AIP | Chaos
An Interdisciplinary Journal of Nonlinear Science

Do dynamical systems follow Benford's law?

Charles R. Tolle, Joanne L. Budzien, and Randall A. LaViolette

Citation: Chaos: An Interdisciplinary Journal of Nonlinear Science **10**, 331 (2000); doi: 10.1063/1.166498
View online: http://dx.doi.org/10.1063/1.166498
View Table of Contents: http://scitation.aip.org/content/aip/journal/chaos/10/2?ver=pdfcov
Published by the AIP Publishing

Do dynamical systems follow Benford's law?

Charles R. Tolle, Joanne L. Budzien, and Randall A. LaViolette[a)]
Idaho National Engineering and Environmental Laboratory, Idaho Falls, Idaho 83415-2208

(Received 23 November 1999; accepted for publication 14 March 2000)

Data compiled from a variety of sources follow Benford's law, which gives a monotonically decreasing distribution of the first digit (1 through 9). We examine the frequency of the first digit of the coordinates of the trajectories generated by some common dynamical systems. One-dimensional cellular automata fulfill the expectation that the frequency of the first digit is uniform. The molecular dynamics of fluids, on the other hand, provides trajectories that follow Benford's law. Finally, three chaotic systems are considered: Lorenz, Hénon, and Rössler. The Lorenz system generates trajectories that follow Benford's law. The Hénon system generates trajectories that resemble neither the uniform distribution nor Benford's law. Finally, the Rössler system generates trajectories that follow the uniform distribution for some parameters choices, and Benford's law for others. © *2000 American Institute of Physics.* [S1054-1500(00)01402-6]

Benford's law states that the distribution of the first digit of decimal numbers compiled from diverse sources should decrease logarithmically as the digit increases. This surprising result contrasts with the intuition that states that the digits would be distributed uniformly. Consequently, Benford's law has been advocated as a test of the diversity of data, and especially as a means to identify falsified data. Dynamical systems, especially chaotic dynamical systems, are widely employed to model all manner of physical and even social systems. We thought that Benford's law might provide another useful criterion for selecting dynamical models: if the data follow Benford's law, then the model dynamical system should do so as well. We could not find any simple cellular automata that follow Benford's law. On the other hand, we found examples of the molecular dynamics of both gas phase and condensed phase that do follow Benford's law. Finally, we looked at popular low-dimensional chaotic models, and found examples of both compliance with and deviance from Benford's law, depending upon the models and the parameters. The fact that chaotic data might not follow Benford's law also suggests that caution should be exercised in using Benford's law to identify falsified data.

I. INTRODUCTION

The frequency of the first digit in a wide variety of data (physical constants, receipts, population of counties, tax returns, area of rivers) is not uniformly distributed. Instead, the frequency $f(d_1)$ of the first digit d_1 is given by the distribution below:

$$f(d_1)=\log_{10}\left(1+\frac{1}{d_1}\right), \quad d_1=1\cdot\cdot\cdot 9. \tag{1}$$

This surprising result, first given in 1881 by Simon Newcomb,[1] and today known as Benford's law,[2] has been thoroughly reviewed in the mathematical[3–6] and popular science literature.[7–10] A rigorous mathematical proof has been supplied only recently.[4] Hill has shown that the first digits follow Benford's law when data are collected, not simply from randomly sampling a distribution, but from randomly sampling a large collection of different distributions.[6,8] The distribution of the second, third, etc., digits are much closer to the uniform distribution than is the distribution of the first digit.

The ubiquity of Benford's law for data compiled from diverse sources suggests that it is worthwhile to examine physical systems for such a feature. Dynamical systems seem ripe for such a study, because they, especially the chaotic systems, are often employed to model a wide variety of physical and even social phenomena. In this work, we ask if the distribution of the first digits of the trajectories generated by various nonlinear dynamical models follows Benford's law, or the uniform distribution (i.e., $f(d_1)=1/9$), or neither. To answer this, besides using visual inspection of the data, we also calculate the "D statistic,"[11] defined by $D=\max|F_{\rm emp}(d_1)-F_{\rm calc}(d_1)|$, i.e., the maximum deviation between the empirical and the calculated cumulative distribution of the first digit. Although the D statistic could be further employed in a variety of statistical tests for significance, e.g., Kolmogorov–Smirnov,[11] in order to quantify the goodness-of-fit between the empirical and the calculated distribution, we employ the D statistic merely to discriminate between the fits to the uniform distribution or Benford's law. To give some perspective, comparing Benford's law with the uniform distribution gives $D=0.27$, corresponding to a strong disagreement. Finally, and only for chaotic systems, we will also compare surrogate data to test the applicability of Benford's law.

The plan for this article follows. In the next section we examine three sorts of dynamical models: one-dimensional two-state cellular automata, three-dimensional molecular dynamics of atoms, both in the gas phase and in amorphous minimum-energy configurations quenched from the liquid,

[a)]Author to whom correspondence should be addressed. Electronic mail: yaq@inel.gov

and finally, low-dimensional chaotic models. For each, we collect the first digit of the coordinates as the system evolves, and score the frequency. This article concludes with a section discussing the significance of these results.

II. DYNAMICAL MODELS

A. Cellular automata

Cellular automata provide some of the simplest models of dynamical systems that fill phase space in a complicated way.[12] We examined elements from a class of one-dimensional two-state cellular automata, discussed thoroughly by Wolfram.[13] For these, we chose a line of sites labeled $0,\ldots,1000$, each having the state 0 or 1, and applied periodic boundary conditions. Thus the sequence is given a fair chance to follow either Benford's law or a uniform distribution for the first digit; with a restricted choice such as $0,\ldots,500$, the first digits 5–9 would have been suppressed compared to the first digits 1–4. A sequence of lines is generated by the cellular automata using only the nearest-neighbors of each site. For each step, we binned the first digit of the label of each site with state 1. The distribution of the leading digit is essentially uniform, for a variety of rules describing cellular automata (rules 18, 30, 110, 126, 184 in Wolfram's notation), and for a variety of initial conditions (single point and disordered). The D statistic in each case is less than 0.10 when compared with the uniform distribution, but is at least 0.25 when compared with Benford's law. The cellular automata generate trajectories whose coordinates have an essentially uniform distribution of the first digit, and do not follow Benford's law.

B. Molecular dynamics

Here we examine the molecular dynamics of atoms, generating classical many-body trajectories in three spatial dimensions from accurate numerical solutions to Newton's equations, using nonlinear pairwise interatomic forces.[14,15] Here, unlike the other systems considered, we distinguish between momenta and positions. At equilibrium, we already know the (Gaussian) distribution of the momenta.[15] Therefore we consider only the spatial coordinates in this section as the only nontrivial case. First, we examined a gas of cesium atoms, and second, the inherent structure of liquid zinc bromide, both generated by molecular dynamics simulations. The first system is hot and dilute, the second is cold and dense. In both cases, periodic boundary conditions were employed, but the coordinates were not remapped into the original cell.

In the first case, the interatomic forces are essentially of the van der Waals type, but with softer repulsive forces than for the rare gases.[16] We initialized a gas of 13 cesium atoms in a periodic cubic cell of length 14 Å, giving a density of 1.05 g/cm^3. Molecular dynamics calculations for gases can be problematic because of the relative infrequency of interatomic collisions. This can result in a non-Gaussian distribution of the fluctuations in the kinetic energy, corresponding to a nonequilibrium state. Therefore, we augmented Newton's equations of motion in order to maintain Gaussian kinetic-energy fluctuations.[17] The resulting ''constant-temperature'' constant-volume simulation was run for 500 ps, after an equilibration run of 100 ps. The average temperature was $T=1700\,°\mathrm{C}$, the average pressure was $P=0.17$ GPa, and the average potential energy was $U=3.3$ kcal/mole. As a final check, we calculated the short-time variation of the velocity autocorrelation function, and found that it fits $\exp(-t/t_0)$, where $t_0=1.0$ ps, a relatively short time. We sampled configurations every 0.5 ps. As Fig. 1 shows, the frequency of the first digit appears to follows Benford's law. Furthermore, when compared with Benford's law, $D=0.07$; when compared with the uniform distribution, $D=0.20$, just the reverse of the cellular automata.

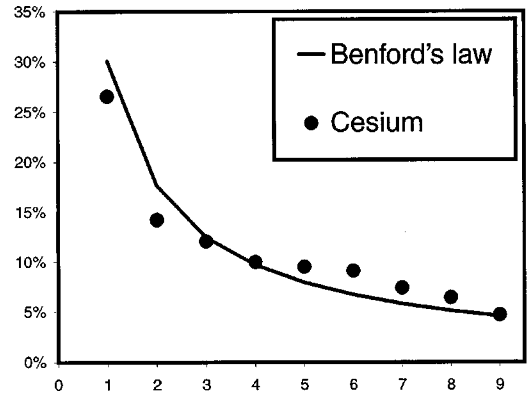

FIG. 1. Comparison of the first digit frequencies for cesium gas with Benford's law.

In the second case, the interatomic forces consist of both van der Waals and Coulomb forces.[18] The latter were calculated with the Ewald summation technique.[14] We initialized a liquid of 486 zinc and 972 bromine ions (the natural formula is $ZnBr_2$) in periodic cubic cells for densities at 2.18, 3.11, and 4.17 g/cm^3, respectively (see Fig. 2). For each density, the initial velocities were adjusted to to give a mean temperature of $2000\,°\mathrm{C}$, well above the melting temperature $(390\,°\mathrm{C})$ at room pressure. However, the dynamics were not augmented, as they were in the simulation of cesium gas, because here the density was sufficiently high, and the collisions sufficiently frequent to produce an equilibrium liquid.

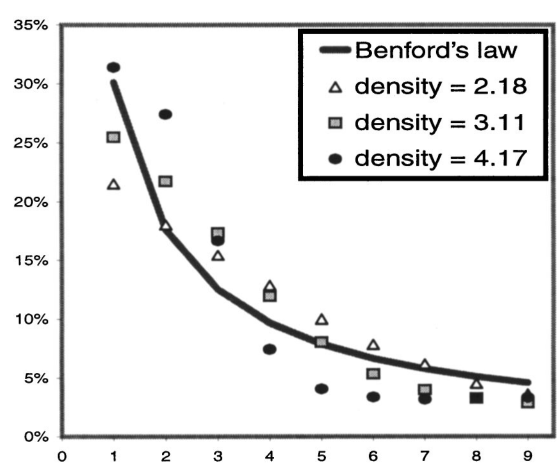

FIG. 2. Comparison of the first digit frequencies for the inherent structure of liquid zinc bromide with Benford's law.

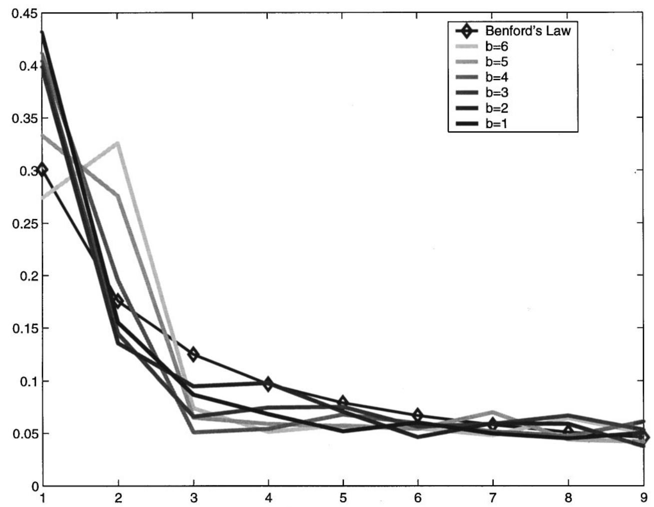

FIG. 3. First digit frequencies of 3000 Lorenz x states ($\sigma=16$, $r=45.92$, and $b=6\cdots1$) compared with Benford's law.

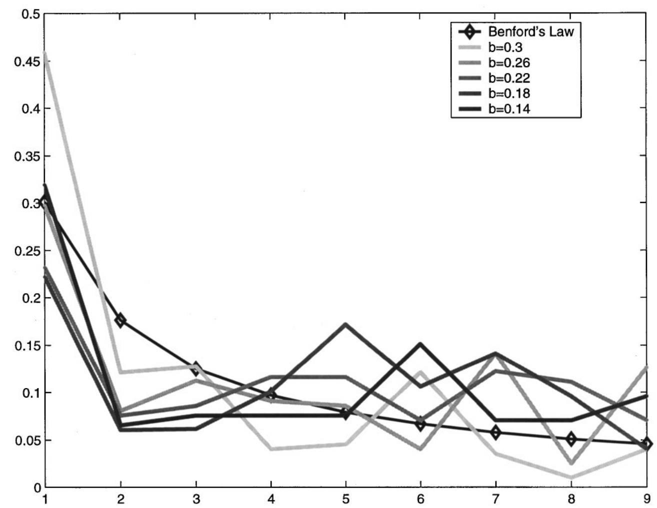

FIG. 4. First digit frequencies of 200 Hénon x states ($a=1.3$, and $b=0.3\cdots0.14$) compared with Benford's law.

The constant-energy constant-volume dynamics resulting from the solution of the unaugmented Newton's equations were run for 22–40 ps, after 30 ps of equilibration, for each density. For each run, 20 quenched structures were collected by applying a steepest-descent minimization to the instantaneous liquid configuration every 1–2 ps. The resulting amorphous minimum-energy configurations constitute the inherent structure of the liquid at this density.[19] The inherent structure has zero temperature, and is independent of the temperature of the equilibrium liquid.[20] The pressure of the inherent structure is positive for the high density structure, near zero for the intermediate density structure, and negative for the low density structure. For each density of the inherent structure, the frequency of the first digit follows Benford's law: $0.07\leqslant D\leqslant0.15$ for all densities, with the better agreement coming at the lower densities. Although this is not great quantitative agreement, the comparison to the uniform distribution is much worse, with $D\geqslant0.23$ for all densities. Furthermore, as with the cesium gas, the decrease in the frequency of the first digit is strictly monotonic. We conclude that the results for the inherent structure follow Benford's law qualitatively. For completeness, we note that the high temperature liquid itself also follows Benford's law qualitatively, with slightly less than or equal statistical significance compared with the inherent structure.

C. Chaotic dynamics

Finally we examined the low-dimensional chaotic models of dynamical systems. We examined the Lorenz, Hénon, and Rössler systems,[21] for a variety of parameters.

Lorenz system:

$$\begin{aligned}\dot{x}(t)&=\sigma(y(t)-x(t)),\\ \dot{y}(t)&=rx(t)-y(t)-x(t)z(t),\\ \dot{z}(t)&=x(t)y(t)-bz(t);\end{aligned}\tag{2}$$

Hénon map:

$$\begin{aligned}x_{n+1}&=y_n+1-ax_n^2,\\ y_{n+1}&=bx_n;\end{aligned}\tag{3}$$

Rössler system:

$$\begin{aligned}\dot{x}(t)&=-(y(t)-z(t)),\\ \dot{y}(t)&=x(t)-ay(t),\\ \dot{z}(t)&=b+z(t)(x(t)-c).\end{aligned}\tag{4}$$

At each sample step in the trajectory, we binned the first digit of the x-coordinate for comparison, similar results hold for the other states as well. Figures 3, 4, and 5 show the distributions of first digits. The D statistic for the sample signals is given in Tables I, II, and III.

The Lorenz System follows Benford's law, at least qualitatively, for all of the parameters considered here; the disagreement with the uniform distribution is consistently strong. However, the data is much more mixed for the other

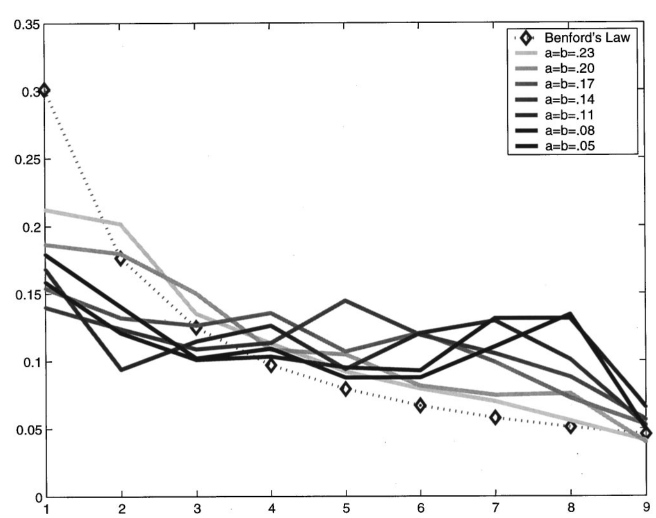

FIG. 5. First digit frequencies for 800 Rössler x states ($a,b=0.23\cdots0.05$, and $c=5.7$) compared with Benford's law.

TABLE I. D statistic for Lorenz signals compared with Benford's law and uniform distribution, respectively. ($\sigma=16$, $r=45.92$, and $b=6\cdots1$.)

b	Benford	Uniform
6	0.12	0.38
5	0.13	0.39
4	0.13	0.38
3	0.10	0.33
2	0.10	0.31
1	0.13	0.13

two systems. It is not clear from either the figures or the D statistic that the Hénon system follows either the uniform distribution or Benford's law. All of the observed distributions display a high occurrence of the first digit, even for the cases where the D statistic favors the uniform distribution. On the other hand, even for the cases that seem to be favorable for Benford's law, the observed frequency of ''2'' is only about half that predicted by Benford's law. Therefore we cannot say much more about the distribution of the first digit for the Hénon system. The situation is clearer for the Rössler system, which produces trajectories that transform from following Benford's law to the uniform digit distribution through a simple parameter change. However, the parameter change is not otherwise accompanied by dramatic changes in the dynamics. Consider the three Rössler attractors in Fig. 6: Which are the two that follow Benford's law? We have been unable to discern anything in Fig. 6 that would indicate in advance which of these should correspond to trajectories that follow Benford's law. We found in fact that only the plots on each of the top and bottom hand panels correspond to trajectories that follow Benford's law; the trajectory corresponding to the middle panel generates an essentially uniform distribution of the first digit.

For our final test of the applicability of Benford's law to chaotic systems, we generated surrogate data for these systems. The concept of surrogate data is widely used to implement null hypothesis testing for chaotic nature of strange signals and data sets.[12,22] Surrogate data consists of random signals whose autocorrelation function is identical to a given signal.[12] One randomizes the phases (i.e., the imaginary part) of the Fourier transform of the signal with white noise, and then inverts the randomized transform to generate the (real) surrogate signal. A plot of a typical surrogate data signal has been time delay embedded and plotted against the time delay embedded phase space reconstruction of the Lorenz system ($\sigma=16$, $r=45.92$, and $b=6$) in Fig. 7. The results for the digit frequency calculation for the surrogate data is given in Fig. 8. The surrogate data in effect magnifies discrepancies between the observed and the calculated distribution, so the qualitative match of the surrogate data with both Lorenz and Benford's law is gratifying. The importance of this example however is to note that the surrogate data is actually a random signal, albeit with correlations, that nevertheless produces a distribution of first digits that resembles Benford's law.

TABLE II. D statistic for Hénon signals compared with Benford's law and uniform distribution, respectively. ($a=1.3$, and $b=0.3\cdots0.14$.)

b	Benford	Uniform
0.3	0.16	0.37
0.26	0.14	0.19
0.22	0.21	0.12
0.18	0.26	0.11
0.14	0.17	0.21

TABLE III. D statistic for Rössler signals compared with Benford's law and uniform distribution ($a,b=0.23\cdots0.05$, and $c=5.7$) and ($a=0.20$, $b=0.14$, and $c=5.7$).

Parameters	Benford	Uniform
$a=b=0.23$	0.09	0.22
$a=b=0.2$	0.11	0.18
$a=b=0.17$	0.19	0.11
$a=b=0.14$	0.23	0.08
$a=b=0.11$	0.23	0.07
$a=b=0.08$	0.22	0.06
$a=b=0.05$	0.18	0.10
$a=0.2$ $b=0.14$	0.11	0.22

III. DISCUSSION

We give two qualifications before completing our presentation of these results. First, no pretense is made here that the agreement with Benford's law, where observed, is anything but qualitative. Second, no pretense is made here that our results are anything other than empirical.

That the cellular automata produced a uniform distribution of the first digit does not seem surprising. Indeed, chaotic cellular automata have been proposed for generating uniform random numbers.[23] Visual inspection of the phase space generated by these cellular automata also suggests a uniform distribution. That the molecular dynamics of the cesium gas did follow Benford's law also does not seem surprising. Although the dynamics for the gas is practically ergodic, which by itself might have suggested a uniform distribution, it is also diffusive. In diffusing systems, the mean-square displacement plays the role of the variance in a Gaussian distribution, and, the mean-square displacement grows (linearly) in time,[15] thereby effectively generating a family of Gaussian distributions for the coordinates. Samples drawn from a large collection of distributions (Gaussian or not) follow Benford's law,[6] so we could have expected that Benford's law would hold for the case at hand. It is more surprising that the distribution of the coordinates in the inherent structure of the liquid also obeys Benford's law, because the structure of the quenches is independent of both time and temperature of the equilibrium liquid.[20] The case of the chaos models is the one that seems the most complicated, and, as Fig. 6 suggests, the least susceptible to intuition. It is also the most interesting, because of the wide variety of behavior that can be produced.

Simple chaotic dynamical systems have been employed or proposed to treat the broad range of applications in the

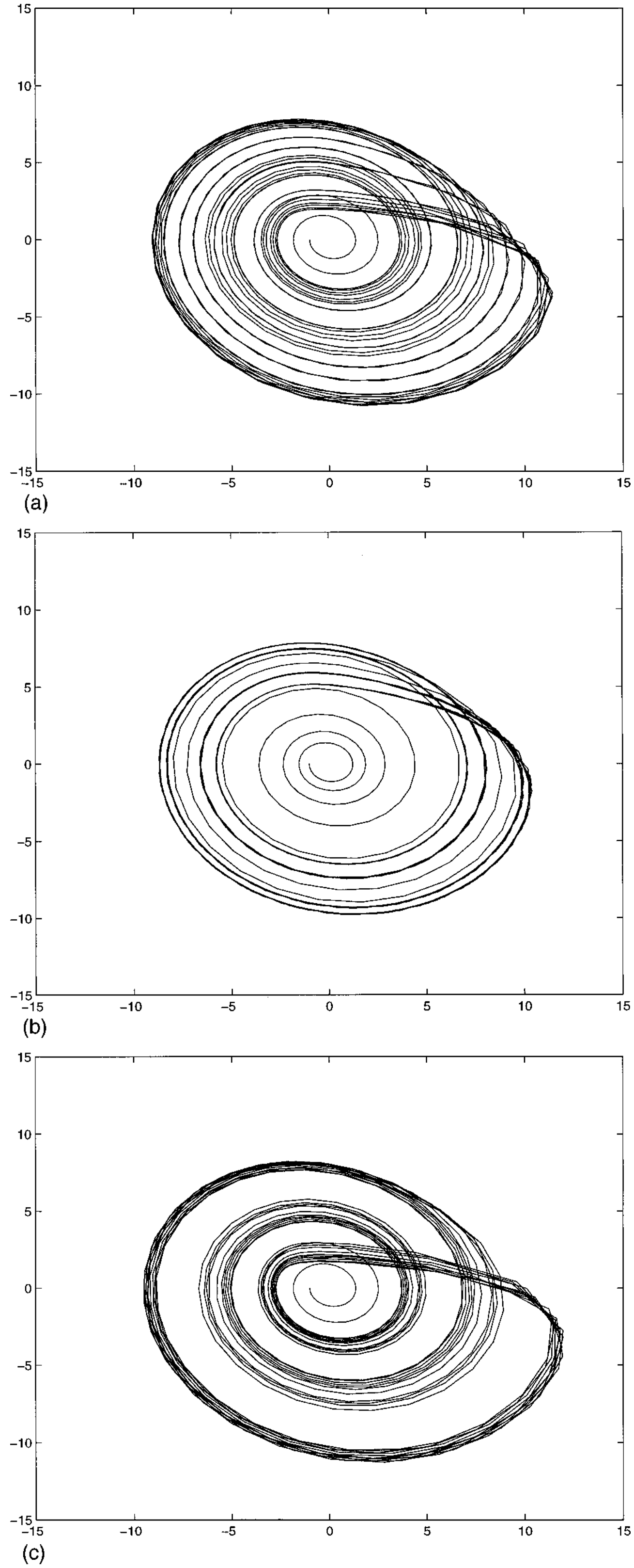

FIG. 6. Plots of the x vs. y states of the Rössler for ($a=0.2$, $b=0.2$, and $c=5.7$), ($a=0.14$, $b=0.14$, and $c=5.7$) and ($a=0.2$, $b=0.14$, and $c=5.7$), top to bottom.

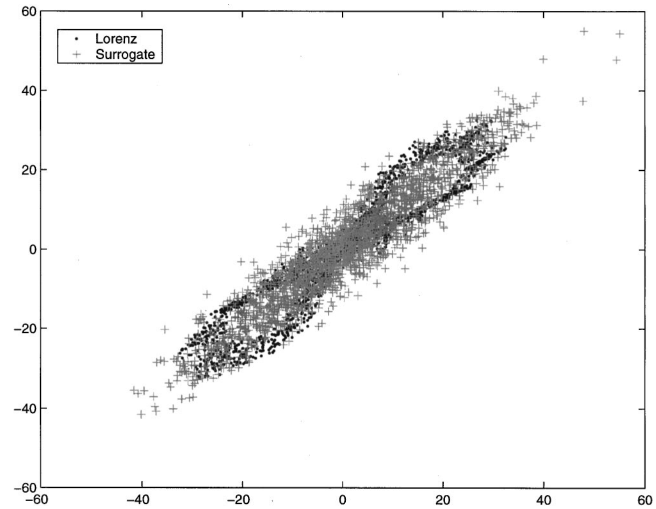

FIG. 7. Phase space reconstruction using time delay embedding $[x(t)$ vs. $x(t+\tau)]$ for Lorenz ($\sigma=16$, $r=45.92$, and $b=6$) and Lorenz surrogate data.

physical, natural, and social sciences.[21] Data from everyday life apparently follow Benford's law, so the fact that some of these models also do, under certain conditions, may add to the explanation of their success. It may be that these results will encourage others to add a test for adherence to Benford's law as part of the diagnostics employed for the application of chaotic dynamical systems to the modeling of data. Finally, one of the principal applications currently advocated for Benford's law is in the detection of fraudulent data, the hypothesis being that ''natural'' data will follow Benford's law.[9,24] However, we have shown, for the Rössler system, that the distinction between the trajectories that do, and do not, emulate Benford's law may be otherwise small. To the extent then, that chaotic dynamical systems can otherwise model the data in question, the application of Benford's law to distinguish ''natural'' data from ''artificial'' data may require more careful thought.

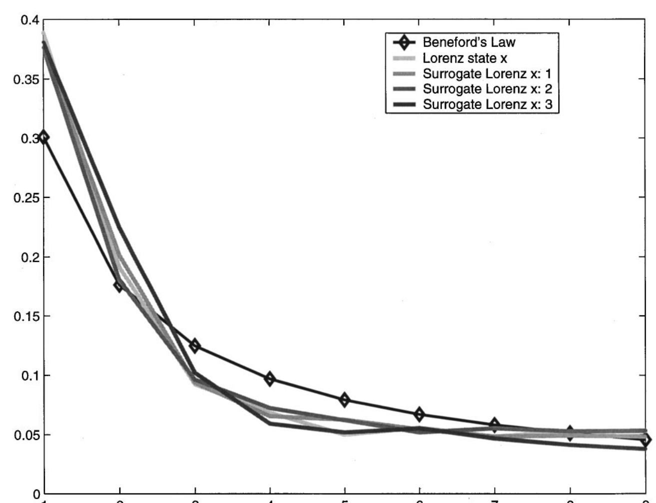

FIG. 8. First digit frequencies for Lorenz ($\sigma=16$, $r=45.92$, and $b=6$) and Lorenz surrogate data plotted against Benford's law.

ACKNOWLEDGMENT

This work is supported by the Environmental Systems Research and Analysis Program, Office of Environmental Management, U.S. Department of Energy, under DOE Idaho Operations Office Contract DE-AC07-94ID13223.

[1] S. Newcomb, ''Note on the frequency of use of the different digits in natural numbers,'' Am. J. Math. **4**, 39–40 (1881).
[2] F. Benford, ''The law of anomalous numbers,'' Proc. Am. Philos. Soc. **78**, 551–572 (1938).
[3] R. A. Raimi, ''The first digit problem,'' Am. Math. Monthly **83**, 521–538 (1976).
[4] T. P. Hill, ''Base-invariance implies Benford's law,'' Proc. Am. Math. Soc. **123**, 887–895 (1995).
[5] T. P. Hill, ''The significant-digit phenomenon,'' Am. Math. Monthly **102**, 322–326 (1995).
[6] T. P. Hill, ''A statistical derivation of the significant-digit law,'' Stat. Sci. **10**, 354–363 (1995).
[7] R. A. Raimi, ''The peculiar distribution of first digits,'' Sci. Am. **221**, 109–119 (1969).
[8] T. P. Hill, ''The first digit phenomenon,'' Am. Sci. **86**, 358–363 (1998).
[9] R. Matthews, ''The power of one,'' New Sci. **163:2194**, 27–30 (1999); see also http://www.newscientist.com/ns/19990710/thepowerof.html
[10] E. W. Weisstein, http://www.treasuretroves.com/math/BenfordsLaw.html, 11 August 1999.
[11] J. V. Bradley, *Distribution-Free Statistical Tests* (Prentice-Hall, Englewood Cliffs, NJ, 1968).
[12] D. Kaplan and L. Glass, *Understanding Nonlinear Dynamics* (Springer-Verlag, New York, 1995).
[13] S. Wolfram, *Theory and Applications of Cellular Automata* (World Scientific, Singapore, 1986).
[14] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, Oxford, 1987).
[15] J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids*, 2nd ed. (Academic, San Diego, 1986).
[16] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff, ''UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations,'' J. Am. Chem. Soc. **114**, 10024–10046 (1992).
[17] H. J. C. Berendsen, J. P. C. Postma, W. F. van Gunsteren, A. Di Nola, and J. R. Haak, ''Molecular dynamics with coupling to an external bath,'' J. Chem. Phys. **81**, 3684–3689 (1984).
[18] M. J. L. Sangster and M. Dixon, ''Interionic potentials in alkali halides and their use in simulation of molten salts,'' Adv. Phys. **25**, 247–342 (1976).
[19] F. H. Stillinger and T. A. Weber, ''Packing structures and transitions in liquids and solids,'' Science **225**, 983–989 (1984).
[20] F. H. Stillinger and R. A. LaViolette, ''Sensitivity of liquid-state inherent structure to details of intermolecular forces,'' J. Chem. Phys. **83**, 6413 (1985).
[21] S. H. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering* (Addison-Wesley, Reading, PA, 1994).
[22] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, ''Testing for nonlinearly in time series: the method of surrogate data,'' Physica D **58**, 77–94 (1992).
[23] S. Wolfram, ''Random sequence generation by cellular automata,'' Adv. Appl. Math. **7**, 123–169 (1986).
[24] M. Nigrini, ''A taxpayer compliance application of Benford's law,'' J. Amer. Tax. Assoc. **18**, 72–91 (1996).