Civil Engineering Infrastructures Journal, 52(2): 323 – 334, December 2019
Print ISSN: 2322-2093; Online ISSN: 2423-6691
DOI: 10.22059/ceij.2019.272005.1534

Application of Benford's Law in Analyzing Geotechnical Data

Alipour, A.[1*] and Alipour, S.[2]

[1] Assistant Professor, Department of Mining and Metallurgical Engineering, Urmia University of Technology, Urmia, West Azerbaijan, Iran.
[2] Assistant Professor, School of Computer Science, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran.

Received: 21 Dec. 2018; Revised: 20 Oct. 2019; Accepted: 20 Oct. 2019

ABSTRACT: Benford's law predicts the frequency of the first digit of numbers met in a wide range of naturally occurring phenomena. In data sets, following Benford's law, numbers are started with a small leading digit more often than those with a large leading digit. This law can be used as a tool for detecting fraud and abnormally in the number sets and any fabricated number sets. This can be used as an effective tool for processing data sets from laboratory tests, site investigation tests data, geotechnical design (both financial and technical) in engineering and specially included in geotechnical engineering and, etc. In this paper, data sets from geotechnical data are gathered and analyzed. It is shown that most of them follow Benford's law. Therefore, we can use this observation for similar applications to detect the validity of data. Also, this can be assumed as an evidence for natural numbers that follow Benford's law.

Keywords: Benford's law, Data Processing, Geotechnical Data.

INTRODUCTION

Geotechnical engineering is incremented for recognizing and analyzing earth from engineering view. Phenomena in nature are illustrated by their quantity or quality properties. Because of the precise scales and standards of illustrations based on quantity, it has many proponents in engineering; numbers are known as indicator of quantity illustration. Note that obtaining geotechnical data is not easy and possible all the times and this leads the geotechnical engineers to model the data based on some parameters and formulas.

Analyzing geotechnical data obtained from laboratory studies and site investigations helps us to study the earth from the engineering view.

There are situations where some part of the data is missed or the data is noisy or the modeled data does not show the behavior of real phenomena. Thus, it is important to determine the correctness of obtained data or the modeled data. In this paper we have shown that some geotechnical datasets have Benford's law behavior and suggest this as a tool for examining the validity of geotechnical data sets.

There are more than 200 publications about Benford's law in the past years, a law found on the frequency of a particular digit occurs in a specific position in numbers (Nigrini, 1999). For example numerous data

* Corresponding author E-mail: aref.alipour@gmail.com

sets from the physical sciences also have been shown to follow Benford, including direct measurement of some physical, geophysical and astronomical quantities. These are described in (Sambridge et al., 2010).

BENFORD'S LAW

In 1881, a 2-page article was published in the American Journal of Mathematics by Newcomb. He described that books of logarithms were dirtier in the beginning and progressively cleaner throughout (Newcomb, 1881). From this, he concluded that scientists that use the logarithm tables were looking for numbers that start with first digit 1 more than numbers starting with 2, and numbers with first digit 2 more often than 3, and so forth. Newcomb stated that the probability that the first non-zero digit of a number is d, *p (d)*, is:

$$p(d) = \log(1+(1/d)); d = 1, 2, \ldots 9 \qquad (1)$$

Table 1, shows value of $p(d) = \log_{10}(1+\frac{1}{d})$ for d = 1,..., 9.

Newcomb's observation was unnoticed. Until 57 years later Frank Benford who was a general electric physicist, concluded the same logarithmic law from his observation about logarithm books (Benford, 1938). But, Benford examine his conjecture on many various types of data from various fields (Table 2).

His resulting table of first significant digits (the number to the left of the decimal point in "scientific notation" fits the logarithmic law exceedingly well. Unlike Newcomb's article, Benford's paper received a lot of attention and Newcomb's contribution became forgotten and the law came to be noticed as Benford's law. There are also generalizations of Benford's law Table 3.

As an example in the general law, the probability that the first three significant digits are 3, 1, and 4, in that order, is $\log_{10}(1+\frac{1}{314}) \cong 0.0014$ and similarly for other significant-digit patterns. Formulas for expected digital frequencies are given as follows.

For first digit of the number:

$$p(d_1) = \log(1+(1/d_1)); d_1 = 1, 2, \ldots 9 \qquad (2)$$

The probability that the second significant digit of a number is d_2, p(d_2), is:

$$p(d_2) = \sum_{d_1=1}^{9} \log(1+(1/\overline{d_1 d_2})); d_2 = 0, 1, 2 \ldots 9 \qquad (3)$$

For two digit combinations:
The probability that the first significant digit of a number is d1 and the second significant digit is d_2, $p(\overline{d_1 d_2})$, is:

$$p(\overline{d_1 d_2}) = \log(1+(1/\overline{d_1 d_2})) \qquad (4)$$

Pinkham (1961) observed that the law is scale-invariant, it does not matter if stock market prices are changed from dollars to pesos, the distribution pattern of significant digits remains the same. Hill (1995) discovered Benford's law is also independent of base, the law holds true for base 2 or base 7. Hill also showed that random samples from randomly chosen various distributions will follow Benford's law.

Benford's law is used in various numbers of areas, such as mathematical modeling and computer design. Of course, many data sets do not follow this logarithmic distribution. However, a surprisingly diverse collection of empirical data follows the law.

The assumption of logarithmically-distributed significant-digits (i.e., floating point numbers) in scientific calculations is widely used and well-established (Knuth et al., 2004). Ley (1996) showed that stock market figures fit with Benford's law closely. Nigrini (1996) showed that in the 1990 U.S. Census, the populations of the three thousand counties in the U.S. follow Benford's law.

In the 80 years, since Benford's article appeared, mathematicians, physicists, statisticians and amateurs tried to prove Benford's law, but there have been two main stumbling blocks.

Table 1. Value of $p(d)=\log_{10}(1+\frac{1}{d})$ for d=1,..,9

First digit	D	1	2	3	4	5	6	7	8	9
Newcomb's formula	$\log_{10}(1+\frac{1}{d})$	30.1	17.6	12.5	9.7	7.9	6.7	5.8	5.1	4.5

Table 2. The distribution of first digits as compiled by Benford in his original paper

Col.	Title	1	2	3	4	5	6	7	8	9	MAE	Samples
1	Rivers, Area	31.0	16.4	10.7	11.3	7.2	8.6	5.5	4.2	5.1	1.1	335
2	Population	33.9	20.4	14.2	8.1	7.2	6.2	4.1	3.7	2.2	1.8	3259
3	Constants	41.3	14.4	4.8	8.6	10.6	5.8	1.0	2.9	10.6	4.4	104
4	Newspapers	30.0	18.0	12.0	10.0	8.0	6.0	6.0	5.0	5.0	0.3	100
5	Specific Heat	24.0	18.4	16.2	14.6	10.6	4.1	3.2	4.8	4.1	2.6	1389
6	Pressure	29.6	18.3	12.8	9.8	8.3	6.4	5.7	4.4	4.7	0.3	703
7	H.P. Lost	30.0	18.4	11.9	10.8	8.1	7.0	5.1	5.1	3.6	0.5	690
8	Mol. Wgt.	26.7	25.2	15.4	10.8	6.7	5.1	4.1	2.8	3.2	2.5	1800
9	Drainage	27.1	23.9	13.8	12.6	8.2	5.0	5.0	2.5	1.9	2.4	159
10	Atomic Wgt.	47.2	18.7	5.5	4.4	6.6	4.4	3.3	4.4	5.5	4.2	91
11	$n^{-1},\sqrt{n}$	25.7	20.3	9.7	6.8	6.6	6.8	7.2	8.0	8.9	2.5	5000
12	Design	26.8	14.8	14.3	7.5	8.3	8.4	7.0	7.3	5.6	1.8	560
13	Cost Data	32.4	18.8	10.1	10.1	9.8	5.5	4.7	5.5	3.1	1.3	741
14	X-Ray Volts	27.9	17.5	14.4	9.0	8.1	7.4	5.1	5.8	4.8	0.8	707
15	Am. League	32.7	17.6	12.6	9.8	7.4	6.4	4.9	5.6	3.0	0.7	1458
16	Addresses	28.9	19.2	12.6	8.8	8.5	6.4	5.6	5.0	5.0	0.6	342
17	$n^1, n^2 ... n!$	25.3	16.0	12.0	10.0	8.5	8.8	6.8	7.1	5.5	1.5	900
18	Death rate	27.0	18.6	15.7	9.4	6.7	6.5	7.2	4.8	4.1	1.2	418
	Average	30.6	18.5	12.4	9.4	8.0	6.4	5.1	4.9	4.7	0.3	1011
	Probable error	± 0.8	± 0.4	± 0.4	± 0.3	± 0.2	± 0.2	± 0.2	± 0.3	± 0.2		

Table 3. Expected frequencies based on Benford's law

Digit	1st place	2nd place	3rd place	4th place
0		0.11968	0.10178	0.10018
1	0.30103	0.11389	0.10138	0.10014
2	0.17609	0.19882	0.10097	0.10010
3	0.12494	0.10433	0.10057	0.10006
4	0.09691	0.10031	0.10018	0.10002
5	0.07918	0.09668	0.09979	0.09998
6	0.06695	0.09337	0.09940	0.09994
7	0.05799	0.09035	0.09902	0.09990
8	0.05115	0.08757	0.09864	0.09986
9	0.04576	0.08500	0.09827	0.09982

The first is very simple -some data sets satisfy the law, and some do not, and there never was a clear definition of a general statistical experiment which would predict which tables would, and which would not. Instead, researchers endeavored to prove that the log law “is a built-in characteristic of our number system;” that is, to show that the set of all numbers follows the Benford's law, and then to suggest that this somehow explains the frequent empirical evidence. For more information see (Fewster, 2009) which contains plausible explanation of why Benford’s law arises in data sets.

Applications of Benford’s law

One of the applications of the significant-digit law is testing mathematical models. Suppose that a new model is proposed to predict future stock indices, census data, or computer usage. If current data follows Benford's law closely or if a hypothesis of unbiased random samples from random distributions seems reasonable, then the predicted data should also follow Benford's law closely (or else perhaps the model should be replaced by one which does). Note that the law says nothing about the raw data itself; e.g., in Benford's law there is not any difference between 20 and 200,000- both have first significant digit 2, and other digits 0. Another application which has been studied is designing of computers. If future users are likely to be performing the calculations taken from many distributions which are unbiased random, as Knuth et al. (2004) and other computer scientists claim is the case today, then their floating-point calculations will be based on data which closely follows Benford's law. Specifically, the numbers will not be uniformly distributed over the floating point numbers, but will rather obey the log law. If this is true, then it is possible to design computers whose designs is based on knowing the distribution of numbers. If 9's are much less frequent than 1's (or the analog for whatever base the computer is using), then it should be possible to design computers which use that information to minimize storage space, or to maximize rate of output printed. A development in the field of accounting is using Benford's law to detect fabrication of data in financial documents or fraud. Nigrini (1999) has found evidence of the occurrence of Benford's law in many areas of demographic data and accounting. He concluded that in a wide variety of accounting situations, the significant digit frequencies of true data follow Benford's law. Benford's law has been extensively used in different kinds of engineering problems.

For example this method has been used for detecting problems in survey data (Judge and Schechter, 2009), Image processing (Pérez-González et al., 2007), fraud detecting in supply chain data (Hales et al., 2009) and analysis of leading digits frequencies in thermal conductivities of liquids data (Whyman et al., 2016). Arshadi and Jahangir (2014) showed the Benford’s law behavior of internet traffic. Joannes-Boyau et al. (2015) used Benford’s law to find natural hazard dataset homogeneity. There are many other recent publication of Benford’s law (e.g. Alves et al., 2016; Wei et al., 2017; Tseng et al., 2017; Nigrini and Wells, 2012). Benford's appearance in numerous areas can be used to detect the onset of an earthquake in seismic data (Sambridge et al., 2010). Benford’s law has role in investigating homogeneity in natural hazard datasets (Joannes-Boyau et al., 2015) and detecting change in a physical phenomenon (Sambridge et al., 2011). Benford’s law was also applied to volcanology (Geyer and Martí, 2012). Besides, the ability of Benford’s law was studied to detect earthquakes (Díaz et al., 2014).

Benford’s law and Geotechnical Data

The engineering properties of soil and rock exhibit depend on uncertain behavior due to the complex and imprecise physical processes associated with the formation of these materials. This is in contrast to the most other civil engineering materials, such as steel, concrete and timber, which exhibit far greater homogeneity and isotropy. Measuring data and analyzing them is important for geotechnical analysis in order to cope with the complexity of geotechnical behaviors, and the spatial variability of these materials, such as: tunneling, slop stability analysis, foundation designing, etc. Thus, designing based on real data is very important. There are situations where the gathered data is uncertain or some part of it is missed. Since gathering geotechnical is hard, for analyzing proposes the data are also modeled in laboratories. In these situations it is important to use valid data. Validity of data is important for any geotechnical analysis (Barkhordari and Entezari Zarch, 2015; Kohestani et al., 2017; Shahraki et al., 2018; Alipour et al., 2012).

In this paper, Benford's law is applied for 10 number sets caught from geotechnical data. We introduce each data set and the methods using for measuring them, some of these data caught by experiment directly and some of them calculated by formulas, in which we measure parameters and calculate them using these formulas specifying their relation; of course the parameters are measured by experimental methods. Note that the data or parameters obtained by experimental calculations may have error and this can affect our results.

INTRODUCTION OF CASE STUDIES

Boreability Index ($(kN / cutter / mm / rev)$)

Boreability is an index to present the difficulty or ease that a rock mass can be penetrated by a tunnel boring machine (Yin et al., 2014). The rock mass boreability index decreases when the penetration rate increases. The index depends on the rock mass conditions and cutterhead design, especially the cutter diameter, cutter spacing and cutter tip width. Consequently, BI is measures the rock mass boreability. It is normal used to analyze the variation of rock mass conditions under the same machine operation parameters (Gong et al., 2007).

Maximum Surface Settlement Caused by Tunneling (mm)

There are numerous empirical and analytical relations between shield tunnel characteristics and surface and subsurface deformation. In the Bangkok MRTA project, data on ground deformation and shield operation are collected. Large numbers of surface settlement markers and settlement arrays are installed to measure surface settlements during excavation and these data have been used in (Suwansawat and Einstein, 2006).

P-Wave Velocity in Rock (m/s)

P-waves are a kind of elastic waves, also called seismic waves, which can travel through gasses (such as sounds), elastic solids and liquids, including the rock (Alipour et al., 2012). The P-wave velocity in rock varies between 1000 to 8000 m/s. Rock samples are collected from geotechnical project for laboratory tests. P-wave velocity values that we analyze them, are determined in the laboratory for the rock and soil dynamic analysis (Chakraborty et al., 2004).

Point Load Strength Index of Rock (MPa)

A fast and convenient way for the strength and fracture toughness detection of an ore is point load test. The Point Load Index can also be used to simultaneously characterize rock for blastability and comminution processes. In a Point Load Test, a sample of rock is mounted between two pointed platens. Then pressure is applied till failure of the sample occurs. The peak applied load is recorded to calculate the Point Load Index. The units of the point load index are *MPa*. The results of the test can be influenced by the size and shape of each particle tested in a Point Load Test. The best known possible shape for the highest accuracy is the cylindrical, such as core. Data obtained from (Broch and Franklin, 1972) are used in this paper.

Shear Strength of Rock (MPa)

A stress state in which the shape of a material starts to change is called shear stress in physics (usually by "sliding" forces - torque by transversely-acting forces) without particular volume change. One can determine the shear strength of rocks is by testing on various rocks. This parameter is the base of many rock mechanics analyses. The rock mechanics literature is rich with a number of shear failure criteria that have been developed. Among these criteria, the most commonly used is the Mohr–Coulomb criterion. Coulomb introduced the most important and simplest failure criterion in 1776. His suggestion was that rock failure would take place in the presence of shear stress; that is extended when a specific plane would reach a value that is sufficient to overcome both the natural cohesion of the rock, plus the frictional force that opposes motion along the failure plane. The criterion can be written as (Al-Ajmi and Zimmerman, 2005; Mogi, 1971):

$$\tau = c + \sigma_n \tan\phi \qquad (5)$$

where σ_n: is the normal stress acting on the failure plane, c: is the cohesion of the material and φ: is the angle of internal friction. Mogi (1971) and Al-Ajmi and Zimmerman (2005) studied shear strength data of some kinds of rocks in experimental test by using the Benford's law.

Shaft Resistance of Piles in Rock (MPa)

In this paper, a rational calculation procedure is introduced to establish the shaft resistance of a pile embedded in rocks, The shaft resistance of the pile can be represented as a relation among mean shear stress throughout the shaft in the rock (τ_m) and some other parameters (Serrano and Olalla, 2006). The variable τ_m which is determined by experiment is used in this paper.

Specific Gravity of Rock (Dimension Less)

Specific gravity is described as the ratio of the density of a given solid or liquid substance to the density of water at a specific temperature and pressure (Alipour et al., 2018). Substances with specific gravity greater than one are denser than water and will sink in it. Specific gravity, SG, is expressed mathematically as:

$$SG = \frac{\rho_{Rock}}{\rho_{H_2O}} \qquad (6)$$

where ρ_{Rock}: is the substance density, and ρ_{H_2O}: is the water density. Specific gravity is dimensionless. Here some samples of rock's specific gravities are obtained from (Chakraborty et al., 2004).

Swelling Pressure of Mudrock (kPa)

After decreasing of the moisture in soil or rock, air enters the pores and when the moisture returns by capillarity, air is captured and its pressure increases in internal pores. This phenomenon can be an important disaggregation mechanism since pressure development can breakdown the integrity of sedimentary rocks (Doostmohammadi, 2016). To increase the level of structural safety we need accurate prediction of swelling pressure. A number of total pressure cells between concrete and shotcrete walls of the powerhouse cavern at Masjed–Soleiman hydroelectric powerhouse project which is in the south of Iran, where mud rock outcrops, confirmed a cyclic swelling pressure on the lining since 1999. In several locations, small cracks are generated which has raised doubts about long term stability of the powerhouse structure. Swelling pressure of mud rock is provided from (Moosavi et al., 2006).

Tunnel Convergence (cm)

A tunnel is an underground passageway which may be used for pedestrians or cyclists, general road traffic, motor vehicles only, rail traffic, or even a canal. A reliable nomination of the extension of tunnel convergence is essential, so that some strategy can be established regarding to stabilize measures and to optimize the support well in advance (during the planning and designing stage) (Ariznavarreta-Fernández et al., 2016). Over 214 convergence measurement stations along Kaligandaki tunnel is obtained. The measured horizontal convergence of this tunnel section is recorded (Panthi and Nilsen, 2007).

Tensile Stress in Lining (MPa)

Tensile strength of a material is the maximum amount of the tensile stress that can be subjected to prior failure. Concrete lining has tensile strength. In the case of the Trasvasur tunnels, tensile stress in lining was measured (Pérez-Romero et al., 2007).

BENFORD'S LAW AND OUR CASE STUDIES

In order to apply Benford's law on these data sets we calculate the frequency of each leading digit in a set and calculate the distribution of numbers with leading digit 1, 2… 9. Table 4 demonstrates the frequency of each set for each leading digit. The results of this method were acceptable and interesting.

We draw the graph of each set and also Benford's distribution to compare each set to Benford's one. In each graph the percentage of each leading number which we calculated and Benford's one are compared to show visually a data set whether is close to Benford's distribution or not. It seems that convergence and swelling pressure are similar to Benford's distribution and point load and shear stress have least similarity to Benford's distribution Graph, Figure 1a-i.

We calculate Mean Absolute Error (MAE) for each set that is defined as:

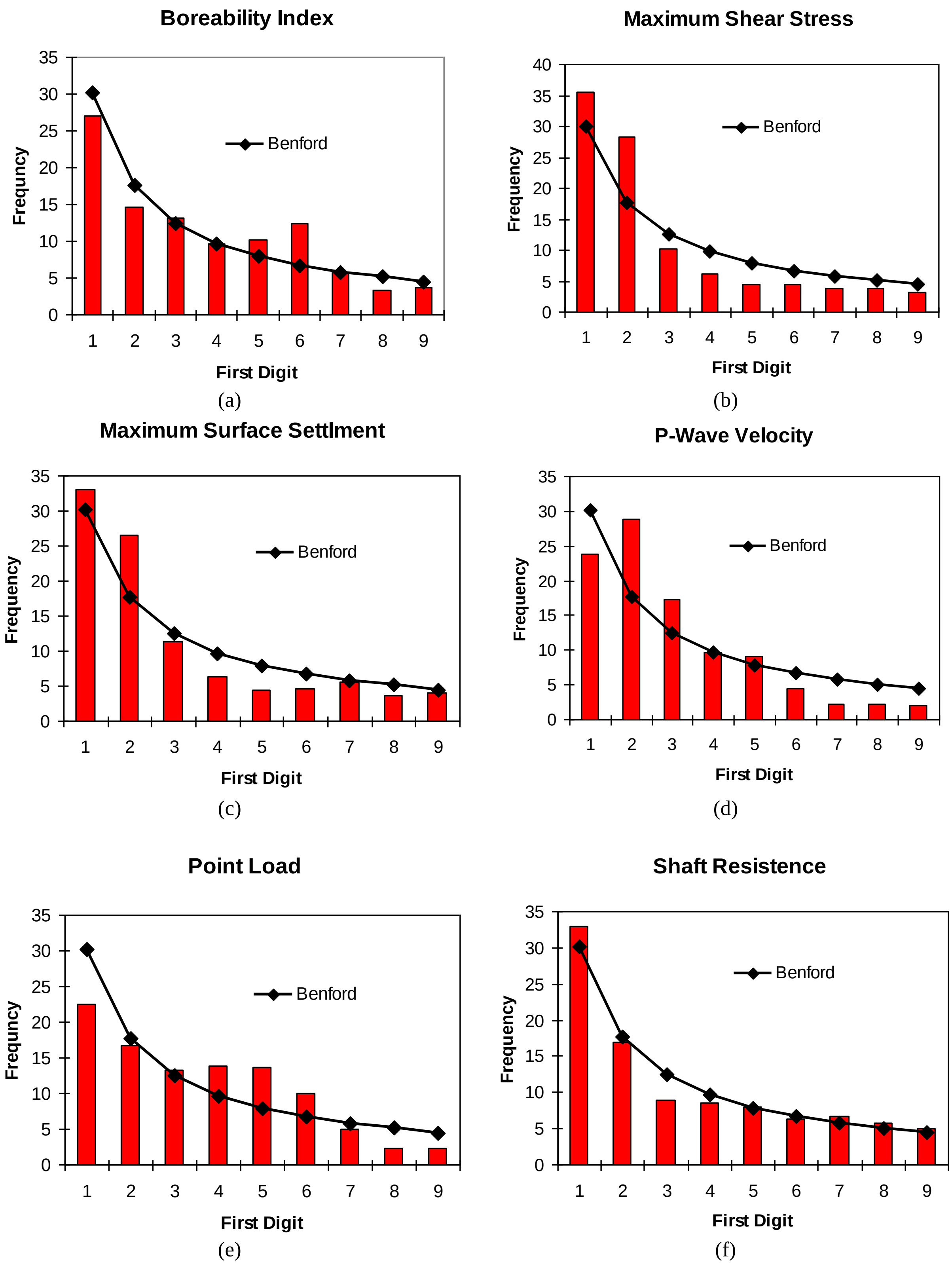
$$MAE = \frac{1}{9}\sum_{i=1}^{9} | b_i - e_i | \quad (7)$$

where e_i : is the observed frequency in each bin in the empirical data and b_i : is the frequency expected by Benford. Note that higher values of MAE imply more error. In addition to merely visually reviewing the data, we use a chi-square χ^2 goodness of-fit test to check the extension to which the data coincide to Benford's law. The χ^2 statistic is calculated as Eq. (8):

$$\chi^2 = \sum_{i=1}^{9} \frac{(e_i - b_i)^2}{b_i} \quad (8)$$

Table 4. Distribution of the first digit of data sets

	Title	1	2	3	4	5	6	7	8	9	Number of samples
1	Boreability index	27	14.6	13.2	9.6	10.2	12.3	5.7	3.4	3.8	41
2	Maximum shear stress of rock	35.6	28.2	10.2	6.1	4.4	4.4	3.9	3.8	3.2	138
3	Maximum surface settlement caused by EPB shield tunneling	33	26.6	11.4	6.4	4.4	4.6	5.6	3.6	4	49
4	P- wave velocity in rock	23.8	28.8	17.3	9.6	9.1	4.5	2.3	2.2	2.1	45
5	Point load strength index	22.5	16.8	13.2	13.9	13.5	9.9	5	2.3	2.4	79
6	Shaft Resistance of piles in rock	33	17	9	8.6	8	6.4	6.7	5.8	5	79
7	Specific gravity of rock	43.8	14.4	9.8	6.9	8.8	4	1.6	4.1	6.4	45
8	Swelling pressure of mud rock	32	16.7	11.2	9.1	7.8	7.7	6.5	6.2	5.1	186
9	Tension stress in lining	37.5	11.1	11.6	10.8	8.2	7.6	5.5	3.9	3.4	28
10	Tunnel convergence	32.7	17.7	12	10.2	8.2	6	4.4	4.3	4.4	129


Boreability Index
Frequncy
0 5 10 15 20 25 30 35
Benford
1 2 3 4 5 6 7 8 9
First Digit
(a)
Maximum Shear Stress
Frequency
0 5 10 15 20 25 30 35 40
Benford
1 2 3 4 5 6 7 8 9
First Digit
(b)
Maximum Surface Settlment
Frequency
0 5 10 15 20 25 30 35
Benford
1 2 3 4 5 6 7 8 9
First Digit
(c)
P-Wave Velocity
Frequency
0 5 10 15 20 25 30 35
Benford
1 2 3 4 5 6 7 8 9
First Digit
(d)
Point Load
Frequncy
0 5 10 15 20 25 30 35
Benford
1 2 3 4 5 6 7 8 9
First Digit
(e)
Shaft Resistence
Frequncy
0 5 10 15 20 25 30 35
Benford
1 2 3 4 5 6 7 8 9
First Digit
(f)

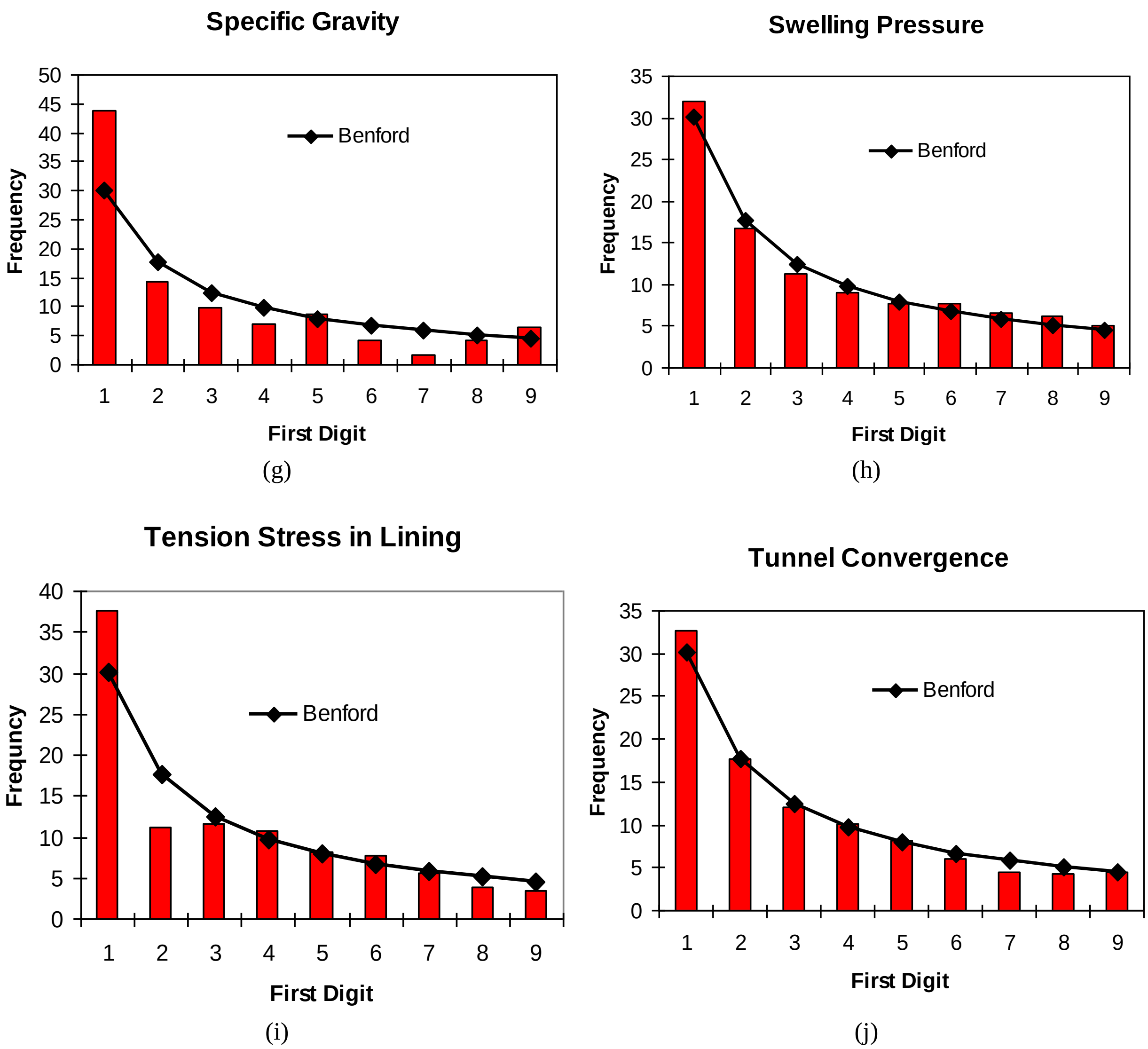


Fig. 1. Percentage of each leading digit in each data set in comparison with Benford's distribution

In Eq. (8) with 8 degrees of freedom, e_i : is the observed number in each bin in the empirical data and b_i : is the number expected by Benford. Note that higher values of χ^2 imply that the empirical data is less similar to Benford's distribution; the 10%, 5%, and 1% critical values for χ^2 are 13.36, 15.51, and 20.09 (Judge and Schechter, 2009).

As it can be seen in Table 5 tunnel convergence has the least value and shear stress has the largest value between our data values. That means by this test, tunnel convergence has the best fitness and shear stress has the worst fitness to Benford's distribution in this test. Arrangement of data sets according to χ^2 is shown in Figure 2.

CONCLUSIONS

• In order to recognize the earth and its components more, some experimental tests are implemented in rock and soil species. Furthermore, field tests are employed for cognizing the earth in a large scale. Experimental results of rock mechanics experimental tests which are investigated in this paper are including rock shear strength, point load strength index, specific gravity of rock, P-wave velocity of rock, various kinds of rocks and some specifies of rocks and soils like value of maximum surface settlement caused by tunneling, swelling pressure of mud rock, tunnel convergence, boreability index, tension stress in tunnel lining and shaft resistance of a pile in rock, which entirely are utilized to describe the geotechnical analysis. Every kind of frauds and fakes in which the experimental results are manipulated, will wash out the experimental results.

The successful experience of utilizing Benford's law in analyzing data which are investigated in this paper, proposes a method of quality control which can be utilized by the side of inspectors and supervisors.

- Since most of our data sets are explanation of earth's specifications and nature it is possible that other specifications of earth follow this law, so we suggest using of Benford's law in similar situations.
- Note that considering the fact that obtaining real geotechnical data is hard, number of samples in our data set was low and we believe that if the number of samples was large the error would be decreased significantly. As it can be seen from the results, the error of data with law number of samples was greater than the other data sets.
- The data obtained by formulas depend on the parameters that were used in those formulas. So, if the real data follow Benford's law, then we expect that the data obtained by formulas should follow as well. This means that if the generated data do not follow Benford's law then we have to modify the proposed formulas or parameters.

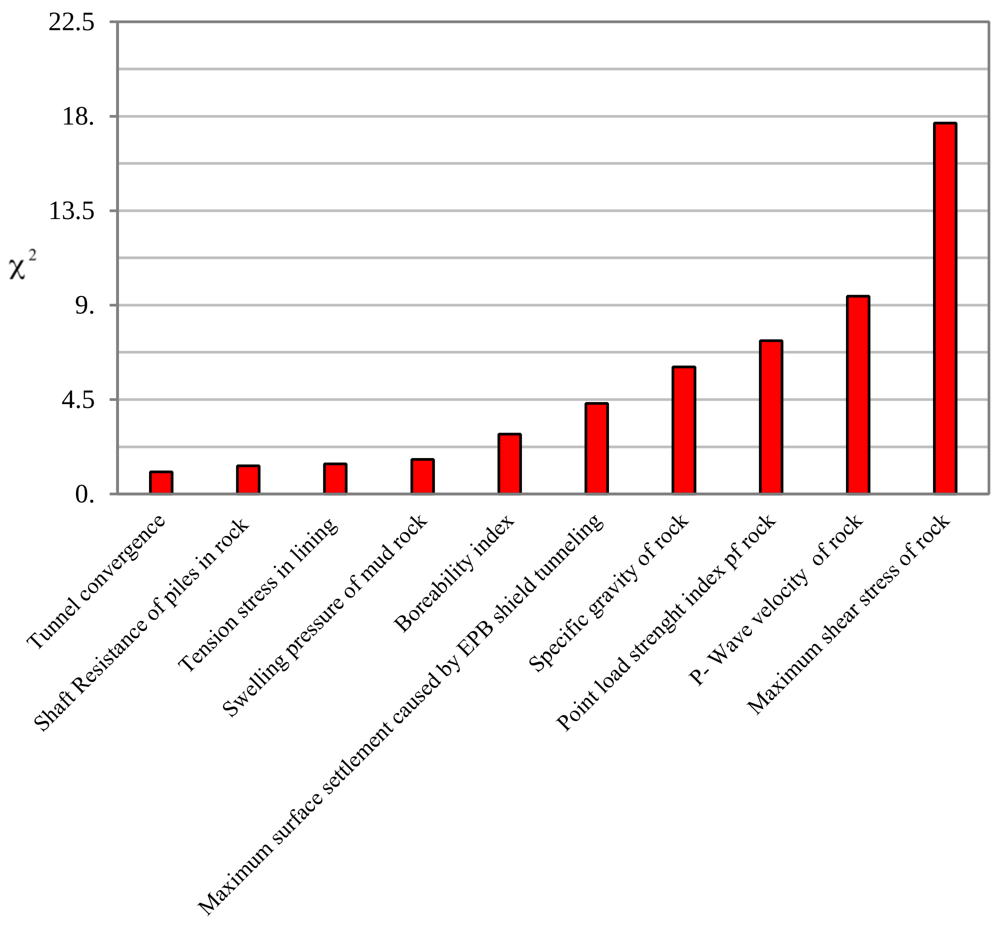


Fig. 2. Arrangement of data sets according to χ^2

Table 5. χ^2 test and MAE between Benford's law and data sets

Col.	Data set	χ^2	MAE
1	Boreability index	2.85	1.91
2	Maximum shear stress of rock	17.68	3.59
3	Maximum surface settlement caused by EPB shield tunneling	4.32	2.67
4	P-Wave velocity in rock	7.3	3.84
5	Point load strength index of rock	9.43	3.08
6	Shaft resistance of piles in rock	1.35	1.16
7	Specific gravity of rock	6.05	3.67
8	Swelling pressure of mud rock	1.65	0.91
9	Tension stress in lining	1.43	2.18
10	Tunnel convergence	1.06	0.77

REFERENCES

Al-Ajmi, A.M. and Zimmerman, R.W. (2005). "Relation between the Mogi and the Coulomb failure criteria", *International Journal of Rock Mechanics and Mining Sciences*, 42(3), 431-439.

Alipour, A., Jafari, A. and Hossaini, S.M.F. (2012). "Application of ANNs and MVLRA for estimation of specific charge in small tunnel", *International Journal of Geomechanics*, 12(2), 189-192.

Alipour, A., Mokharian, M. and Chehreghani, S. (2018). "An application of fuzzy sets to the blastability index (BI) used in rock engineering", *Periodica Polytechnica Civil Engineering*, 62(3): 580-589.

Alipour, A., Mokhtarian, M. and Abdollahei Sharif, J. (2012). "Artificial neural network or empirical criteria? A comparative approach in evaluating maximum charge per delay in surface mining, Sungun copper mine", *Journal of the Geological Society of India*, 79(6), 652-658.

Alves, A.D., Yanasse, H.H. and Soma, N.Y. (2016). "An analysis of bibliometric indicators to JCR according to Benford's law", *Scientometrics*, 107(3), 1489-1499.

Ariznavarreta-Fernández, F., González-Palacio, C., Menéndez-Díaz, A. and Ordoñez, C. (2016). "Measurement system with angular encoders for continuous monitoring of tunnel convergence". *Tunnelling and Underground Space Technology*, 56, 176-185.

Arshadi, L. and Jahangir, A.H. (2014). "Benford's law behavior of Internet traffic", *Journal of Network and Computer Applications*, 40, 194-205.

Barkhordari, K. and Entezari Zarch, H. (2015). "Prediction of permanent earthquake-induced deformation in earth dams and embankments using Artificial Neural Networks", *Civil Engineering Infrastructures Journal*, 48(2), 271-283.

Benford, F. (1938). "The law of anomalous numbers", *Proceedings of the American Philosophical Society*, 551-572.

Broch, E. and Franklin, J. (1972). "The point-load strength test", *International Journal of Rock Mechanics and Mining Sciences and Geomechanics*, Abstracts, 9(6), 669-676, Pergamon.

Chakraborty, A., Raina, A., Ramulu, M., Choudhury, P., Haldar, A., Sahoo, P. and Bandopadhyay, C. (2004). "Development of rational models for tunnel blast prediction based on a parametric study", *Geotechnical and Geological Engineering*, 22(4), 477-496.

Díaz, J., Gallart, J. and Ruiz, M. (2014). "On the ability of the Benford's Law to detect earthquakes and discriminate seismic signals", *Seismological Research Letters*, 86(1), 192-201.

Doostmohammadi, R. (2016). "Investigation of swelling pressure of weak rocks in vicinity of support systems", *Journal of Mining Science*, 52(3), 473-480.

Fewster, R.M. (2009). "A simple explanation of Benford's Law", *The American Statistician*, 63(1), 26-32.

Geyer, A. and Martí, J. (2012). "Applying Benford's law to volcanology", *Geology*, 40(4), 327-330.

Gong, Q., Zhao, J. and Jiang, Y. (2007). "In situ TBM penetration tests and rock mass boreability analysis in hard rock tunnels", *Tunnelling and Underground Space Technology*, 22(3), 303-316.

Hales, D.N., Chakravorty, S.S. and Sridharan, V. (2009). "Testing Benford's Law for improving supply chain decision-making: A field experiment", *International Journal of Production Economics*, 122(2), 606-618.

Hill, T. P. (1995). "Base-invariance implies Benford's law", *Proceedings of the American Mathematical Society*, 123(3), 887-895.

Joannes-Boyau, R., Bodin, T., Scheffers, A., Sambridge, M. and May, S.M. (2015). "Using Benford's law to investigate Natural Hazard dataset homogeneity", *Scientific Reports*, 5, 12046.

Judge, G. and Schechter, L. (2009). "Detecting problems in survey data using Benford's Law", *Journal of Human Resources*, 44(1), 1-24.

Knuth, D.E., Saitou, H., Nagao, T., Matui, S., Matui, T. and Yamauchi, H. (2004). *The art of computer programming*, Vol. 2, *Seminumerical algorithms*, (Japanese Edition), ASCII.

Kohestani, V.R., Vosoghi, M., Hassanlourad, M. and Fallahnia, M. (2017). "Bearing capacity of shallow foundations on cohesionless soils: A random forest based approach", *Civil Engineering Infrastructures Journal*, 50(1), 35-49.

Ley, E. (1996). "On the peculiar distribution of the US stock indexes' digits", *The American Statistician*, 50(4), 311-313.

Mogi, K. (1971). "Fracture and flow of rocks under high triaxial compression", *Journal of Geophysical Research*, 76(5), 1255-1269.

Moosavi, M., Yazdanpanah, M. and Doostmohammadi, R. (2006). "Modeling the cyclic swelling pressure of mudrock using artificial neural networks", *Engineering Geology*, 87(3), 178-194.

Newcomb, S. (1881). "Note on the frequency of use of the different digits in natural numbers", *American Journal of Mathematics*, 4(1), 39-40.

Nigrini, M.J. (1996). "A taxpayer compliance application of Benford's law", *The Journal of the American Taxation Association*, 18(1), 72.

Nigrini, M.J. (1999). "Adding value with digital analysis", *Internal Auditor*, 56(1), 21-24.

Nigrini, M.J. and Wells, J.T. (2012). *Applications for forensic accounting auditing, and fraud detection*, Vol. 586, John Wiley & Sons.

Panthi, K. and Nilsen, B. (2007). "Uncertainty analysis of tunnel squeezing for two tunnel cases from Nepal Himalaya", *International Journal of Rock Mechanics and Mining Sciences*, 44(1), 67-76.

Pérez-González, F., Heileman, G.L. and Abdallah, C.T. (2007). "Benford's law in image processing", *International Conference on Image Processing (ICIP 2007)*, IEEE.

Pérez-Romero, J., Oteo, C. and De la Fuente, P. (2007). "Design and optimisation of the lining of a tunnel in the presence of expansive clay levels", *Tunnelling and Underground Space Technology*, 22(1), 10-22.

Pinkham, R.S. (1961). "On the distribution of first significant digits", *The Annals of Mathematical Statistics*, 32(4), 1223-1230.

Sambridge, M., Tkalčić, H. and Arroucau, P. (2011). "Benford's law of first digits: From mathematical curiosity to change detector", *Asia Pacific Mathematics Newsletter*, 1(4), 1-6.

Sambridge, M., Tkalčić, H. and Jackson, A. (2010). "Benford's law in the natural sciences", *Geophysical Research Letters*, 37(22), 1-5.

Serrano, A. and Olalla, C. (2006). "Shaft resistance of piles in rock: Comparison between in situ test data and theory using the Hoek and Brown failure criterion", *International Journal of Rock Mechanics and Mining Sciences*, 43(5), 826-830.

Shahraki, M., Rafiee-Dehkharghani, R. and Behnia, K. (2018). "Three-dimensional Finite Element modeling of stone column-improved soft saturated ground", *Civil Engineering Infrastructures Journal*, 51(2), 389-403.

Suwansawat, S. and Einstein, H.H. (2006). "Artificial Neural Networks for predicting the maximum surface settlement caused by EPB shield tunneling", *Tunnelling and Underground Space Technology*, 21(2), 133-150.

Tseng, H.-C., Huang, W.-N. and Huang, D.-W. (2017). "Modified Benford's law for two-exponent distributions", *Scientometrics*, 110(3), 1403-1413.

Wei, L., Sundararajan, A., Sarwat, A.I., Biswas, S. and Ibrahim, E. (2017). "A distributed intelligent framework for electricity theft detection using benford's law and stackelberg game", *2017 Resilience Week (RWS)*.

Whyman, G., Ohtori, N., Shulzinger, E. and Bormashenko, E. (2016). "Revisiting the Benford law: When the Benford-like distribution of leading digits in sets of numerical data is expectable?", *Physica A: Statistical Mechanics and its Applications*, 461, 595-601.

Yin, L.J., Gong, Q.M. and Zhao, J. (2014). "Study on rock mass boreability by TBM penetration test under different in situ stress conditions", *Tunnelling and Underground Space Technology*, 43, 413-4