MDAG.com Internet Case Study 54

MDAG-com Case Study 54 - Fun with Numbers: Benford's and Zipf's Laws at Work in Minesite Drainage

by K.A. Morin

www.mdag.com/case_studies/cs54.html

TABLE OF CONTENTS

Abstract

Benford's Law and Zipf's Law describe the statistical distributions expected, under certain conditions, for the first significant digit (first non-zero digit to the left) in values of widely varying datasets. In other words, the first significant digit can have a value from 1 up to 9. For example, the three concentrations of 21.5, 2.6, and 0.0298 mg/L all have "2" as their first non-zero, significant digit.

Full-scale-minesite Case Studies 1 and 3 of Morin (2016) provided more than 1000 values, up to nearly 6000 values, for flow and for each aqueous element and parameter. These values spanned many years to decades, and were collected from sampling locations across the entire minesites. Put simply, all the aqueous data for each of these two minesites were "lumped together" for each parameter. Aqueous pH was automatically ignored, because its values did not range from 1 through 9 at these minesites.

Despite the major spatial and temporal "lumping", the first-digit frequencies for flow, and for most aqueous elements frequently above detection, at each minesite did indeed resemble the frequency distributions provided by Benford's Law, and Zipf's Law with s = 1. Those that did not had their aqueous concentrations limited by known geochemical processes, primarily aqueous gypsum saturation. Such limitations prevent their first significant digits from ranging from 1 through 9.

While the correspondence to Benford's and Zipf's Laws is very "eye catching", it was really not surprising for these two Case Studies. This is because these laws can be followed by large datasets, such as geochemical analyses, that range widely across values, particularly large ranges of logarithmic values. Moreover, datasets that follow power laws and display scale invariance, which had been documented for these Case Studies up to decades ago, can also follow Benford's and Zipf's Laws.

Part of the explanation for these laws applying to some large ranges of logarithmic values lies in the logarithmic transformation itself. This transformation is similar to Benford's Law and to Zipf's Law with s = 1. While that may take some magic out of the results, it is not all that simple. There remain unexplained aspects for the frequent appearances of these two laws across many fields of science, mathematics, and art. Thus, some mathematicians still find Benford's Law to be "mysterious".

1. Introduction

Based on past decades of in-field monitoring (e.g., Morin et al, 1993, 1994, and 1995), Morin (2016) showed that full-scale drainage from highly reactive geologic materials and from minesite components:

- can display lognormal distributions and follow power laws when decades of monitoring data from many sampling locations at a single minesite were assembled into a single database; and

- can display scale invariance, fractal patterns, and 1-over-f slopes in high-frequency and long-term time series at individual sampling locations at a minesite.

This MDAG investigation carries the numerical analyses further, by comparing two large, full-scale minesite databases (Case Studies 1 and 3 of Morin, 2016) to Benford's Law (Wikipedia, 2018a) and Zipf's Law (Wikipedia, 2018b).

These two laws describe interesting statistical distributions of the first significant digit (first non-zero digit to the left). For example, the three concentrations of 21.5, 2.6, and 0.0298 mg/L all have "2" as their first non-zero, significant digit.

While one might initially expect all first non-zero digits from 1 to 9 have an equal probability of occurring, this is not always so. Benford's Law indicates the probability of the first digit being 1 is 0.301 or 30.1%, the probability of the first digit being 2 is 0.176, etc., with the probability of the first digit being 9 at only 0.046. Numerically, Benford's Law is:

$$\text{Probability of the first digit being n (where n is 1 through 9)} = \log_{10}[1 + 1/n)] \qquad \text{(Eq. 1-1)}$$

Zipf's Law also addresses the probabilities of the first significant digits, but includes an exponent (s) that leads to additional distributions. When s = 1, the probabilities from Benford's and Zipf's Laws are close. Numerically, Zipf's Law for the first non-zero digit is:

$$\text{Probability of the first digit being n (when n is 1 through 9)} = \{[(1/n^s) / [(1/1^s) + (1/2^s) + (1/3^s) + (1/4^s) + (1/5^s) + (1/6^s) + (1/7^s) + (1/8^s) + (1/9^s)]\} \qquad \text{(Eq. 1-2)}$$

For comparison to Case Studies 1 and 3 of Morin (2016), all flows and aqueous concentrations were truncated, not rounded, to extract the first non-zero digit. All values below the human artifacts of detection limits were deleted. Aqueous elements whose concentrations were typically or consistently below detection were deleted entirely.

Also, pH was deleted from consideration. This is because pH at these sites predominantly ranged from roughly 2 to 8. Over this range that does not include 1 or 9, Benford's and Zipf's Laws could not possibly apply.

The result was first digits for flow and many aqueous elements, based on at least 1000 to nearly 6000 analyses each (Section 2).

2. Examples with More Than a Thousand Analyses at Full-Scale Minesites

As explained in Section 1, the first significant digits for measured values of full-scale flows and aqueous concentrations from Case Studies 1 and 3 of Morin (2016) were compiled. For each parameter, there were more than 1000 values, up to nearly 6000 analyses.

For Case Study 1 (Figures 2-1a and 2-1b) and Case Study 3 (Figures 2-2a and 2-2b), many histograms generally followed both Benford's Law (Equation 1-1) and Zipf's Law with s = 1 (Equation 1-2). Sulphate for Case Study 3 (Figure 2-2a) was better matched by Zipf's Law with s = 2. While this may seem remarkable, there are some reasons why this should be expected, although it remains "mysterious" in some ways (Section 3).

Some exceptions, where some first digits were under-represented, included:

- conductivity (e.g., digit "5"), acidity (e.g., digit "1"), sulphate, calcium, and magnesium for Case Study 1, and
- hardness for Case Study 3.

Many of these are related by gypsum solubility at these two minesites. Mineral saturation with gypsum at these sites limited the aqueous variability of sulphate. In turn, this limited the variabilities of conductivity, calcium, magnesium, hardness (which is basically calcium plus magnesium), and, in some cases, acidity.

As explained in Section 1, pH was dismissed, because it was known in advance that pH varied across less than 1 to 9. Similarly, these linked elements and parameters could not follow Benford's and Zipf's Laws, because their potentials to attain a first significant digit of 1 through 9 were limited by geochemical processes. This is a hint at some characteristics of datasets needed to follow these laws (Section 3).

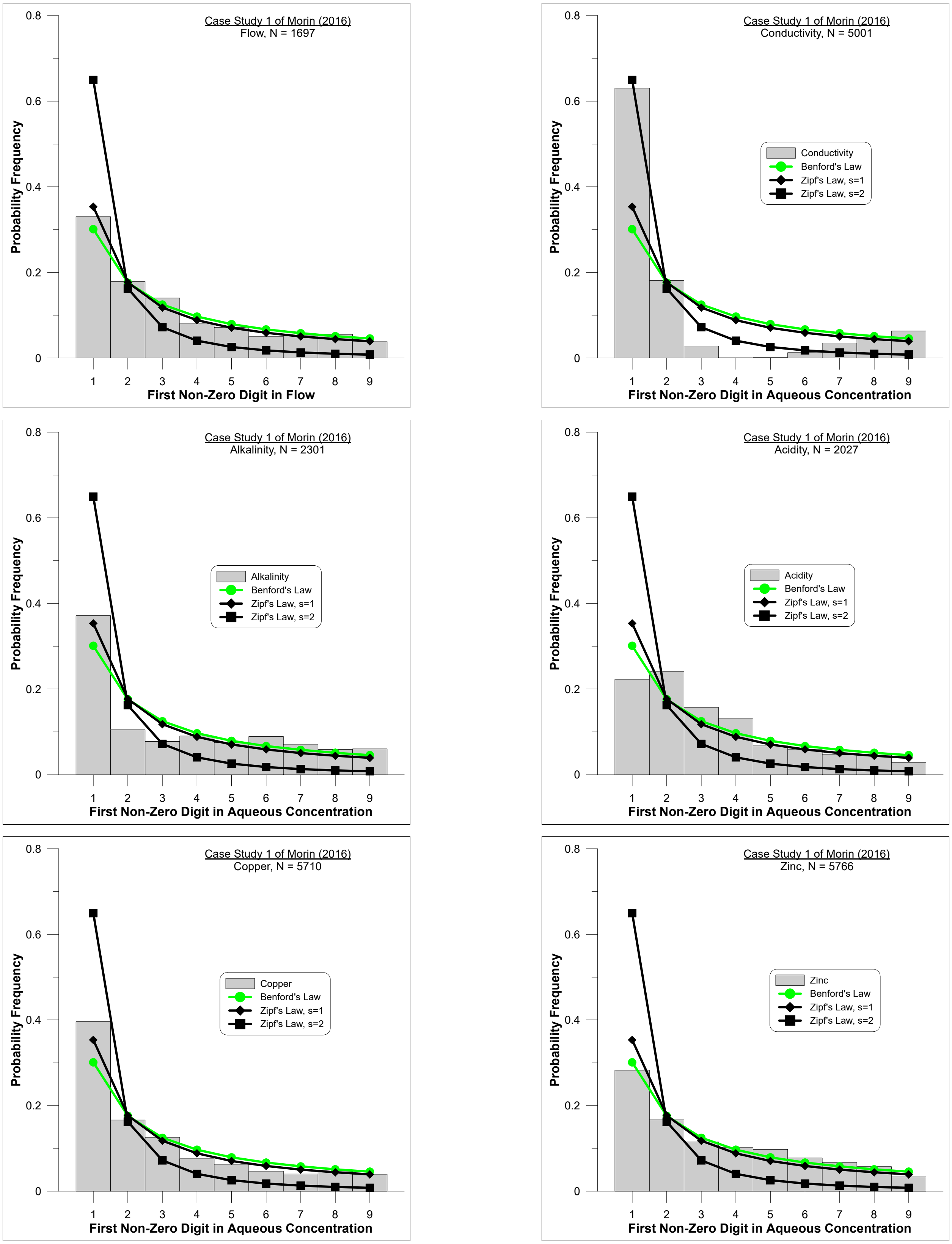

Figure 2-1a. Frequency histograms of the first non-zero digit in measured values of full-scale flow, electrical conductivity, alkalinity, acidity, copper, and zinc from Case Study 1 of Morin (2016), with overlaid plots of Benford's and Zipf's Laws.

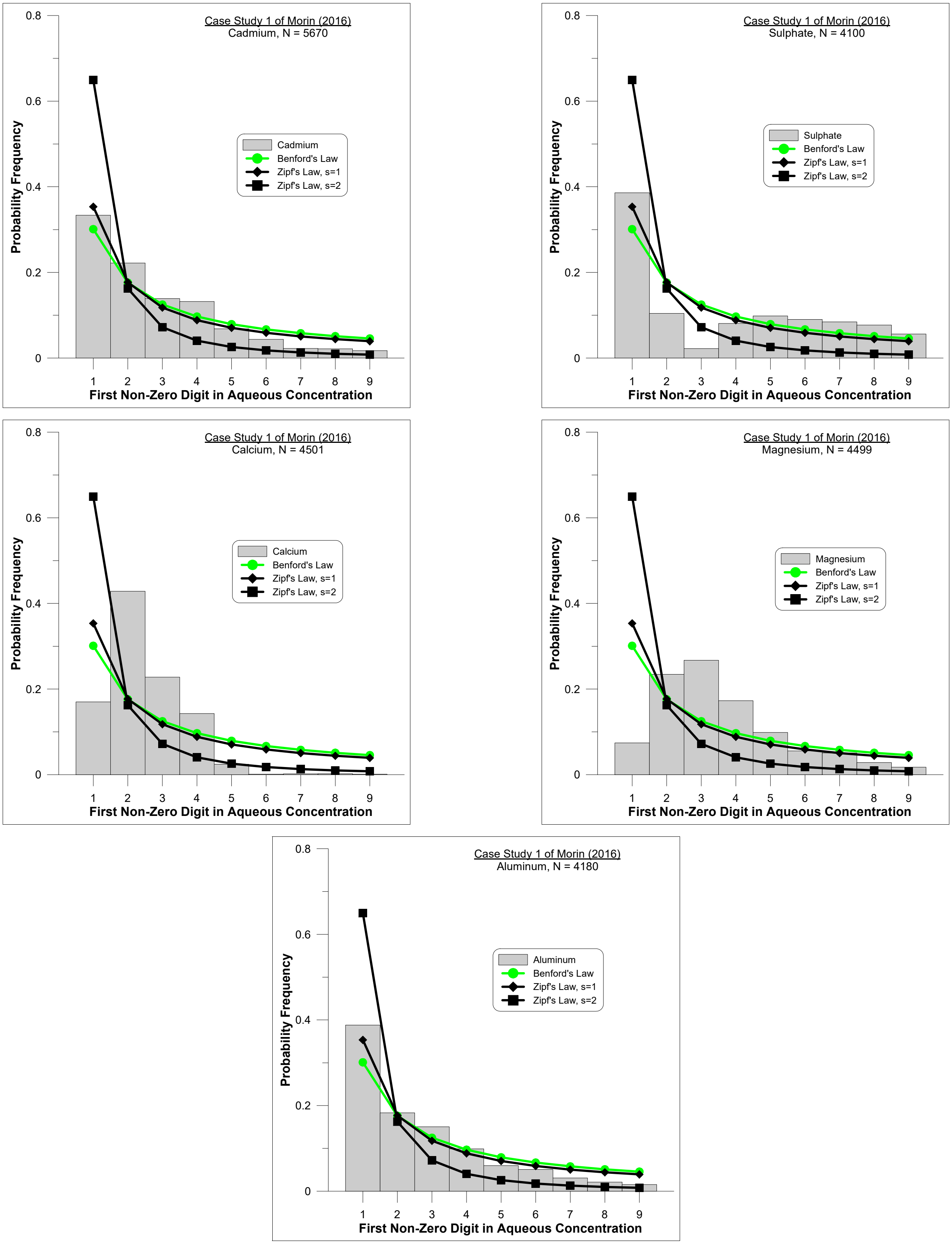

Figure 2-1b. Frequency histograms of the first non-zero digit in measured values of full-scale cadmium, sulphate, calcium, magnesium, and aluminum from Case Study 1 of Morin (2016), with overlaid plots of Benford's and Zipf's Laws.

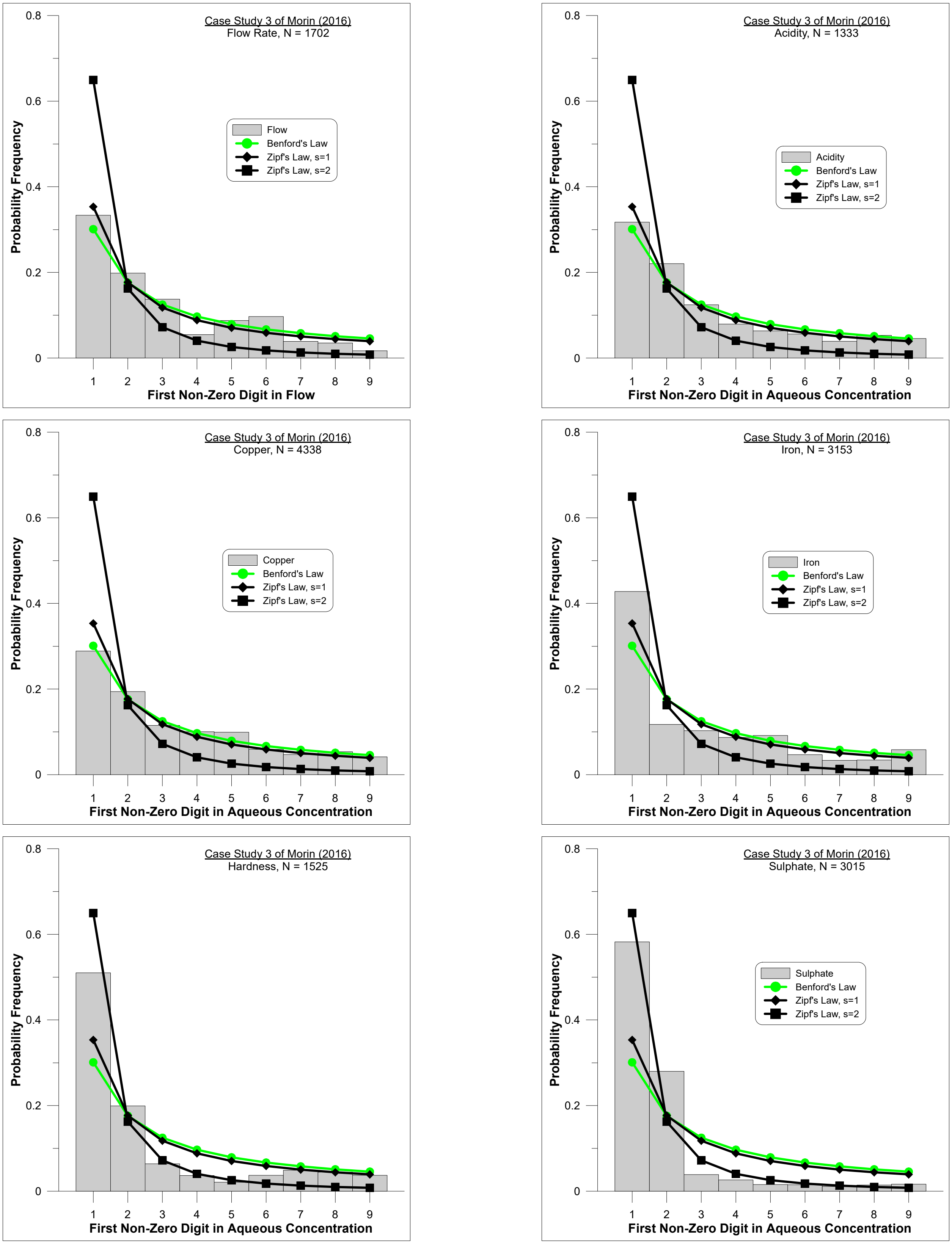

Figure 2-2a. Frequency histograms of the first non-zero digit in measured values of full-scale flow, acidity, copper, iron, hardness, and sulphate from Case Study 3 of Morin (2016), with overlaid plots of Benford's and Zipf's Laws.

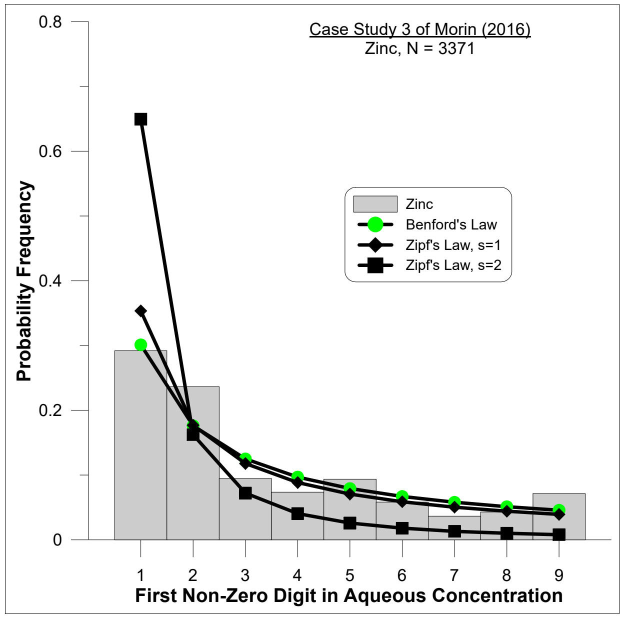

Figure 2-2b. Frequency histograms of the first non-zero digit in measured values of full-scale zinc from Case Study 3 of Morin (2016), with overlaid plots of Benford's and Zipf's Laws.

3. Conclusion

In reality, many diagrams in Figures 2-1 and 2-2 should generally resemble Benford's Law and Zipf's Law with s = 1. This is because these laws can be approximated by large datasets, such as geochemical analyses, that range widely across values, particularly large ranges of logarithmic values. Moreover, datasets that follow power laws and display scale invariance can follow these laws. Power laws and scale invariance had been documented for these Case Studies (Section 1) up to decades ago.

Part of the explanation for these laws applying to some large ranges of logarithmic values lies in the logarithmic transformation itself. This transformation is, in fact, similar to Benford's Law and Zipf's Law with s = 1 (Table 3-1).

While that may take some magic out of the results, it is not all that simple. Berger and Hill (2011) explained that a dataset with a large range of logarithmic values can resemble the laws, but that is not sufficient. It can also depend on aspects such as the base of the logarithm, the number of values, and the statistical distribution within the range.

> "Although many facets of BL [Benfords's Law] now rest on solid ground, there is currently no unified approach that simultaneously explains its appearance in dynamical systems, number theory, statistics, and real-world data. In that sense, most experts seem to agree . . . that the ubiquity of BL, especially in real-life data, remains mysterious." (Berger and Hill, 2011)

In some ways, the frequent appearances of Benford's Law, and of 1-over-f slopes, across many fields of science, mathematics, and art are similarly interesting and "mysterious".

Table 3-1. Comparison of frequency probabilities for digits 1 through 9 from Benford's Law, Zipf's Law, and log10 transformation

	Frequency Probability based on . . .		
Digit	Benford's Law	Zipf's Law with s = 1	log10 Transformation
1	0.30	0.35	0.30
2	0.18	0.18	0.18
3	0.13	0.12	0.12
4	0.097	0.088	0.097
5	0.079	0.071	0.079
6	0.067	0.059	0.067
7	0.058	0.051	0.058
8	0.051	0.044	0.051
9	0.046	0.039	0.046

4. References

Berger, A., and T.P. Hill. 2011. Benford's Law Strikes Back: No Simple Explanation in Sight for Mathematical Gem. The Mathematical Intelligencer, 33. Available at: works.bepress.com/tphill/74/

Morin, K.A. 2016. *Spectral Analysis of Drainage from Highly Reactive Geologic Materials*. ISBN 978-0-9952149-1-0. Free ebook available at: http://www.mdag.com/spectral-book.html

Morin, K.A., N.M. Hutt, and I.A. Horne. 1995. Prediction of future water chemistry from Island Copper Mine's On-Land Dumps. IN: Proceedings of the 19th Annual British Columbia Mine Reclamation Symposium, Dawson Creek, British Columbia, Canada, June 19-23, p. 224-233.

Morin, K.A., I.A. Horne, and D. Riehm. 1994. High-frequency geochemical monitoring of toe seepage from mine-rock dumps, BHP Minerals' Island Copper Mine, British Columbia. IN: Proceedings of the Third International Conference on the Abatement of Acidic Drainage, Pittsburgh, Pennsylvania, USA, April 24-29, Volume 1, p.346-354

Morin, K.A., N.M. Hutt, and D. Flather. 1993. The appropriate geochemical monitoring of toe seepage from a mine-rock dump. IN: Proceedings of the 17th Annual Mine Reclamation Symposium, Port Hardy, British Columbia, May 4-7, p.119-129. Mining Association of British Columbia.

Wikipedia.org. 2018a. Benford's Law. Accessed March 2018 at: en.wikipedia.org/wiki/Benford%27s_law

Wikipedia.org. 2018b. Zipf's Law. Accessed March 2018 at: en.wikipedia.org/wiki/Zipf%27s_law