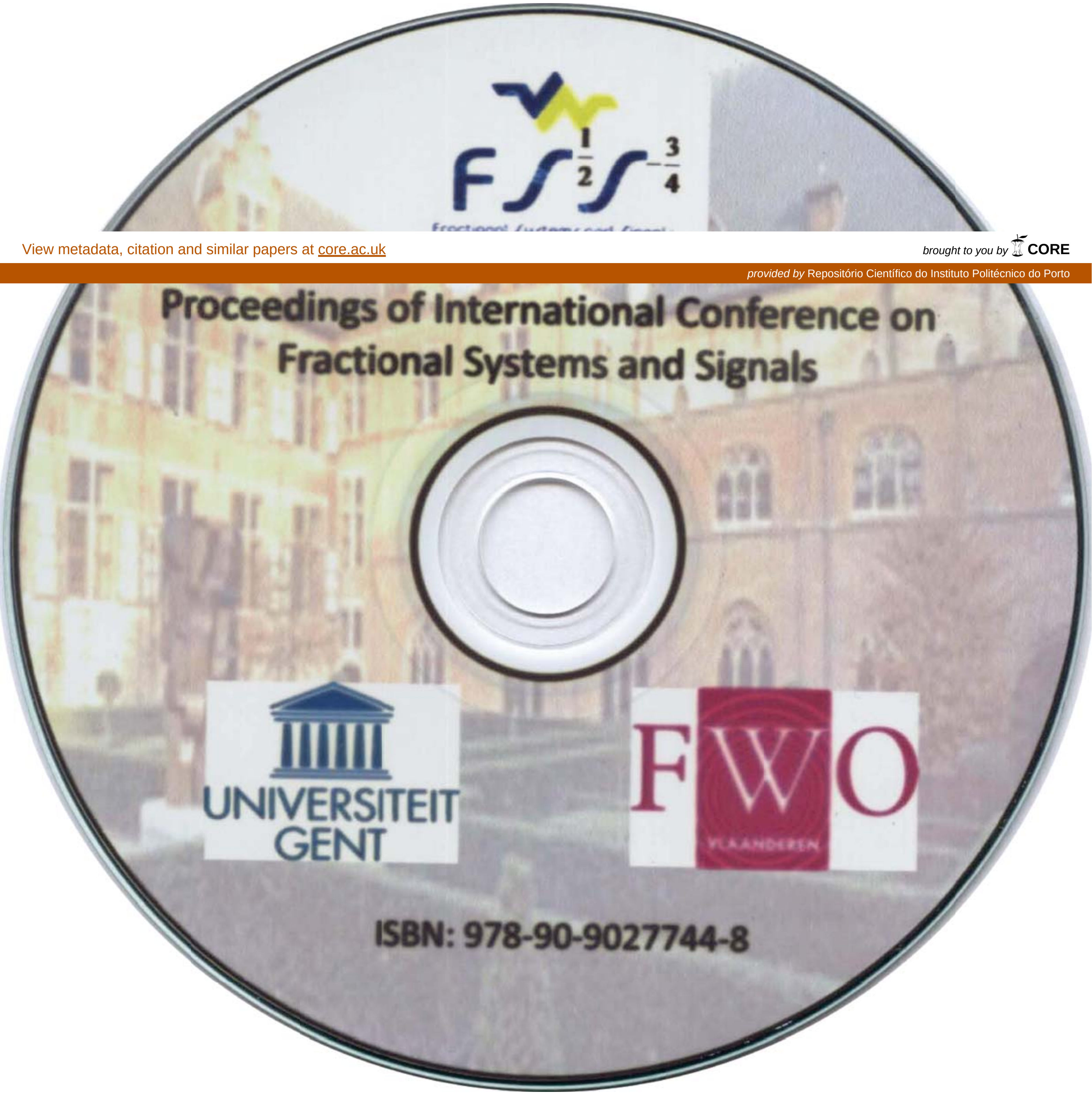
Proceedings of International Conference on
Fractional Systems and Signals
UNIVERSITEIT
GENT
FWO
VLAANDEREN
ISBN: 978-90-9027744-8

NEW FEATURES TO LOOK AT NATURAL PHENOMENA

António M. Lopes[a], Carla Pinto[b], and J.A. Tenreiro Machado[c]

[a] *UISPA, IDMEC - Polo FEUP*
Faculty of Engineering, University of Porto
Rua Dr. Roberto Frias, 4200-465 Porto PORTUGAL
[b] *School of Engineering, Polytechnic of Porto,*
Centre of Mathematics, University of Porto,
GECAD - Knowledge Engineering and Decision Support Research Center
Rua Dr António Bernardino de Almeida, 431, 4200-072 Porto, PORTUGAL
[c] *Department of Electrical Engineering*
School of Engineering, Polytechnic of Porto
Rua Dr António Bernardino de Almeida, 431, 4200-072 Porto, PORTUGAL

***Abstract* — The paper focuses the patterns seen in the number of victims from natural catastrophic phenomena. We consider the number of victims of storms from 1900 up to 2013 in 11 countries and study the distributions of the events with more than 30 deadly victims. The similarities among events across the 11 countries are analysed using agglomerative hierarchical clustering. Countries belonging to the same cluster are similar with respect to fatalities. Power laws and hierarchical clustering provide comparable results for the data. Future work is needed in order to explore these numerical tools in more countries and in victims of other hazards.**

***Keywords:* catastrophic events, power laws, hierarchical clustering**

I – Introduction

In 1896, Vilfredo Pareto introduced power laws (PLs)to model the distribution of individuals incomes [13]. Since then, studies of applications of PLs to real world phenomena have largely increased. Examples are in wealth distribution and expenditure [11, 5], city size distribution [19, 6], number of articles' citations and scientific production [10, 15], number of victims in wars, terrorist attacks, and earthquakes [16, 3], and words' frequency [19]. Interesting reviews on PL behavior and applications can be found in [17, 14].

The most well known examples of PL distributions are the Pareto [13] and the Zipf [18, 19] laws. The later is also known as rank-size rule.

Let X be a non-negative discrete random variable following a PL distribution. Then, its complementary cumulative distribution function is of the form $F(x) = P(X \geq x) = \frac{C}{\alpha-1}x^{-(\alpha-1)}$, where $\alpha > 0,\ C > 0$. In the text, we will consider $\tilde{\alpha} = \alpha - 1$ and $\tilde{C} = \frac{C}{\tilde{\alpha}}$. The probability function of a discrete random variable following Pareto distribution is given by:

$$P(X = x) = Cx^{-\alpha} \tag{1}$$

Zipf law is a special case of the Pareto law with exponent $\tilde{\alpha} = 1$.

Application of PL behavior in natural or human-made phenomena usually comes with a log-log plot, where the axes represent the size of an event and its frequency. The log-log plot is asymptotically a straight line with negative slope.

In this paper, we apply PLs and hierarchical clustering to the study of casualties occurring in natural hazards. In Section II, we review some results in the literature concerning patterns of the number of victims in distinct natural hazards. Numerical results are presented in Section III. Finally, we state the main conclusions of this work and some future research directions.

II – Power laws in natural hazards

In the last decades, researchers have been focusing in the patterns of natural and human made hazards. Examples of the later are wars, terrorist attacks, tornadoes, earthquakes, landslides, floods, and other severe occurrences [16, 7, 9, 8, 3].

Many attentive explanations have arisen in the literature. Nevertheless, the understanding of these phenomena has always to consider political, geographical, historical, and, even cultural, factors. Predicting the number of casualties in natural or human-made disasters is extremely important in developing pre-disaster strategies. Aspects like rationalization of medical supplies and food, gathering emergency teams, organize shelter spaces, amongst others, have to be dealt with, in order to minimize the damage.

Guzzetti [7] considers landslide events in specific periods in different countries, such as Italy, Canada, Alps, Hong Kong, Japan, and China. He shows that the plot of the cumulative distribution function of the number of landslide events vs the number of casualties is well approximated by a straight line. This result suggests a PL distribution of the data.

In 2005, Jonkman [9] studied the number of human deaths caused by three types of floods (river floods, flash floods and drainage issues), between January 1975 and June 2002. Highest average mortality was computed for flash floods. The author plotted the global frequency of events with N or more deaths vs N. He observed a PL behavior in all data but the flood data. Becerra *et al* [2] use the same data set as Jonkman [9], but consider all disasters combined, both globally and disaggregated by continent. They obtained straight-line log-log plots for all disasters combined. The slopes of the casualties PL distributions were smaller than those for modern wars and terrorism. The explanation for this remained an open question. Another unsolved issue was the existence of PL behavior in combined disasters and not in individual disasters, such as floods. Here it is worth mentioning that casualties in earthquakes verified a PL distribution [9, 2].

III – Results and discussion

In this section we apply PLs and hierarchical clustering to analyze data from natural hazard phenomena.

We collected data from the website `http://www.emdat.be`, concerning victims of storms. We consider the period from 1900 till 2013 and study the distribution of the events with more than 30 deadly victims.

Figure 1 represents the normalized rank/frequency log-log plot of the variable *"Number of deadly victims in USA in the period* $[1900, 2013]$*"*. To construct the plot, we first sort the number of victims in descending order, and number them, consecutively, starting from one [12]. Then a normalization of the values is carried out, meaning that, the numbers of victims (x-axis) are divided by the corresponding highest value, and the ranks (y-axis) are divided by the rank of the smallest number of victims. Finally, a PL is adjusted to the data using a least squares algorithm. It is observed a PL behavior with parameters $(\tilde{C}, \tilde{\alpha}) = (0.0059, 0.975)$.

We repeat the study presented in Figure 1 to the 11 countries listed in Table 1 that shows the corresponding parameters $(\tilde{C}, \tilde{\alpha})$.

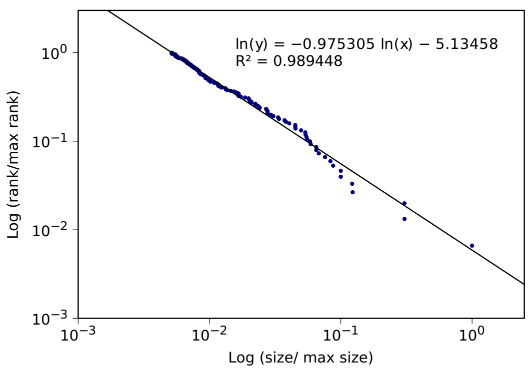

Figure 1: *Rank/frequency log-log plot for the number of deadly victims occurred in storms in the USA in the period* $[1900, 2013]$.

Figure 2 shows the locus of the PLs parameters $(\tilde{C}, \tilde{\alpha})$ for the 11 countries in Table 1.

Similarities between countries, in terms of victims caused by storms, are determined using agglomerative hierarchical clustering analysis. Firstly, the Euclidean distance, s, between each pair of $(\tilde{C}, \tilde{\alpha})$ points is computed as:

Table 1: *Parameter values of a PL distribution for casualties of storms in 11 countries.*

	$\tilde{C}$	$\tilde{\alpha}$	min	max	Events
USA (US)	0.0059	0.975	6000	31	151
Philippines (PH)	0.0117	0.874	5956	31	153
India (IN)	0.0124	0.6	60000	31	97
Vietnam (VN)	0.024	0.747	7000	42	39
Taiwan (TW)	0.0669	0.751	1046	31	27
Mexico (MX)	0.0909	0.789	960	34	21
Madagascar (MG)	0.0553	1.319	363	32	23
Bangladesh (BD)	0.0252	0.498	61000	31	79
China (CN)	0.0043	0.741	50000	31	92
Japan (JP)	0.0389	0.648	5098	32	68
Korea (KR)	0.0507	0.897	1104	32	23

$$s = \sqrt{(\tilde{C}_1 - \tilde{\alpha}_1)^2 + (\tilde{C}_2 - \tilde{\alpha}_2)^2} \quad (2)$$

producing a 11×11 symmetrical matrix. In a second step, based on that metric, the single linkage criterion is used to create the hierarchical cluster tree and, for a given threshold value measured by the Euclidean distance, the cluster tree is pruned and the clusters created.

Figure 3 shows the cluster tree (dendrogram). Pruning the tree in order to take four clusters we obtain a first region comprising the USA, Philippines and Korea. The second region consists of India, Japan, Vietnam, China, Taiwan and Mexico. Bangladesh and Madagascar appear isolated from every other country. This means that clusters are essentially dictated by parameter $\tilde{\alpha}$. The single country clusters correspond to the extreme values of that parameter.

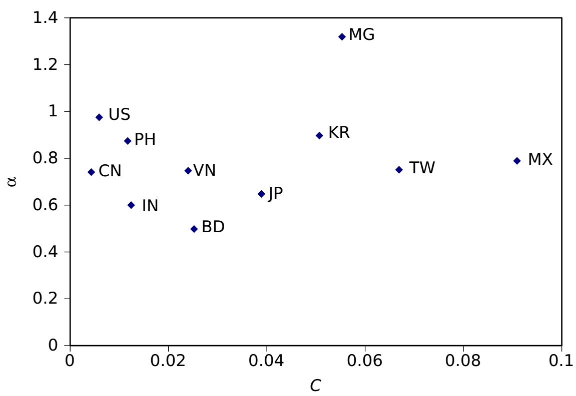

Figure 2: *Locus of the PLs parameters* $(\tilde{C}, \tilde{\alpha})$ *for the 11 countries in Table 1.*

To complement the analysis a second metric, d, defined as:

$$d = |\tilde{\alpha}_1 - \tilde{\alpha}_2| + |log(\tilde{C}_1/\tilde{C}_2)| \quad (3)$$

is considered, based on the parameters $(\tilde{C}, \tilde{\alpha})$, and hierarchical clustering analysis is again applied based on that metric.

Figure 4 shows the hierarchical tree for the 11 countries. It can be seen that, pruning the dendrogram in order to take four clusters, the United States and

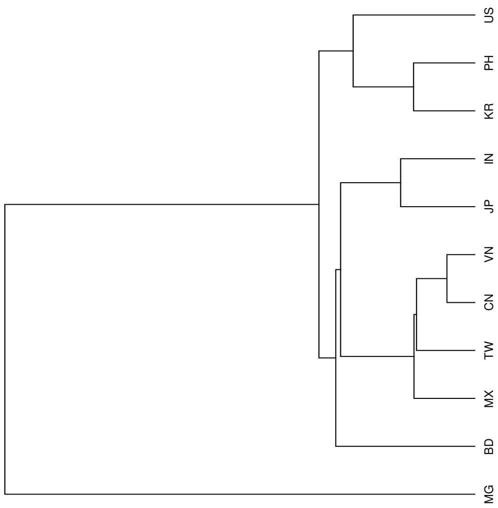

Figure 3: *Dendrogram corresponding to agglomerative hierarchical clustering based on Euclidean distance metric, s.*

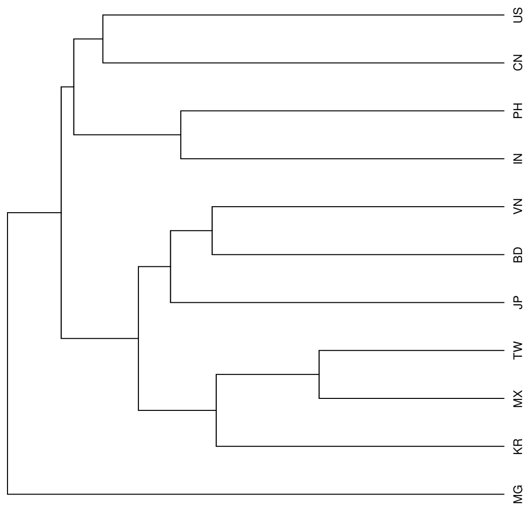

Figure 4: *Dendrogram corresponding to agglomerative hierarchical clustering based on metric d.*

China are included in a cluster. This means the corresponding storm events are similar with respect to fatalities. Philippines and India form another cluster, in this case justifiable by the geographic proximity and similar urbanism. The third cluster comprises Vietnam, Bangladesh, Japan, Taiwan, Mexico and Korea. The geographic proximity may again justify the similarities between these countries, except Mexico, that is more awkward. Finally, Madagascar appears apart from the rest.

It is observed that the PL analyses and the agglomerative hierarchical clustering provide comparable results for the data. Nevertheless, more work is needed in order to explore these numerical tools and more data has to be analysed.

IV – Conclusion In this paper we focus on the patterns seen in the number of victims from natural catastrophic phenomena. We considered the number of victims of storms from 1900 to 2013 in 11 countries and we studied the distribution of the events with more than 30 deadly victims. The similarities among events across the 11 countries were also analysed using agglomerative hierarchical clustering. It is observed that, pruning the hierarchical cluster tree in order to take five clusters, the USA, China, Philippines and India, are included in one cluster. This means that these countries are similar with respect to fatalities. The other 7 countries are divided in the remaining 4 clusters. Geographic proximities seem to explain the obtained cluster division. Power laws and hierarchical clustering provide comparable results for the data. Future work is needed in order to explore these numerical tools in more countries and in victims of other hazards.

Acknowledgments

Authors which to thank the Polytechnic of Porto, through the *PAPRE Programa de Apoio à Publicação em Revistas Científicas de Elevada Qualidade* for financial support.

References

[1] P. Bak, K. Chen, C. Tang. A forest-fire model and some thoughts on turbulence. *Physics Letters A* **147** Issues 5–6 (1990) 297–300.

[2] O. Becerra, N. Johnson, P. Meier, J. Restrepo, and M. Spagat. Natural Disasters, Casualties and Power Laws: A Comparative Analysis with Armed Conflict. *Proceedings of the annual meeting of the American Political Science Association*, Marriott, Loews Philadelphia, and the Pennsylvania Convention Center, Philadelphia, PA, (2006).

[3] J. C. Bohorquez, S. Gourley, A. R. Dixon, M. Spagat, N. F. Johnson. Common Ecology Quantifies Human Insurgency. *Nature* **462** No. 7275 (2009) 911–914.

[4] J.H. Brown, V.K. Gupta, B.-L. Li, B.T. Milne, C. Restrepo, and G.B. West. The fractal nature of nature: power laws, ecological complexity and biodiversity. *Phil. Trans. of the Royal Society London B* **357** (2002) 619–626.

[5] F.C. Figueira, N.J. Moura, M.B. Ribeiro. The Gompertz-Pareto income distribution. *Physica A* **390** (2011) 689–698.

[6] X. Gabaix. Zipf's Law and the Growth of Cities. *American Economic Review Papers and Proceedings* **LXXXIX** (1999) 129–132.

[7] F. Guzzetti. Landslide Fatalities and the evaluation of landslide risk in Italy. *Engineering Geology* **58** (2000) 89–107.

[8] N.F. Johnson, M. Spagat, J.A. Restrepo, O. Becerra, J.C. Bohorquez, N. Suarez, E. M. Restrepo, R. Zarama. Universal patterns underlying ongoing wars and terrorism. *arXiv:physics/0605035v1* (2006).

[9] S.N. Jonkman. Global Perspectives on Loss of Human Life Caused by Floods. *Natural Hazards* **34** (2005) 151–175.

[10] A.J. Lotka. The frequency distribution of scientific productivity. *J Washington Acad Sci* **16** (1926) 317–324.

[11] B. Mandelbrot. The Pareto-Levy Law and the Distribution of Income. *International Economic Review* **I** (1960) 79–106.

[12] M.E.J. Newman. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics* **46** (2005) 323–351.

[13] V. Pareto. *Cours d'Economie Politique*, Geneva, Switzerland: Droz (1896).

[14] C.M.A. Pinto, A.M. Lopes, and J.A. Tenreiro Machado. A review of power laws in real life phenomena. *Communications in Nonlinear Science and Numerical Simulations*, accepted for publication, January (2012).

[15] D. J. de S. Price. Networks of scientific papers. *Science* **149** (1965) 510-515.

[16] D. C. Roberts, D. L. Turcotte. Fractality and self-organized criticality of wars. *Fractals* **6** (1998) 351–357.

[17] D. Sornette. *Critical Phenomena in Natural Sciences*, chapter 14. Springer, Heidelberg, 2nd edition (2003).

[18] G. Zipf. *Selective Studies and the Principle of Relative Frequency in Language*. Harvard University Press, cambridge, MA (1932).

[19] G. Zipf. *Human Behavior and the Priciple of Least Effort*. Addison-Wesley, Cambridge, MA (1949).