Eur. Phys. J. B **25**, 245–251 (2002)
DOI: 10.1140/epjb/e20020028

THE EUROPEAN PHYSICAL JOURNAL B

EDP Sciences

Self-similarity in highway traffic

E.G. Campari[1,a] and G. Levi[2]

[1] Physics Department of Bologna University and Istituto Nazionale per la Fisica della Materia viale Berti-Pichat 6/2 40127 Bologna, Italy

[2] Physics Department of Bologna University and Istituto Nazionale di Fisica Nucleare viale Berti-Pichat 6/2 40127 Bologna, Italy

Received 23 April 2001 and Received in final form 28 September 2001

Abstract. Highway traffic as simulated with a simple cellular automata model has been analysed in a search for self-similarity in the behaviour of car density and car flow as a function of space. Fractal dimensions between 1.5 and 1.6, depending on the simulation characteristics, have been measured with a box counting algorithm. The self-similarity spans over about 2 orders of magnitude. A comparison with experimental data is suggested.

PACS. 05.45.Df Fractals – 45.70.Vn Granular models of complex systems; traffic flow

1 Introduction

Traffic flow is a problem which is increasingly attracting the interest of scientists. This is due partly to the great economical and social relevance of the problem, partly to the interesting features which are emerging from its study. Let us consider highways, for instance: different kinds of traffic can be found and transitions between them which resemble phase transitions [1–3]. There are density and flow waves, which can born spontaneously or be triggered by roadworks or ramps [4–7]. Some of these waves are non-dispersive and could even be solitons [8]. More generally, traffic reveals itself as an highly non linear system [9] and therefore exhibits most of the features which are commonly found in those systems.

A peculiar feature of many non linear systems is the generation of structures which appear self-similar under varying degrees of magnification. The observation of a self-similar behaviour on a number of scales lets us define and measure a fractal dimension. Strictly speaking, the fractal dimension of a set is just a geometrical feature of that set. Nevertheless, when the set is originated by measurable quantities from a physical phenomenon, like earthquake tremors, certain chemical reactions or fluid turbulence, this geometrical feature reveals something about the system dynamics. The presence of a fractal dimension first of all confirms the system's non linearity. But also can be a signature of deterministic chaos: a behaviour of systems whose dynamics is attracted towards a strange attractor, which is self-similar [10,11].

Computing a value for D, the fractal dimension, has at least a double meaning: (a) It can be a powerful tool to investigate the dynamical properties of car traffic treated as a physical system, detecting for example different flow conditions. (b) The computed value can be compared to a measured value from real data in order to have a better understanding about the validity of the model.

The state of a highway is usually given by the value of car density, ρ, and car speed, v, as a function of the distance x from the road beginning or as a function of time t at a fixed point along the road. Car flow, f, is then computed as $f = \rho v$. Therefore, the variables defining the system are functions of both time and space. A 3-dimensional plot of these variables yields highly fragmented patterns. This suggests that it is possible to measure their eventual fractal dimension, which should be a number between 2 and 3. As stated by Mandelbrot [16], a linear section of a surface like $\rho(x,t)$ should still be a fractal, with dimension reduced to $D-1$. A well known example of this behaviour is Earth landscape, whose fractal dimension is between 2 and 3, while coastlines-landscape sections at constant height- have a dimension between 1 and 2. In the rest of this article we will describe how we analyzed simulated highway traffic data, obtained with a simple cellular automata model, in search for a fractal dimension. Following the suggestion mentioned above, we will limit our analysis to one dimensional curves, that is to car flow and car density along the highway at fixed time. This choice makes it possible to employ for the data analysis algorithms like the Yardstick method (see Sect. 3) which cannot be used for surfaces. Furthermore, fixing time or space helps having more homogeneous data, which is what we need in order to look for differences in the fractal dimension of the phases of traffic. To our knowledge, a similar task was previously done only on a single lane round-about without entrance or exit [12]. This is so unreal as a highway model as to cast doubts on its usefulness for any comparison

[a] e-mail: campari@df.unibo.it

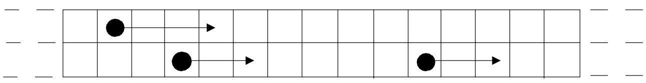


Fig. 1. Cars move from left to right along the two lane highway. At each iteration every car position, given by the occupation of a cell, and car speed are updated according to a defined set of rules.

with experimental data. We plan in the near future to accomplish a similar analysis on real data, which will be collected on Italian highways, to check the simulated results.

2 Highway model and traffic phases

In order to simulate traffic for a 2 lane highway, we built a cellular automata model [8,13], which provides a simple physical picture of the system and is easily implemented on computers. In our model, the highway lanes are divided into cells which can be either empty or occupied by a car (see Fig. 1). The road starts at cell 1; cars, which enter the road at its beginning and at on-ramps along the highway, move in discrete steps (named time ticks in the following) in the direction of increasing cell number. Since on and off-ramps are equivalent in many effects, like jams generation [6,8], in this model there are no off-ramps except for the highway end, where cars leave the road. On-ramps are modeled as a single entrance cell in lane 1.

A set of rules, described in details in reference [8], specify the time and space evolution of the system. These rules were derived from those originally developed by Schreckenberg and coworkers [14] and by Nagel and coworkers [5] to implement a multilane highway without ramps and with periodic boundary conditions. These rules specify at each step and for every car: (a) the speed with which cars move along the road, (b) how faster cars overtake slower ones and (c) how cars slow down or accelerate according to the behavior of neighbouring vehicles. The model is asymmetric: one lane is for normal cruise, the other for overtaking. Cars have an unbounded braking ability in order to avoid car accidents.

In each traffic simulation a starting highway configuration is chosen. Street length, on-ramps number and location, position and speed of cars already inside the highway are set up. New cars enter the highway at a chosen rate as such: for all on-ramps, at each iteration the program checks if the cell is empty. If empty, a random number in the [0,1] interval is generated and if it is less than the threshold chosen for that simulation, a car is generated in that cell with a speed of 2 cell/time tick. At each iteration a number of local and average quantities are recorded. For each car, position and speed; for each cell, mean speed, density and flow, averaging over the 100 cells before and after the given cell. The use of averages is due first of all to the discreteness of cellular automata models. Cars in the model can only have as speed an integer number between 0 and 9, while car speeds are real numbers in nature. Furthermore, experimental data are averages over time and/or space [1,3], so that a proper comparison again requires one to average simulated data. A comparison, of

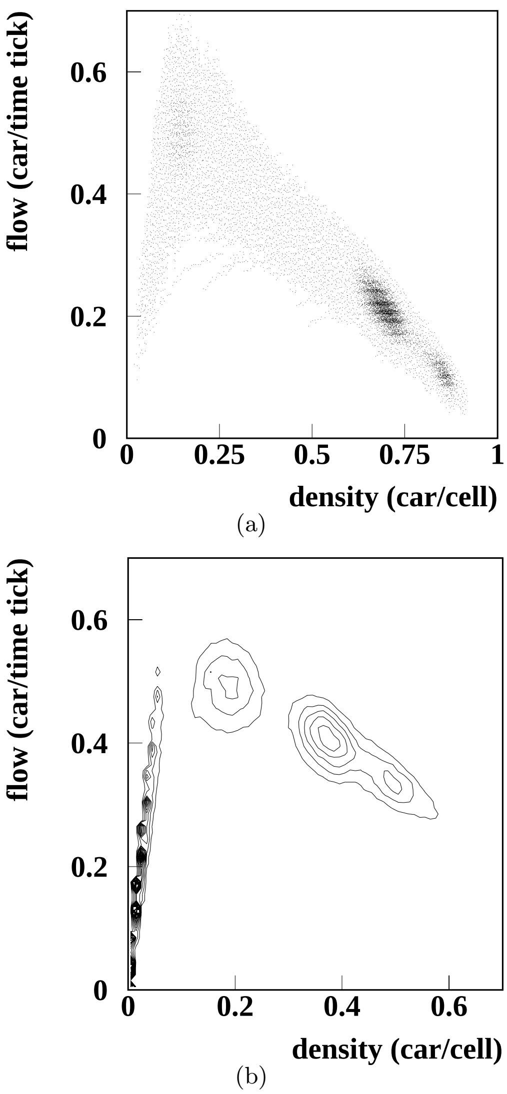


Fig. 2. (a) Flow as a function of density as obtained in a cellular automata simulation. Points on the left correspond to cars in free flow conditions while those on the right are due to congested flow conditions. (b) The results of a different simulation are reported using a population level contour plot. In the contoured regions are found about 99% of the points. Moving from left to right are found a region with cars in free flow, in synchronized flow and in congested flow.

course, can be done only after a length is attributed to a cell and a time tick is turned into time. With the reasonable choice of cell length = 5 m and time tick = 1 s, a speed of 6 cell/time tick corresponds to 108 km/h.

As reported elsewhere [8], this model proved to be able to reproduce some of the experimental data available on highway traffic. A noteworthy example is reported in Figures 2a and b , where it is shown the so-called fundamental plot of traffic as obtained in two simulations with different parameters. In Figure 2a every point corresponds to a value of car flow as a function of car density from the first simulation. At low density, points ideally align along a first curve with positive slope, while at higher densities they stack on a negative slope curve. In Figure 2b a

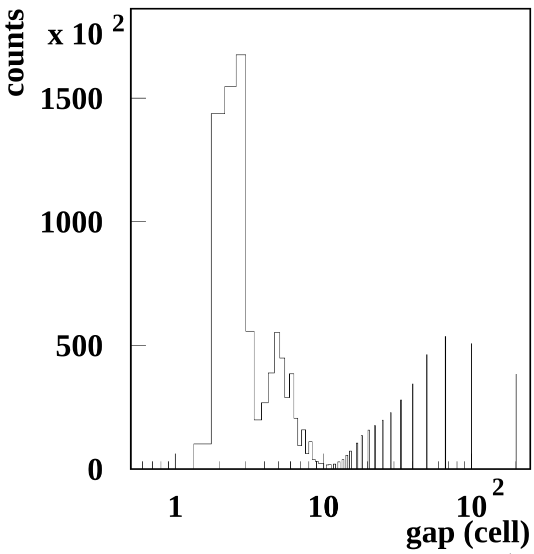
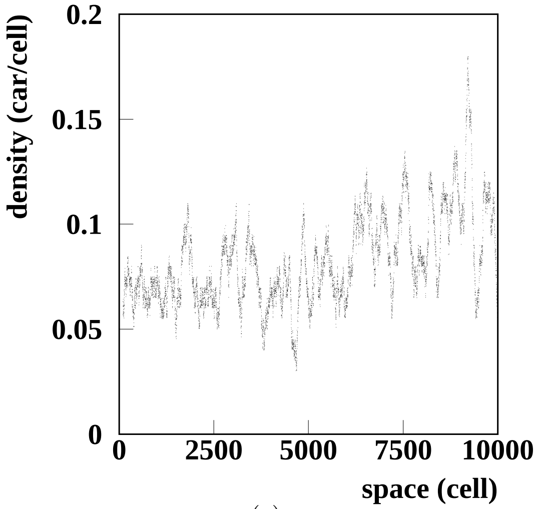


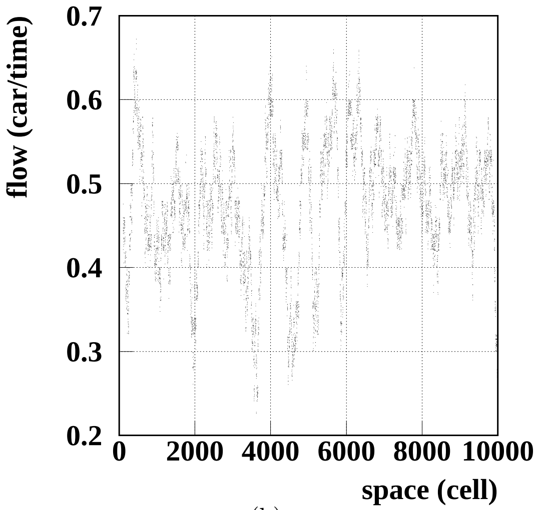
Fig. 3. Histogram reporting the average distance (gap) between cars for the same simulation of Figure 2b. The peak centered between 2 and 3 cell is that of congested flow; the one at about 5 cell that of synchronized flow while the last peak, at about 68 cell is that of free flow. The peculiar shape of the histogram at high gap values is due to the use of a discrete model. At low densities the graininess of the model becomes evident.

population level contour plot is used to put into evidence how traffic concentrates into three regions, corresponding, from left to right, to free flow, synchronized flow and congested flow. These are the main phases into which traffic can be subdivided. In the following, we will give a brief description for each of these phases. For a more complete description the reader is referred to [1] for a general review and to [4] for some experimental data showing the appearance of these phases in real highway traffic.

Free flow is a condition where fast cars can easily overtake slow ones, there are large gaps between cars and traffic flows easily. It is, in a sense, the ideal traffic condition, for cars are almost free to run along the highway as they like. This phase occurs at low car density (below 0.1 car/cell in the computer simulations of this work), where the more cars there are in the road the higher is the flow, producing the elongated region with positive slope in the fundamental diagram of Figures 2a and b. At the opposite range of density, is found the so-called congested flow. It is a condition of heavy traffic, with flow decreasing with increasing car density (negative slope region at the right of Fig. 2b) and easy formation of jams. Overtaking is extremely difficult and car speed can be nearly zero. Synchronized flow is intermediate between the former and the latter. It is defined as a state of traffic in multilane roads in which the vehicles in different lanes move with almost the same speed, about 1/2 that of free flow. In this region flow can be high in spite of an increasing density, but the linear correlation between flow and density is lost and the two quantities becomes totally non-correlated. At each phase of traffic corresponds a different value of the average distance between cars [15]. This is reported in the car gap histogram of Figure 3. Data comes from the same simulation used to produce Figure 2b. Three peaks are shown in this histogram: the peak centered between 2 and 3 cells is that of congested flow; the one at about 5 cell that of synchronized flow while the last peak, at about 68 cells is that of free flow.



(a)



(b)

Fig. 4. (a) Car density as a function of space at time $t = 4800$ time ticks from the beginning of the simulation. It is shown a highway portion 10000 cell long. Traffic is in free flow conditions. (b) Car flow corresponding to the same conditions reported for Figure 4a. A grid is superimposed on the figure to illustrate how the box counting algorithm works.

3 Fractal analysis

A typical plot as obtained from our simulations for car density as a function of space at fixed time, is shown in Figure 4a. In Figure 4b car flow is plotted in the same conditions. Similar curves were obtained for ρ, v and f, plotted as a function of time at fixed space. The repeated fragmented look common to all of these plots when viewed

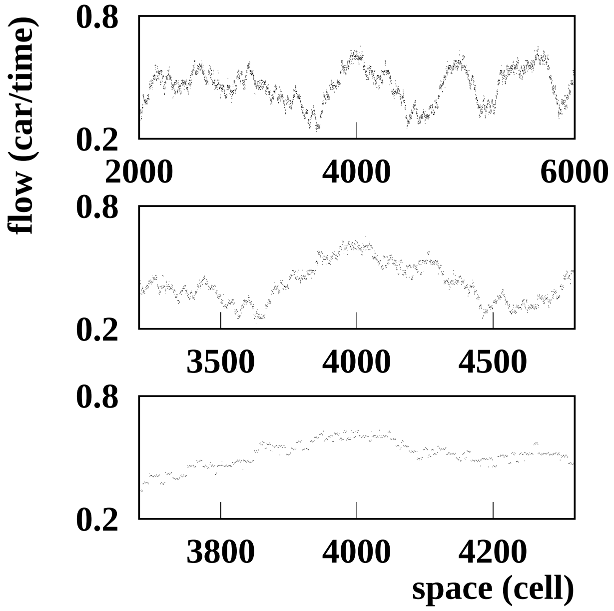


Fig. 5. Three successive magnifications by a factor 2.5 of that part of Figure 4b centered about cell 4000. At every magnification a point distribution like the original one is obtained, with the partial exception of the bottom figure. Here the scale is comparable to the number of cells used to compute averages (200 cells).

at different scales is evidenced in Figure 5, where portions of Figure 4b at increasing magnification are shown in sequence. In the following will be described results obtained on car density and car flow in simulations with different starting conditions.

The curves are analysed with the box counting algorithm [11,12,17,18]. This is probably the most widely used algorithm to extract a non integer dimension from some data. Let consider for example Figure 4b. The plot area is divided into 25 boxes with side length 1/5 of the abscissa and ordinate range of variation. This can be done again and again using n^2 boxes of decreasing side length ($1/n$ of the whole length). An estimate of the fractal dimension D of the curve is obtained from a linear fit of $\log(N)$ as a function of $\log(1/n)$, where N is, for any value of the side length, the number of squares which contain at least one point of the curve. With a simple computer program the value of N is readily obtained. In Figures 6a and b are shown the results of such an analysis for two simulations with cars in free flow and congested flow conditions. The capability of the program to correctly yield the fractal dimension of several point sets was tested with fractal objects taken from mathematics and with non fractal curves. Some results of these tests are reported in Table 1. A further analysis, based on the discrete Fourier transform was accomplished, revealing that our traffic data do not have periodic components.

Only a subset of the points of Figures 6a and b can be fitted by a line whose slope is computed to be $D = (1.55 \pm 0.01)$ for the free flow condition and $D = (1.49 \pm 0.01)$ for the congested flow condition. These subsets roughly correspond to a box side $\Delta x \geq$ (200 cell), 200 being the number of cells used in the simulations to compute average quantities. The limitation occurs because when the square side is lower than the value used to compute averages, the averaging process starts smoothing away every eventual fractal scaling. Unfortunately, the number of cells used to compute averages, cannot be reduced too much below 200, since this would produce density, speed and flow values too rough to have any interest for the study of traffic.

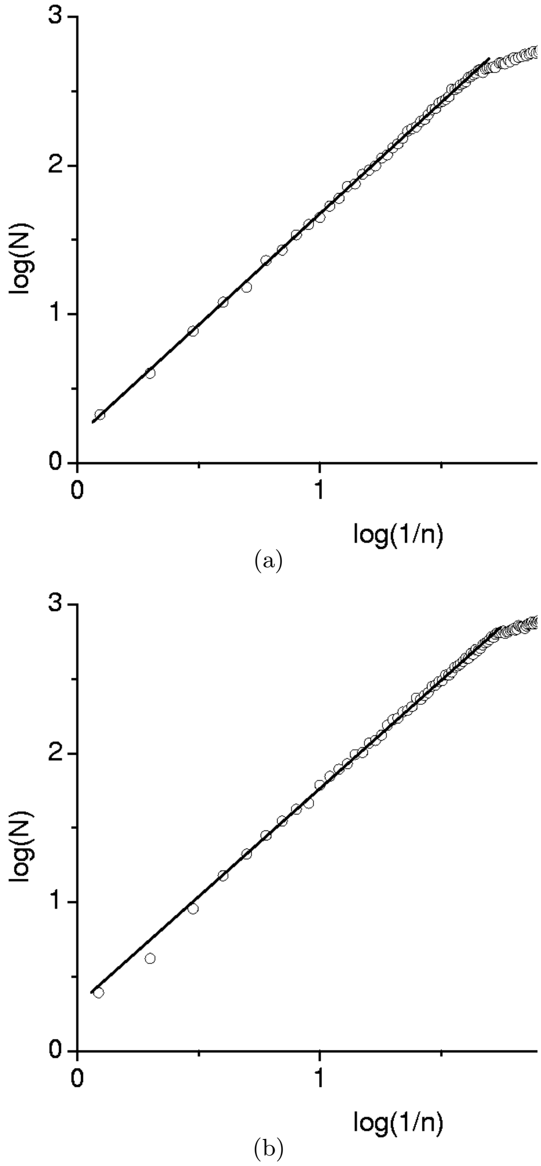


Fig. 6. (a) Box counting analysis of car density for a simulation in free flow conditions. Data sticks to a line whose slope $D = 1.55 \pm 0.01$ is taken as the fractal dimension of that traffic simulation. (b) Box counting analysis of car density for a simulation in congested flow conditions. Here is found a fractal dimension $D = 1.49 \pm 0.01$. The slope change for the points at the extreme right of the figures occurs when the box side becomes smaller than the average number of cells used to compute the averages.

Table 1. Some well known fractal sets (first column) are analysed with the box counting algorithm. In the second and third columns, respectively, are reported the exact value and that obtained with the program used to analyse traffic data.

Fractal	$D_{\rm th}$	$D_{\rm box}$
Sierpinski gasket	1.585	1.59±0.01
Sierpinski carpet	1.893	1.87±0.02
Kock curve	1.2618	1.269±0.007
quadratic Kock curve	1.5	1.481±0.005
sin	1	1.001±0.001

The existence of a lower boundary to the feasibility to analyse our simulated data in search for a fractal dimension is peculiar to the use of a cellular automata algorithm. Despite simulations based on differential equations or other continuous models, here a discreteness is introduced from the beginning. Therefore, even if this is a numerical work, the granularity of the system forbid to obtain a fractal range larger than that of real data. The fractal dimension of most real systems usually spans no more than 1 or 2 logarithm decades, like in the well known case of country boundaries data [19] or stock market prices [20] or fractal slip in crystals [21,22]. Here, the measured fractal range, for the best cases analysed up to now, span 1.96 decimal decades.

Confidence in the goodness of our results is provided by two facts. First of all the quality of the model, which is surprisingly close (being so simple) to real highway data for other measured quantities. The second fact is the comparison with mathematical fractals, like the Kock or Sierpinski curves. These sets exhibits self similarity at every scale. Starting with their construction rule, we generated subsets of these objects containing the same number of points as the sets used to compute the fractal dimension of traffic data. As reported in Table 1, their fractal dimension is correctly computed, but due to the finite number of points in the set the self-similarity never spans more than 2.5 decades. Taking into account the absolute noiseless of these data sets, we believe that up to 1.96 decades for highway traffic is a good result.

In summary, the search for a self-similar scaling in a cellular automata traffic simulation suffers from a double limitation. Firstly, the limited range over which every natural phenomenon exhibits self-similarity. Secondly, the above mentioned smoothing effect, due to the use of a discrete model.

The simulations analysed with the box counting show D values scattered in the 1.4–1.6 range both for car density and car flow. This spread in D values coming from simulations performed with different traffic conditions, suggested us to consider if any significant difference in fractal dimension could be found between phases of traffic. This would represent a way to distinguish between them. In order to test this hypothesis, we analyzed simulated data from the phases which are most different between them in terms of traffic flow. These are free flow and congested flow, which should exhibit the greatest difference. Table 2 reports the average results of 12 different analysis of each phase for car density and car flow as a function of space at fixed time. Despite our hope, the results of the simulations analysed up to now with the box counting method do not reveal a significant enough D difference, even if the average D value of free flow is slightly greater than that of congested flow. It is to say that traffic conditions along our highway, usually 10 000 to 20 000 cells long, are never completely uniform. As it happens in a long highway, the prevalence of congested traffic do not exclude the presence of sections with a more fluid traffic and *vice versa*. This effect would mix up D values, decreasing their difference. It is also possible that highway traffic flow is multifractal in nature, that is an object whose different regions have different fractal properties [23]. Again, our analysis would compute an average D value.

Table 2. Average values of fractal dimension D from simulations with free flow conditions or congested flow conditions. The D values are obtained with the Box counting algorithm. Synchronized flow was not analysed also because of the difficulty in having large portions of the street with this kind of flow. At each value is associated the standard error of its mean.

Traffic phase	D(free flow)	D(congested flow)
Car flow	1.565±0.012	1.547±0.016
Car density	1.546±0.019	1.515±0.025

Searching to understand if the little D differences measured with the box counting could really be a signature of a difference in fractal dimension between free flow and congested flow, we reanalysed the same data with another algorithm: the yardstick or structured walk method [11,16]. With this new algorithm, evidences of a difference in fractal dimension between traffic phases were clearly found.

The yardstick algorithm prescribes to repeatedly measure the length of a curve using yardsticks of always decreasing length L and can be outlined as follows:

1) set the yardstick (or compass) at a step length L;

2) take the initial point of the curve;

3) draw an arc, centered at the initial point, which crosses the curve;

4) the point where the arc first cross the curve becomes the centre of a second arc;

5) draw a new arc centred at the crossing point and repeat the above procedure until the curve end is reached.

6) take a shorter step length L and start again the measuring procedure.

As for the box counting method, an estimate for D is obtained from the slope of $\log(N)$ *versus* $\log(1/L)$. N is the number of times the yardstick of length L is used to measure the curve length and increases as L gets shorter and shorter. This algorithm is usually equivalent to the box counting one, but is less versatile, being limited to the study of curves in the plane. The case of our traffic data is one of those where the method is not equivalent to the box counting and provides dissimilar D values. Another

Table 3. Average values of fractal dimension D from simulations with free flow conditions or congested flow conditions. The D values are obtained with the Yardstick algorithm. At each value is associated the standard error of its mean.

Traffic phase	D(free flow)	D(congested flow)
Car flow	1.93±0.02	1.75±0.03
Car density	1.89±0.02	1.76±0.02

case is that of self-affine fractals, which are self-similar only if magnified by a different factor along different directions. The problem, which affects all those methods to compute D which rely on a measure of length (the correlation dimension, for example) can be outlined as follows. The original traffic data are (x, y) couples, with x the cell number and y car density or car flow. The yardstick length L is computed as $L = \sqrt{\Delta x^2 + \Delta y^2}$. Now, x is a number typically $0 < x < 10\,000$, whereas y is $0 < y < 1$, and therefore, $\Delta x > 200$ (recall the above discussion on the need to compute averages on a minimum number of cells) while Δy is usually below 0.1. It is clear that with Δx so great with respect to Δy, $\Delta x^2 + \Delta y^2 \approx \Delta x^2$ and the value of L is practically given by that of Δx. To put things another way, let us consider Figure 4a but now imagine using for the y axis the same scale used for the x axis, that is 0 to 10 000. In the plot, the fragmented curve would be flattened to a straight line, undistinguishable from the x axis. Then no matter how fragmented the values of y may be, D will be measured as 1. The only way to put into evidence the fragmented nature of the curve and to measure a D value different from 1 is to perform a data rescaling. Therefore, the original (x, y) data were normalised to 1 before being used to compute D with the yardstick method. A normalization at 1 of the variables expands a little y-values and shrinks about 4 orders of magnitude x-values. As a result of this transformation the ups and downs of the curve are enhanced and its fractality is increased. It should be also evident that the result of a box counting analysis is insensitive to such a normalization, for this algorithm is not based on length calculations.

Summarising, the yardstick algorithm reveals itself useless to compute the correct value of D. But the method, being sensitive to scale changes, is useful to enhance the little differences between phases of traffic. The normalised data, once analysed with the yardstick method, provides D values higher than those obtained with the box counting. Interestingly, free flow still has a greater D value than congested flow and the difference between the two is increased beyond statistical uncertainty, as reported in Table 3. This somehow supports the hypothesis that a difference in fractal dimension exists for the various phases of traffic. In Figure 7 it is reported the plot of $\log(N)$ *versus* $\log(1/L)$ for one of the data sets used to obtain the values reported in Table 3. As for the box counting, only a subset of the points is used to measure the fractal dimension.

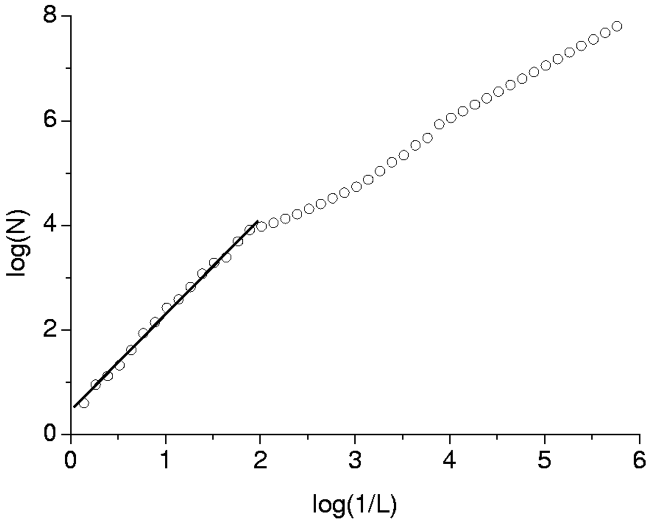


Fig. 7. Yardstick analysis of car density data as a function of space from a simulation with free flow conditions. Values of $\log(1/L)<2$ correspond to a yardstick length greater than the average number of cell used to compute averages. In this region points stick to a line with slope $D = 1.87 \pm 0.04$. Above $\log(1/L) = 2$, a departure from linearity is observed, which is attributed to the lack of self-similarity due to the averaging. The last part of the curve (above 4), is again a straight line, whose slope is 1. Here the yardstick length is equal to or smaller than the (normalized) distance between 2 cell. The yardstick is now so small that the measured value of the curve length is no more dependent on L.

4 Conclusions

The following considerations can be drawn from the above analysis. Highway traffic exhibits self-similarity in both car density and car flow. This is a non obvious result: there is no direct relationship between non-linearity and self-similarity, for a non-linear system could well be non self-similar. Furthermore, the fractal dimension increases moving from free flow towards congested flow. This suggests that in congested traffic there is a smoothing of car variables with respect to free flow, which seems reasonable in a state where all cars are slowly moving along the highway, without many possibilities to overtake other cars or to accelerate. Such a valuable result supports the view of each state of traffic as a real phase and provides a further distinction between traffic conditions. In this regard, it would be interesting to check if D changes monotonously, as does car density, or suddenly, as does the overtaking probability, at each change of traffic condition. This is really difficult to verify by means of simulations and we are currently trying to improve our model in order to obtain good enough data.

Unlike other systems, described with the use of continuous variables, we work with discrete quantities from the very beginning. For this reason the range over which our simulated data exhibits self-similar behaviour is slightly less than 2 decades. This is a value typical of many real systems. An important point is that we do not claim in this article to fully demonstrate the fractal nature of traffic. We rather want to suggest to the scientific community and especially to those who study highway traffic to look into their experimental data for the presence of self-similarity. This has never been done before and our simulations at least suggest that the search could yield a positive result.

As described in the previous section, car density and car flow have D values greater in free flow than in congested flow. Since these results comes mainly from the use of the yardstick method, where data are stretched before being analysed, we cannot exclude that the different values for D are a byproduct of this operation, without relation to the real fractal dimension. This stretching is also responsible for D values close to (but definitely less than) 2. Again, we address to the analysis of experimental data in order to clarify this point.

References

1. B.S. Kerner, Physics World **12**, 25 (1999).
2. T. Nagatani, Physica A **271**, 200 (1999).
3. D.E. Wolf, Physica A **263**, 438 (1999).
4. B.S. Kerner, H. Rehborn, Phys. Rev. E **53**, R1297 (1996).
5. K. Nagel, D.E. Wolf, P. Wagner, P. Simon, Phys. Rev. E **58**, 1425 (1998).
6. G. Diedrich, L. Santen, A. Schadschneider, J. Zittartz, Int. J. Mod. Phys. **11**, 335 (2000).
7. A. Schadschneider, Physica A **285**, 101 (2000).
8. E.G. Campari, G. Levi, Eur. Phys. J. B **17**, 159 (2000).
9. M. Bando, K. Hasebe, A. Nakayama, A. Shibata, Y. Sugiyama, Phys. Rev. E **51**, 1035 (1995).
10. I. Stewart, *Does God play dice?* (Penguin Books 1990, London), Chap. 6.
11. P. Addison, *Fractals and Chaos* (IOP 1997, London).
12. J. Woinowski, in: *Fractal Frontieres '97*, edited by M.M. Novak, T.G. Dewey (World Scientific 1997, Singapore), pp. 321–327.
13. S. Wolfram, Nature **311**, 419 (1984).
14. W. Knospe, L. Santen, A. Schadschneider, M. Schreckenberg, Physica A **265**, 614 (1999).
15. A.D. May, *Traffic flow fundamentals* (Prentice-Hall 1990, New Jersey.).
16. B.B. Mandelbrot, *The fractal geometry of nature* (W.H. Freeman and Company, 1983, N.Y.), Chaps. 4 and 14.
17. K.T. Alligood, T.D. Sauer, J.A. Yorke, *Chaos, an introduction to dynamical systems* (Springer-Verlag 1996, New York), Chap. 4.
18. L. Lam, *Nonlinear physics for beginners* (World Scientific 1998, Singapore), pp. 237–240.
19. L.F. Richardson, Gen. Syst. Yearbook **6**, 139 (1961). The results of this article are also reported and discussed in the cited book of B.B. Mandelbrot.
20. B.B. Mandelbrot, J. Business **36**, 394 (1963).
21. B. Sprusil, F. Hnilica, Czech. J. Phys. **35**, 897 (1985).
22. T. Kleiser, M. Bocek, Z. Metallkde **77**, 582 (1986).
23. H.E. Stanley, P. Meakin, Nature **335**, 405 (1988).