DO LIGHTNING DATA OBEY THE BENFORD'S LAW?

Pooyan Manoochehrnia[1], Farhad Rachidi[1], Marcos Rubinstein[2],
Gerhard Diendorfer[3], Wolfgang Schulz[3]

[1] Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland

[2] Western University of Applied Sciences, Yverdon, Switzerland

[3] ALDIS OVE, Vienna, Austria

E-mail (corresponding author): Farhad.Rachidi@epfl.ch

Abstract – Benford's law applies to a wide variety of natural and man-made data sets and it has been successfully applied to test the integrity of data and to detect possible fraud and anomalous results in economy, politics and management studies. In this paper, we investigate the applicability of Benford's law to lightning data. To do this, we used lightning data in Switzerland obtained using the European Cooperation of Lightning Detection (EUCLID) network. The considered data set consists of the total number of negative cloud-to-ground (CG) flashes per day in Switzerland for the period from 1999 till 2007. It is shown that the obtained distribution is in very good agreement with Benford's law. The same analysis was repeated considering in the data set only the flashes containing return strokes with absolute peak currents lower than 2 kA, for which the detection efficiency of the lightning location network is expected to be lower. The resulting distribution shows less agreement with the Benford's law, especially for the first, third and eighth digits, for which significant differences are obtained. The obtained results suggest that Benford's law may find an interesting application in the evaluation of detection efficiency of a given lightning location network.

1 INTRODUCTION

In this paper, we investigate the applicability of Benford's law to lightning data obtained using lightning location systems (LLS). Parameters that are generally derived from LLS observations are the ground flash density (GFD), flash multiplicity, polarity, and peak current distribution.

The paper is organized as follows: In Section 2, we present Benford's law and its applications. Section 3 describes briefly the lightning data set used in this study. Section 4 presents an application of Benford's law to the considered data set. Finally, the conclusions are given in Section 5.

2 BENFORD'S LAW

It was empirically observed that in many naturally occurring sets of data, the first significant digit[1] is not uniformly distributed as it might be expected. This was first observed in 1881 by Simon Newcomb [1] who noted that the first pages in his book of logarithm tables relating to numbers with low first digits showed more wear than the ending pages relating to numbers with high first digits. Later in 1938, Benford [2] hypothesized that the logs of low first digit numbers were looked up more often because there were more low first digit numbers in the world [3]. Benford tested his hypothesis by analyzing the first digits of numbers from 20 different tables such as the numbers appearing in a newspaper, physical constants, area of rivers, American league baseball statistics, population, addresses, etc. and found that the data are in remarkable fit with the following logarithmic law

$$\text{Prob}(\textit{first significant digit} = d) = \log_{10}(1 + \frac{1}{d}) \quad (1)$$

with d=1, 2, …, 9.

In other words, the first digit is 1 almost one third of the time, and larger digits occur with lower and lower frequency (9 as a first digit occurs less than 5%).

In 1961, Prinkham [4] proved that Benford's law was scale invariant. For example, if the digits of the lengths of all rivers conformed to Benford's law, then it should not matter whether they are measured in miles or kilometres [3]. He also showed that Bernford's law is the only

---

[1] The first significant digit of a number is the leftmost, nonzero digit of the number. For example the first significant digit of 32.865 is 3 and the first significant digit of 0.00462 is 4.

distribution which is invariant under scale change. In 1996, Hill [5] demonstrated Benford's law is scale and base invariant. He also derived a statistical limit law similar to the central-limit theorem but applied to the first digit.

This counterintuitive law applies to a wide variety of data sets and it has been successfully applied to test the integrity of data and to detect possible accounting fraud and anomalous results [3, 6].

3 DATA

The European Cooperation for Lightning Detection (EUCLID) has served as the database of Lightning statistics in Switzerland used in this study. The lightning statistic data were obtained in Switzerland using EUCLID data from 1999 to 2007 [7].

Fig. 1 shows the ground flash density (GFD) in Switzerland during an 8-year period from 1999 to 2006 based on a 1 km x 1 km grid.
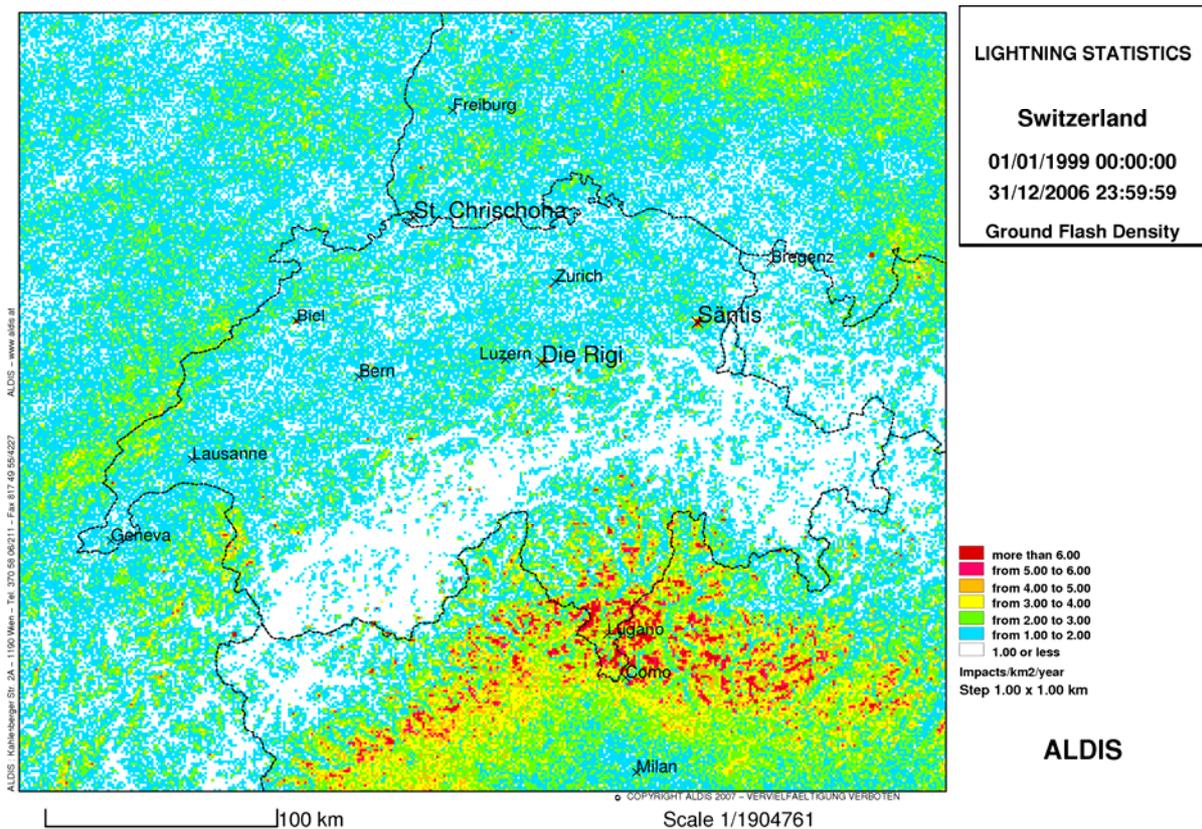

Fig. 1: Ground Flash density (Impacts.km^{-2}.year^{-1}), 1999-2006.
(white: less than 1, blue: 1-2, green: 2-3, yellow: 3-4, orange: 4-5, red: 5-6)
(Adapted from [7])

The GFD varies between 1 and 3 flashes km^{-2} yr^{-1} across most of the Swiss geographical regions. This value is quite similar to the GFD reported for Austria by Schulz et al. [8].

4 APPLICATION OF BENFORD'S LAW TO DATA FROM LIGHTNING LOCATION SYSTEMS

4.1 Application to negative cloud-to-ground flashes

The first considered data set consists of the number of negative cloud-to-ground (CG) flashes per day in Switzerland for the period from 1999 till 2007.

Fig. 2 presents the frequency of the first significant digit plotted against Benford's law. It can be seen from this graph that the obtained distribution is in very good agreement with Benford's law.

A chi-square goodness-of-fit test is also applied to quantify how well lightning statistics follow Benford's law. Results of Chi-statistics for the given dataset confirmed that the null hypothesis is acceptable with 5% significance level, indicating that the observed distribution is a random sample of Benford's distribution.
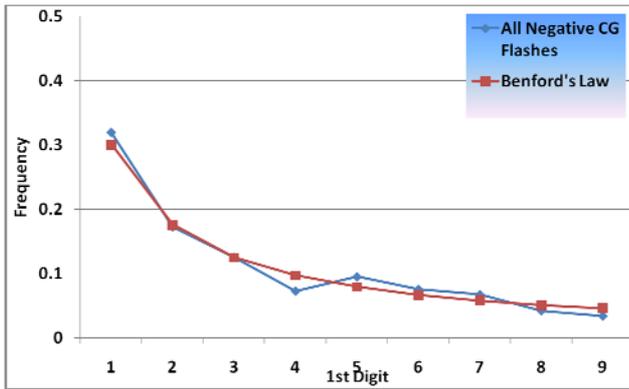
Fig. 2 - Frequency of the first significant digit associated with the cumulative number of negative cloud-to-ground flashes per day in Switzerland during the period 1999 to 2007.

4.2 Efficiency of LLS and Benford's law

Benford's law has been widely used for the detection of unnatural contaminations in various datasets. It is well known that the detection efficiency of LLS decreases significantly for flashes containing return-strokes with small current peaks [9]. We have repeated the same analysis presented in Sect. 4.1 considering in the dataset only the flashes containing return strokes with absolute peak currents less than 2 kA. The LLS is expected to be less efficient in the detection of these flashes. The results are shown in Fig. 3.
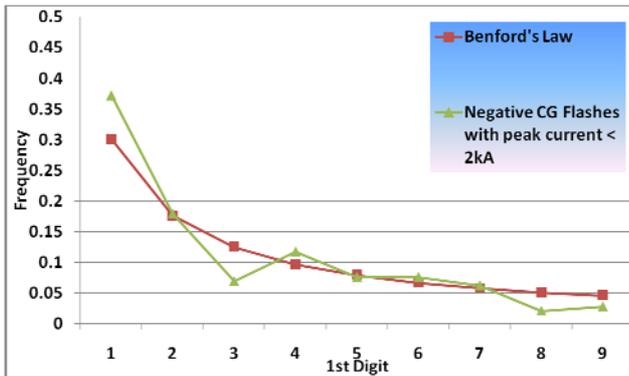

Fig. 3 - Frequency of the first significant digit associated with the cumulative number of less-than-2 kA (absolute peak value), negative cloud-to-ground flashes per day in Switzerland during the period 1999 to 2007.

It can be seen that the agreement with Benford's law worsens, especially for the first, third and eighth digits, for which significant differences are obtained. Applying a chi-square goodness-of-fit test to this dataset still does not reject the null hypothesis but the chi-statistics value (10.6) is closer to the limit of rejection (15.5).

5 CONCLUSIONS

Benford's law applies to a wide variety of natural and man-made data sets and it has been successfully applied to test the integrity of data and to detect possible fraud and anomalous results in economy, politics and management studies.

In this paper, we investigated the applicability of Benford's law to lightning data. To the best of the authors' knowledge, this is the first time that Benford's law is applied to lightning. To do this, we used lightning data in Switzerland obtained using the European Cooperation for Lightning Detection (EUCLID) network. The data set consisted of the number of negative cloud-to-ground (CG) flashes per day in Switzerland for the years from 1999 till 2007. It was shown that the obtained distribution is in very good agreement with Benford's law.

The same analysis was repeated considering in the data set only the flashes containing return strokes with absolute peak currents less than 2 kA. The lightning location network is expected to be less efficient in the detection of these flashes. It was shown that the new distribution's agreement with Benford's law worsens, especially for the first, third and eighth digits for which significant differences were observed.

The obtained results are promising because they suggest that Benford's law may find an interesting application in the evaluation of detection efficiency of a given lightning location network. Further research is ongoing in this respect.

6 ACKNOWLEDGMENTS

7 REFERENCES

[1] S. Newcomb, "Note on the frequency of use of the different digits in natural numbers," *American Journal of Matehmatics,* vol. 4, pp. 39-40, 1881.

[2] F. Benford, "The law of anomalous numbers," *Proc. Amer. Phil. Soc.,* vol. 78, pp. 551-572, 1938.

[3] M. Nigrini, "The peculiar patterns of first digits," *IEEE Potentials,* pp. 24-27, 1999.

[4] R. Pinkham, "On the Distribution of First Significant Digits," *Ann. of Mathematical Statistics,* pp. 1223-1230, 1961.

[5] T. Hill, "A statistical derivation of the significant-digit law," *Stat. Sci.,* vol. 10, pp. 354-363, 1996.

[6] M. Nigrini, "I've got your number," *J. Accountancy (http://www.aicpa.org/pubs/jofa/may1999/nigrini.htm),* 1999.

[7] P. Manoochehrnia, F. Rachidi, M. Rubinstein, and W. Schulz, "Lightning statistics in Switzerland," in *9th International Symposium on Lightning Protection (SIPDA)*, Foz do Iguaçu, Brazil, 2007.

[8] W. Schulz, K. L. Cummins, G. Diendorfer, and M. Dorninger, "Cloud-to-ground lightning in Austria: A 10-year study using data from a lightning location system," *Journal of Geophysical Research Letters,* vol. 110, D09101,doi:10.1029/2004JD005332, 2005.

[9] "CIGRE Working Group C4.404: Cloud-to-ground lightning parameters derived from lightning location systems," CIGRE Technical Brochure No 376, 2009.