



## A superstatistical model of vehicular traffic flow



Caglar Kosun<sup>a,\*</sup>, Serhan Ozdemir<sup>b,1</sup>

<sup>a</sup> Department of City and Regional Planning, Izmir Institute of Technology, 35430 Urla, Izmir, Turkey

<sup>b</sup> Artificial Intelligence & Design Lab., Department of Mechanical Engineering, Izmir Institute of Technology, 35430 Urla, Izmir, Turkey

### HIGHLIGHTS

- It is shown that superstatistics may be an important tool in traffic flow analysis.
- Short and long time scales are computed for the vehicular flow time series data.
- The  $q$  value computed from  $q$ -Gaussians characterizes the highway segment.
- The  $q$  value obtained from the beta distribution defines a given time series data.
- Traffic flow may be characterized by long-range interactions.

### ARTICLE INFO

#### Article history:

Received 17 August 2015

Received in revised form 13 October 2015

Available online 23 October 2015

#### Keywords:

Superstatistics

Highway traffic

$q$ -Gaussian

Vehicle speeds

Inverse local variance

Nonadditivity

### ABSTRACT

In the analysis of vehicular traffic flow, a myriad of techniques have been implemented. In this study, superstatistics is used in modeling the traffic flow on a highway segment. Traffic variables such as vehicular speeds, volume, and headway were collected for three days. For the superstatistical approach, at least two distinct time scales must exist, so that a superposition of nonequilibrium systems assumption could hold. When the slow dynamics of the vehicle speeds exhibit a Gaussian distribution in between the fluctuations of the system at large, one speaks of a relaxation to a local equilibrium. These Gaussian distributions are found with corresponding standard deviations  $1/\sqrt{\beta}$ . This translates into a series of fluctuating beta values, hence the statistics of statistics, superstatistics. The traffic flow model has generated an inverse temperature parameter (beta) distribution as well as the speed distribution. This beta distribution has shown that the fluctuations in beta are distributed with respect to a chi-square distribution. It must be mentioned that two distinct Tsallis  $q$  values are specified: one is time-dependent and the other is independent. A ramification of these  $q$  values is that the highway segment and the traffic flow generate separate characteristics. This highway segment in question is not only nonadditive in nature, but a nonequilibrium driven system, with frequent relaxations to a Gaussian.

© 2015 Elsevier B.V. All rights reserved.

### 1. Introduction

Real life dynamics of a certain system, e.g. vehicular traffic flow, is often an interwoven dynamics of multiple of those superimposed on one another. If the system demonstrates driven nonequilibrium system characteristics, with multiple time scale separation, then one may suspect of the presence of superstatistics. Currently, a well-separated two or more time scales is an indication of the theory. The superstatistics theory suggests that there should be a fast time scale where there are

\* Corresponding author. Tel.: +90 232 750 7026.

E-mail addresses: [cglrksn@gmail.com](mailto:cglrksn@gmail.com), [caglarkosun@iYTE.edu.tr](mailto:caglarkosun@iyte.edu.tr) (C. Kosun), [serhanozdemir@iYTE.edu.tr](mailto:serhanozdemir@iYTE.edu.tr) (S. Ozdemir).

<sup>1</sup> Tel.: +90 232 750 6585; fax: +90 232 750 6505.

many equilibrium states whose dynamics may very well be described by the Boltzmann–Gibbs statistical mechanics, whose parameter distributions could be described by Gaussians. Then, there is another time scale, a much slower one, whose slow dynamics is accounted for by the environmental changes, for example. To go along with the theory, this change is caused by a parameter, known as the local inverse temperature,  $\beta$ . This inverse temperature  $\beta$  could be any parameter that is representative of the local variance, found by windowing the whole time series data. Since this  $\beta$  is a statistical parameter out of local windows, and whose long term change could also be defined by another global statistics, the term statistics of the statistics, or superstatistics would fit well.

If one overviews the traffic flow in a thermodynamic context, even before delving into the theory of superstatistics, Ref. [1] is a good starting point. Here the authors mention a continuous particle system that is under psychological forces, or interactions. All the interactions are assumed to be of short-range, hereby shearing much of the connection with the Tsallis thermostatistics [2,3], over a closed-loop ring. The two-body repulsion potential is given as a function of the vehicle locations around this ring, and a Hamiltonian of speeds and locations of the particles is provided. The kinetic energy of the Hamiltonian is specified with respect to an optimal speed. An inverse temperature parameter is named mental strain coefficient, and taken constant for a given part of a freeway. This study claims that the steady-state properties are in good agreement with the relevant properties of traffic flow. No long-range interactions are considered. Three different analyses are performed for three distinct (constant) pseudo-temperature parameters, or mental strain coefficients, generating specific distance clearance distributions for vehicular congestions.

Ref. [4] deals with the traffic flow, in a similar vein, determining interaction potentials through steady-state statistics. Only that this time real-life data of Dutch two-lane freeway A9 is used. Pseudo-temperature values are computed and tabulated. But all of these parameters reflect the specific cases, and not a fluctuating  $\beta$  as in superstatistical theory.

Another study [5] has the vehicles distributed not over a closed-loop band but over the surface of a sphere, called a unit sphere. The novelty is that the time spectral rigidity is found via Random Matrix Theory, since it provides a general overlook into the eigenvalues of the random matrix ensembles. This time, headway distributions are obtained for specific inverse temperatures. And the study [6] follows a scenario similar to Refs. [1,4], only that inter-vehicle gap statistics is analyzed at signal-controlled crossroads.

The traffic flow has also been modeled through thermostatistical models as well as traditional techniques that are not mentioned in this paper. A typical unconventional approach is by Ref. [7]. Here, the driven systems in a thermodynamic framework are being discussed. Particles are identified as vehicles. The traffic flow is described by Bando's optimal velocity model in terms of forces. Bando's optimal velocity model allows an internal energy of the car system in a thermodynamic framework. The size of the car cluster is assumed to be a stochastic variable. Hence the solutions range from a fixed-point solution to a free-flow phase in a limit cycle. A detailed analysis on distinguishing the characteristics of a traffic flow is provided, where the system is specified as either driven or equilibrium systems.

As for the superstatistics concept, it is a relatively new idea, whose origins and the theory may be found in Refs. [8–10]. A more comprehensive discussion is found in Ref. [11]. Some of the key points were duly explained. Some of these key points are the fingerprints of the superstatistical dynamics, the ratio of the time scales, and the distribution of the inverse temperature  $\beta$ . The case study is a turbulent Taylor–Couette flow, where the longitudinal velocity differences are processed. One practical finding for the log-normal distribution of the parameter  $\beta$  is that the Tsallis  $q$  variable is linked to the flatness of this distribution.

Another fundamental work in the theory of superstatistics is by Ref. [12]. This work sets the standards for superstatistical analysis of any time series. Inspired by the theory of statistical hypothesis testing, the method proposed in the manuscript computes the parameters of a data set. But most importantly, the proposed method checks for the validity of the superstatistical model assumptions. The practical outcome of this cornerstone study is that, for any superstatistical assumption to be valid, two conditions must be satisfied. The first condition states that the ratio of the slow time scale to the fast one should tend to zero. The latter condition is on the deviation of the Gaussianity, designated by  $\epsilon$ , and dictates that it too should tend to zero.

A recent study [13] on share price returns has focused on determining a proper superstatistical model on a range of time scales. By allowing the time scale to move from minutes to a larger scale of days, a transition from lognormal to chi-square superstatistics has been observed. Even a general model was proposed over a control parameter  $\kappa$ , interpolating a hybrid superstatistical structure. There is a mention that, by varying the system parameters, a range of  $q$  values may be obtained. The factors such as time scale of returns, intraday periodicity, and the sector of shares all affect this range of  $q$  values.

There are numerous applications of superstatistical theory on data analyses, such as sea-level fluctuations [14], high energy physics [15], modeling train delays [16], environmental atmospheric turbulence [17], temperature fluctuations [18], metastasis and cancer survival [19], currents in complex polymers [20], and headway distributions [21].

Of the superstatistical and the traffic flow literature, there are two research papers that are the most relevant: [21] and [22]. Ref. [22] studies the clearance distributions of the traffic flow and the time-headway distributions. Using an analogy, traffic states are considered to be phase changes of matter, as a function of vehicle density. An analytic expression is derived for the spacing distributions that interpolate from the Poisson distribution. Here, the fast dynamics is represented by the vehicle velocity and the slow dynamics by the traffic density. It is stated that any statistic of a congested traffic may be expressed as an average of another statistic for a Gaussian random ensemble over the local mean level spacing.

In this paper, we first investigate the probability distributions of the speed variable at a local point selected from a highway segment. If the space is Euclidean, the memory is Markovian and the time is continuous; the distribution of a given

system would then be considered in a Boltzmann–Gibbs formalism. However, for a real system, such as vehicular traffic flow on the highway segment, this may not be so. Since the systems may exhibit (multi)fractal, non-Markovian properties and the time is discrete, we expect, at least occasionally, that the distributions become  $q$ -Gaussian and Tsallis formalism could be the underlying statistical mechanics. We model the vehicular traffic flow through  $q$ -Gaussian distributions. Additionally, distinct from the literature, this paper also proposes that, specific to chi-square distributions, two  $q$  values may be computed, not just one on the following premises. A  $q$ -Gaussian fit is performed over the vehicle speeds, whereas a chi-square distribution is over the  $\beta$  values. A detailed discussion is provided in Section 3.

## 2. The concept of superstatistics

As briefly mentioned in the introduction, superstatistics is the superposition of at least two statistics. It is derived from the Boltzmann factor  $e^{-\beta E}$  of statistical mechanics and an intensive parameter  $\beta$  [10]. This intensive parameter can be inverse temperature, chemical potential, energy dissipation, etc. The main aspect of the superstatistics concept is the fluctuations in the quantity of the selected intensive parameter in a driven nonequilibrium system. Consider that the system is composed of a series of artificially generated windows and in each window the  $\beta$  is approximately constant, thus the system is described by an ordinary Boltzmann–Gibbs statistics. In the long-term, the local intensive quantities in the system would fluctuate.

Suppose that there is a given time-series data and it has two different time scales  $\tau$  and  $T$ .  $T$  designates the length of the equally partitioned windows of the time series. To reach the local equilibrium, the condition  $\tau \ll T$  should be satisfied where  $\tau$  determines how fast the local equilibrium is reached in each window of the time-series so that when a different dynamics of the following window starts, the dynamics of the previous window has already reached an equilibrium.

Let us consider the function of an averaged Boltzmann factor ( $BF$ ) in Ref. [10]

$$BF(E) = \int_0^\infty f(\beta) e^{-\beta E} d\beta \quad (1)$$

where  $E$  is the energy,  $f(\beta)$  is the probability distribution of  $\beta$ , so that

$$p(E) = \frac{1}{Z} BF(E) \quad \text{and} \quad Z = \int_0^\infty BF(E) dE \quad (2)$$

where  $\frac{1}{Z}$  is the normalization constant of  $e^{-\beta E}$ .

For a  $\beta$ -dependent normalization constant, the probability distribution becomes

$$p(E) = \int_0^\infty f(\beta) \frac{1}{Z(\beta)} e^{-\beta E} d\beta. \quad (3)$$

It is timely to point out an interesting fact that if  $f(\beta)$  from data analysis turns out to be a chi-square distribution ( $\chi^2$ ),  $p(E)$  assumes a Tsallis distribution.

### 2.1. Three widely occurring superstatistical distributions

In the literature, there are three widely known superstatistics distributions. These are chi-square superstatistics, inverse chi-square superstatistics, and log-normal superstatistics. The fluctuations in  $\beta$  of an experimental data can be described by one of these three distributions. More on the three universal classes of these distributions can be found in Ref. [11].

The  $\chi^2$  distribution of  $\beta$  is given as

$$f(\beta) = \frac{1}{\Gamma\left(\frac{n}{2}\right)} \left( \frac{n}{2\beta_0} \right)^{n/2} \beta^{n/2-1} e^{-\frac{n\beta}{2\beta_0}} \quad (4)$$

where  $\beta_0$  is the average of  $\beta$  and  $n$  is the number of degrees of freedom.

The inverse  $\chi^2$  distribution of  $\beta$  is given as

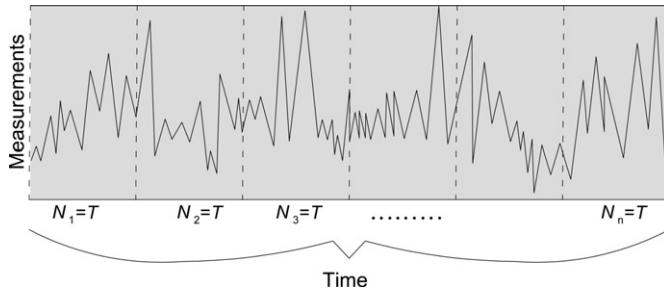
$$f(\beta) = \frac{\beta_0}{\Gamma\left(\frac{n}{2}\right)} \left( \frac{n\beta_0}{2} \right)^{n/2} \beta^{-n/2-2} e^{-\frac{n\beta_0}{2\beta}} \quad (5)$$

where  $\beta_0$  is the average of  $\beta$  and  $n$  is the number of degrees of freedom.

The last one is given by the log-normal superstatistics as below

$$f(\beta) = \frac{1}{\sqrt{2\pi}\sigma\beta} \exp\left(-\frac{\left(\ln\frac{\beta}{m}\right)^2}{2\sigma^2}\right) \quad (6)$$

where  $m$  and  $\sigma$  are the suitable mean and variance parameters.



**Fig. 1.** Representation of equally partitioned windows of a time series data.

## 2.2. Description of time scales for the superstatistical model

In this paper, our data set is defined as  $v$ . Consider  $v = \{v_1, v_2 \dots v_n\}$  and each value in the data set corresponds to single point speed observations.

Let us assume that the local equilibrium is represented by approximately a Gaussian distribution. When the local characteristics of each partitioned cell (window) of the time series exhibits Gaussian behaviour with zero mean and variance  $1/\beta_{T,u}$ , the formulae of the local moments are given as in Ref. [12]

$$\langle v \rangle_{T,u} = 0, \quad \langle v^2 \rangle_{T,u} = \frac{1}{\beta_{T,u}}, \quad \langle v^3 \rangle_{T,u} = 0, \quad \langle v^4 \rangle_{T,u} = \frac{3}{\beta_{T,u}^2}. \quad (7)$$

And the condition for the kurtosis  $\kappa_{T,u} = 3$  hints at local Gaussians.

The correlation function of a time series  $v$  can be obtained as in Ref. [12]

$$C_{k,t}(v) = \frac{1}{k-t} \sum_{i=1}^{k-t} v_i v_{i+t}. \quad (8)$$

The short time scale  $\tau$  of the time series is described by the exponential decay of  $C_{k,\tau}(v)$  and is shown as follows [12]:

$$C_{k,\tau}(v) = e^{-1} C_{k,0}(v). \quad (9)$$

Consider the kurtosis to determine the long time scale [12]:

$$\kappa_T = \frac{1}{N} \sum_{i=1}^N \kappa_{T,i} \quad \text{with} \quad \kappa_{T,i} = \frac{\langle u^4 \rangle_{T,i}}{\langle u^2 \rangle_{T,i}^2} \quad (10)$$

where  $N$  is the total number of identical windows and  $T$  is the length of the windows.

Given the whole time series  $v = \{v_1, v_2 \dots v_n\}$ , the goal is to obtain the average kurtosis. The average kurtosis stands for the average kurtosis of the windows when the whole time series is divided into windows. Based on the average kurtosis formula, the calculation is performed for the different window lengths to obtain the time scale  $T$ . Let the kurtosis quantity for a time window of the time series data corresponds roughly to 3. This indicates that the time windows contain data which are approximately normally distributed, Fig. 1.

## 2.3. An auxiliary condition

Let us suppose that the local behaviour of the windows of length  $T$  does not have perfect Gaussian distribution, then  $\varphi_{T,u}$  is introduced to determine the deviations from the fourth moment of 3, and shown as [12]

$$\varphi_{T,u} = \langle v^4 \rangle_{T,u} - 3 \langle v^2 \rangle_{T,u}^2. \quad (11)$$

Gaussian approximation is obtained when  $|\varepsilon| \ll 1$  by

$$\varepsilon = \frac{1}{3} \frac{\sum_{i=1}^N \varphi_{T,i}}{\sum_{i=1}^N \langle v^2 \rangle_{T,i}^2}. \quad (12)$$

For a suitable Gaussian approximation, the numerator in the formula would tend to zero. Thus, for the Gaussian approximation of a given time series, the value  $|\varepsilon|$  should be small.

**Table 1**

Results of analysis for the speed, speed differences and shuffled speed data.

Variable	$\tau$	$T$	$\tau/T$	$ \varepsilon $	$q(q\text{-Gaussian fit})$	$q(\beta\text{-fit})$
Speed	2.4	18	0.13	0.01	1.3	1.8
Speed differences	1.5	16	0.09	0.001	1.2	1.7
Shuffled speed	1.7	22	0.07	0.04	1.3	1.2

## 2.4. $q$ -Gaussian formulations

In superstatistics, the marginal probability  $p(v)$  is given as, [10]

$$p(v) = \int f(\beta)p(v|\beta)d\beta. \quad (13)$$

In Eq. (13), as stated, if  $f(\beta)$  is a  $\chi^2$  distribution and the  $p(v|\beta)$  is a Gaussian one, then the  $p(v)$  becomes a Tsallis distribution.

Under  $\chi^2$  distributed  $\beta$ , the generalized canonical distributions of nonextensive statistical mechanics become, [10]

$$p(v) \sim \left(1 + \frac{1}{2}\beta^\circ(q-1)v^2\right)^{\frac{1}{1-q}} \quad (14)$$

where  $\beta^\circ$  is the inverse temperature,  $q$  is the entropic index.

Consider the  $\chi^2$  distribution Eq. (15) in the integration of Eq. (13),

$$f(\beta) = \frac{1}{\Gamma\left(\frac{n}{2}\right)} \left(\frac{n}{2\beta_0}\right)^{n/2} \beta^{n/2-1} e^{-\frac{n\beta}{2\beta_0}}. \quad (15)$$

The relation between the parameters  $q$  and  $n$  is given as, [8,10]

$$q = 1 + \frac{2}{n+1} \quad (16)$$

where  $n$  is the number of degrees of freedom.

## 3. Traffic data analyses

One of the fingerprints of superstatistics is the two distinct and well-separated time scales. So long as these time scales are well-separated, there is an indication of a well-formed superstatistics. In this section, first objective is to determine the two time scales  $T$  and  $\tau$  for the data of the vehicle speeds, the vehicle speed differences and the shuffled vehicle speeds. For those three cases, probability distributions of the inverse temperature variables are also plotted. Then the  $q$ -Gaussians are computed and the specific  $q$ -values are obtained of the time series data. The auxiliary condition is also checked for each case and the results are given on Table 1.

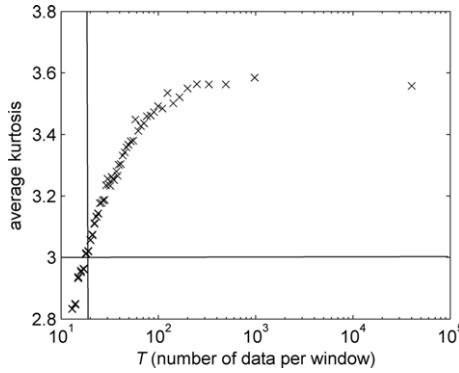
### 3.1. Analysis of vehicle speeds

The speed data have been collected from a surveillance point at Çanakkale-İzmir highway in Izmir. Approximately 37-hours' data are considered and one directional individual speed measurements are processed in our analyses. The division of the relevant time series, speed or the differences, into equally partitioned windows would reveal  $T$  through Eq. (10). The determination of the number of windows depends on the kurtosis value. Kurtosis should be roughly equal to 3 due to the stipulation that the system locally relaxes to a Gaussian distribution. The optimal time scale  $T$  of time series data for the vehicle speeds is obtained at around 18 where the kurtosis value is 3 as seen in Fig. 2. Fig. 2 also shows the other kurtosis values changing roughly between 2.8 and 3.6 for different window partitioning for our time series.

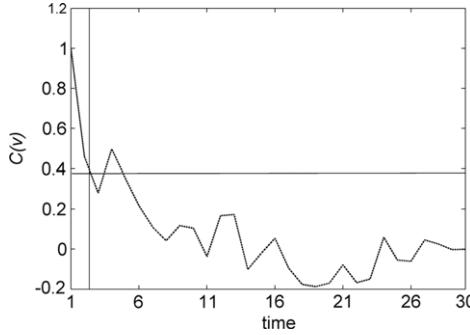
Let us determine the short time scale  $\tau$  for the vehicle speed time series data. The condition of exponentially decaying  $C_{k,\tau}(v)$  has been given in Section 2 to obtain the  $\tau$ . The  $\tau$  value is computed by using Eqs. (8) and (9). As it is displayed in Fig. 3, it corresponds to a  $\tau$  of 2.4. The initial portion of the time series is presented in Fig. 3. Please note that the correlation graph crosses  $1/e$  at two additional points later. We consider only the first crossing at 2.4.

According to our analyses, the ratio of the short time scale to the long time one is obtained to be  $\tau/T = 0.13$  for the vehicle speed time series data. The ratio is relatively small enough for a moderate proof of the theory. Thus, this partially verifies that our time series is superstatistical.

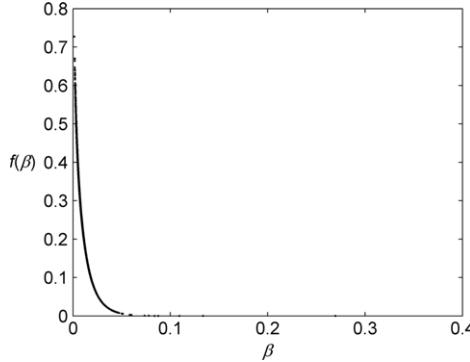
In the previous steps, we have divided the time series into windows to obtain the time scale  $T$ . Now, the number of windows corresponds to the number of inverse temperature parameters i.e. inverse local variances of the time series. Thus, the inverse temperature values for each window are calculated. The probability distribution of these values is plotted in



**Fig. 2.** Average kurtosis diagram of vehicle speeds.



**Fig. 3.** Correlation function of vehicle speeds.

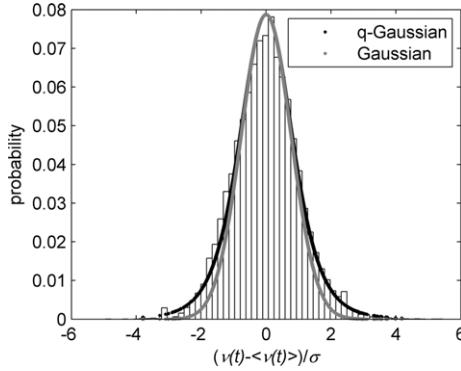


**Fig. 4.** Probability distribution of beta values for vehicle speeds.

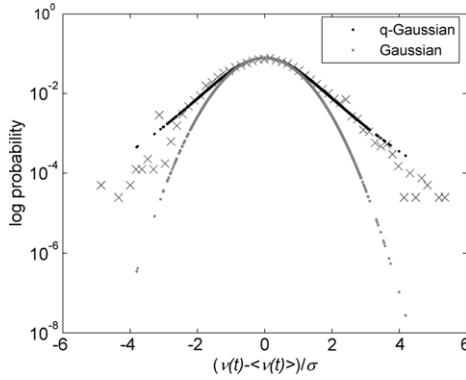
**Fig. 4.** The distribution is a  $\chi^2$ , Eq. (4), which fits well the fluctuating inverse temperature. This, in turn, follows the Tsallis  $q$ -statistics.

Hence, Eq. (14) is employed here since the parameter  $\beta$  is distributed according to  $\chi^2$ . Histogram of the whole time series data well matches the  $q$ -Gaussian form. One could see the  $q$ -Gaussian fit for the vehicle speed data in Figs. 5a and 5b. Please note that two different  $q$  parameters are calculated. One of them is acquired from the  $\chi^2$  distribution of  $\beta$  parameter and thus from the degrees of freedom Eq. (16), and the other is generated from  $q$ -Gaussian fitting. It is seen that these two values do not coincide, since one is directly related to the sequence of the measurements, whereas the other is not influenced by the sequence. We classify them in two categories i.e. time-dependent and time-independent  $q$  values.

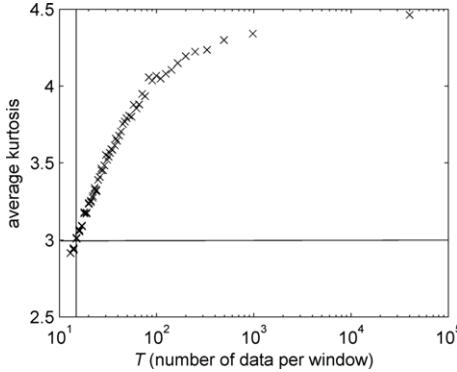
Our speed time series data exhibit  $q$ -Gaussian behaviour with the  $q$  value of 1.3. The normalized linear-linear and semi-log plots are illustrated in Figs. 5a and 5b, respectively. The semi-log plot makes tails more noticeable. In Figs. 5a and 5b, the difference from the Gaussian distribution where the  $q$  value is unity is also displayed and this difference is more pronounced in Fig. 5b. This explicitly implies that there is a deviation from complete randomness of the speeds in the traffic flow by means of correlations. It is expected that the long-range interactions, long-term memory may be observed in the traffic flow for the deviation from the Gaussianity mentioned above.



**Fig. 5a.** Histogram of the speed time series, fitting plots of Gaussian and  $q$ -Gaussian forms.



**Fig. 5b.** Semi-log plot of speed time series and fitting plots of Gaussian and  $q$ -Gaussian forms.



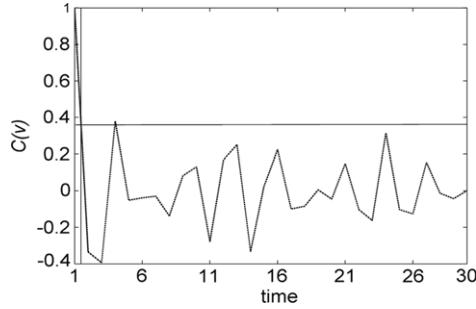
**Fig. 6.** Average kurtosis diagram of vehicle speed differences.

From the probability distribution of the inverse temperature, we obtain yet another  $q$  of 1.8. As this study points out that the difference between the  $q$  values of 1.3 and 1.8 stems from the sequence i.e. the specific time history of the time series.

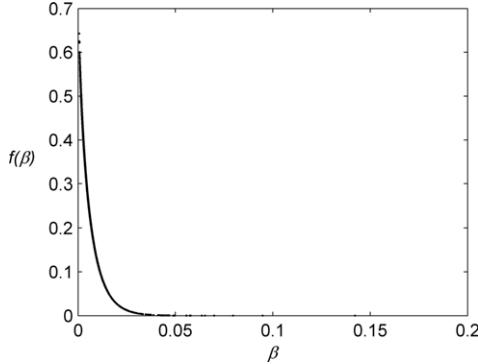
### 3.2. Analysis of vehicle speed differences

In this section, the vehicle speed differences are now processed for superstatistical analyses. As we discussed in Section 3.1, the kurtosis value is an indicator of the long time scale  $T$ . By stipulating the Gaussian approximation,  $T$  is obtained as 16 and the results are plotted in Fig. 6. The procedure for the correlation analysis is also carried out for the speed differences and the short time scale yielding a  $\tau$  of 1.5. One can check the relation between the short time scale and the correlation values in Fig. 7.

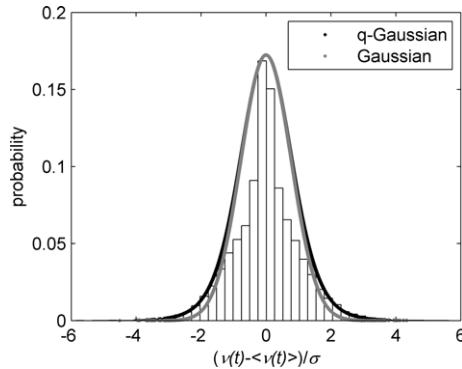
It is found that  $\tau$  is shorter than the  $T$ , as expected, fairly upholding the superstatistical condition of  $\tau \ll T$  where the ratio  $\tau/T$  is 0.09.



**Fig. 7.** Correlation function of vehicle speed differences.



**Fig. 8.** Probability distribution of beta values for vehicle speed differences.



**Fig. 9a.** Histogram of the speed difference time series, fitting plots of Gaussian and  $q$ -Gaussian forms.

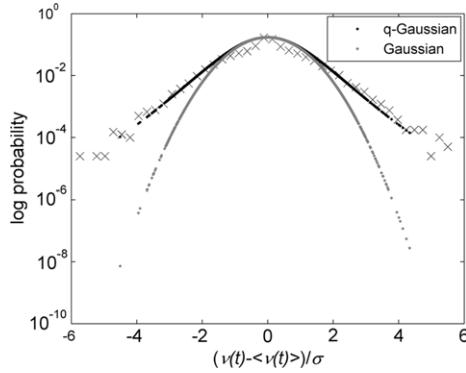
When we consider the probability distribution of the beta parameter,  $\chi^2$  distribution is again found to be the relevant one. Fig. 8 displays the fitting plot of the beta distribution extracted from the vehicle speed differences.

The speed difference time series data show  $q$ -Gaussian form with  $q$  value of 1.2. This is relatively lower than the one of the speed time series. As it is seen in Figs. 9a and 9b, linear-linear and semi-log plots of the speed differences can express the tails and the peak markedly. For example, the deviation from the Gaussian distribution could be observed in Fig. 9b with the help of the  $q$ -Gaussian. The  $\chi^2$  distribution of the inverse temperatures yields another  $q$  value of 1.7 for the speed differences. As it is expected from the previous subsection, two computed  $q$  values do not agree.

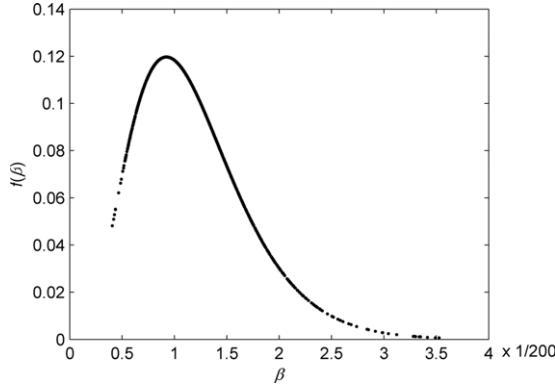
### 3.3. Influence of shuffled vehicle speed data

Early on, the authors of this study have claimed that the sequence of the time series is an important parameter in determining the so-called time-dependent  $q$  parameter. In this subsection we use the completely shuffled vehicle speed data in order to show the influence of the shuffling i.e. sequence of the data on the  $q$  parameter. This is now discussed again by virtue of the inverse local variances (see Fig. 10).

Table 1 provides a summary of all the three cases, speed, speed differences and the shuffled speed data. Note that  $\tau/T$  is a minimum at shuffled speed data, with relatively larger but still small enough an  $|\varepsilon|$  (Eq. (12)), making the superstatistical



**Fig. 9b.** Semi-log plot of speed difference time series, fitting plots of Gaussian and  $q$ -Gaussian forms.



**Fig. 10.** Probability distribution of beta values for shuffled vehicle speed.

modeling a good choice. The worst seems to be the speed data where the time scale separation is not the strongest. But a small  $|\epsilon|$  may alleviate the problem posed by the time scale separation. The thesis that the time-dependent  $q$  is a property of how the time series is realized is supported by different  $q$  values on Table 1. The very realization of the traffic flow for those certain days of data acquisition has generated a  $q$  of 1.8 out of inverse temperature fit. When shuffled, this value shrinks down to 1.2. One does not expect a change in  $q$  values from speed histograms, since the histograms are independent from how that traffic flow has occurred. Next, this issue is discussed further.

#### 4. Discussion: the coexistence of different $q$ values

The authors of this paper claim that there ought to be two distinct  $q$  values, specifying the highway segment with a certain traffic flow, and the interactions. Our analysis has shown that the inverse temperature distribution is distributed with respect to a chi-square distribution Eq. (4), allowing a mapping to a  $q$ -Gaussian distribution of vehicle speeds Eq. (14). Some work in the literature assume that once a chi-square distribution manifests itself, a fit from chi-square distribution of inverse temperature values generates a  $q$  value that could be assumed to be the same coming out of Eq. (14), which we believe is a fallacy on very simple grounds.

The  $q$  values from the fits of Eqs. (4) and (14) must be different values, pointing at different characteristics. The  $q$  parameter from the fit out of a  $q$ -Gaussian would and should ultimately reflect the highway or traffic properties, since it is a fit based on speeds (or speed differences). It has a certain histogram, and once recorded, it may be assumed to be time-independent in the sense that given speed distribution is a representation of the given highway segment. Note that no history is involved, and hence time-independent.

A fit by virtue of Eq. (4) is a distribution strictly of the inverse temperature values of segments of windows, which is highly time-dependent, that is the time history of the traffic flow. Thus the time history of the vehicle flow would affect the  $q$  value, once the histogram of the speeds is specified. Namely, when the speed distribution is given, depending on the time history of the flow, one could obtain infinitely many  $q$  values by Eqs. (4) and (16). It must be noted that, a similar mechanism albeit through the time scales, is hinted at in Ref. [13].

For this reason, we claim that the  $q$  out of Eq. (14) is time-independent parameter, reflecting the characteristics of the highway segment. The  $q$  index out of Eq. (16) is time-dependent in the sense that different flows would generate different  $\beta$ , since it is a function of the time history of the flow.

## 5. Conclusions

This paper has dealt with a superstatistical analysis of a given highway traffic flow data. It was seen that the data were comprised of two distinct time scales,  $\tau$  and  $T$ , where  $T$  ought to be sufficiently large enough for the system to reach a local equilibrium until the system dynamics requires another local settling point. The whole time series thus was divided into  $N$  windows of width  $T$ , with their respective inverse variance parameters  $\beta$ . The analyses were repeated for the speeds, the speed differences, and the shuffled speed values, given on Table 1. The windows generated inverse temperatures, whose distribution has turned out to be a chi-square one. Once this distribution is known, a  $q$ -Gaussian fit was performed first on the speeds, then the speed differences, and finally on the shuffled speed values.

As it was claimed earlier in the study by the authors the literature lacks a warning on how to approach a chi-square distribution data, once it is found to be so. Since Eq. (4) pinpoints a  $q$  value over a distribution fit, where all the related inverse temperature statistics is known. But a parallel work on the  $q$ -Gaussian fit through Eq. (14) is surmised to yield exactly the same value, which was accepted to be so in a number of papers in the literature. We claim that the  $q$  values obtained through fits of Eqs. (4) and (14) would most certainly be different. We also claim that Eq. (14) provides a  $q$  index that is a characteristic of the traffic flow, but not of its history, hence it is independent of the history of the flow. On the other hand, Eq. (4) is a function of the inverse temperatures. The statistics of the inverse temperatures is a strict function of the time history of the traffic flow, hence a characteristic of how the flow happened, making it time-dependent. As a proof of our claim, the speed data have been shuffled and reanalyzed to verify that indeed the time history affects the Tsallis  $q$  index out of superstatistical analyses. Not only Eqs. (4) and (14) have yielded different values, but finally a shuffled speed data analysis too produced a third and different  $q$  value, emphatically different from the one by Eq. (14).

As a final point, it is seen that data shuffling has reduced the variances of the windows, increasing the inverse temperature range. The shuffled data have also been found to be chi-square in nature through another fit, only that this time having a larger degrees of freedom. By virtue of Eq. (16), an increased degrees of freedom reduces the Tsallis  $q$  value, making the distribution approach a Gaussian. A series of shuffling trials always proved to be of low variance, large degrees of freedom distributions. This helps frame a tentative rule that if the shuffling i.e. randomizing the data still holds a chi-square distribution, it should in general decrease the variance over a larger degrees of freedom, always bringing down the Tsallis  $q$  value.

## References

- [1] M. Krbálek, K. Kittanová, Lattice thermodynamic model for vehicular congestions, Procedia-Soc. Behav. Sci. 20 (2011) 398–405. <http://dx.doi.org/10.1016/j.sbspro.2011.08.046>.
- [2] C. Tsallis, Possible generalization of Boltzmann–Gibbs statistics, J. Stat. Phys. 52 (1–2) (1988) 479–487. <http://dx.doi.org/10.1007/BF01016429>.
- [3] C. Tsallis, *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World*, Springer, New York, NY, 2009.
- [4] M. Krbálek, D. Helbing, Determination of interaction potentials in freeway traffic from steady-state statistics, Physica A 333 (2004) 370–378. <http://dx.doi.org/10.1016/j.physa.2003.10.059>.
- [5] M. Krbálek, Analytical derivation of time spectral rigidity for thermodynamic traffic gas, Kybernetika 46 (6) (2010) 1108–1121.
- [6] M. Krbálek, Inter-vehicle gap statistics on signal-controlled crossroads, J. Phys. A: Math. Theor. 41 (20) (2008) <http://dx.doi.org/10.1088/1751-8113/41/20/205004>.
- [7] R. Mahnke, J. Kaupuž, J. Hinkel, H. Weber, Application of thermodynamics to driven systems, Eur. Phys. J. B 57 (4) (2007) 463–471. <http://dx.doi.org/10.1140/epjb/e2007-00182-7>.
- [8] C. Beck, Dynamical foundations of nonextensive statistical mechanics, Phys. Rev. Lett. 87 (18) (2001) <http://dx.doi.org/10.1103/PhysRevLett.87.180601>.
- [9] C. Beck, E.G.D. Cohen, Superstatistics, Physica A 322 (2003) 267–275. [http://dx.doi.org/10.1016/S0378-4371\(03\)00019-0](http://dx.doi.org/10.1016/S0378-4371(03)00019-0).
- [10] C. Beck, Superstatistics: Theory and applications, Contin. Mech. Thermodyn. 16 (3) (2004) 293–304. <http://dx.doi.org/10.1007/s00161-003-0145-1>.
- [11] C. Beck, E.G. Cohen, H.L. Swinney, From time series to superstatistics, Phys. Rev. E 72 (5) (2005) 056133. <http://dx.doi.org/10.1103/PhysRevE.72.056133>.
- [12] E. Van der Straeten, C. Beck, Superstatistical fluctuations in time series: Applications to share-price dynamics and turbulence, Phys. Rev. E 80 (3) (2009) <http://dx.doi.org/10.1103/PhysRevE.80.036108>.
- [13] D. Xu, C. Beck, Transition from lognormal to chi-square superstatistics for financial time series, 2015. ArXiv Preprint <arXiv:1506.01660>.
- [14] P. Rabassa, C. Beck, Superstatistical analysis of sea-level fluctuations, Physica A 417 (2015) 18–28. <http://dx.doi.org/10.1016/j.physa.2014.08.068>.
- [15] C. Beck, Superstatistics in high-energy physics, Eur. Phys. J. A 40 (3) (2009) 267–273. <http://dx.doi.org/10.1140/epja/i2009-10792-7>.
- [16] K. Briggs, C. Beck, Modelling train delays with q-exponential functions, Physica A 378 (2) (2007) 498–504. <http://dx.doi.org/10.1016/j.physa.2006.11.084>.
- [17] S. Rizzo, A. Rapisarda, Environmental atmospheric turbulence at Florence airport, in: The 8th Experimental Chaos Conference: AIP Conference Proceedings, Italy, 2004. <arXiv:cond-mat/0406684>, <http://dx.doi.org/10.1063/1.1846475>.
- [18] G.C. Yalcin, C. Beck, Environmental superstatistics, Physica A 392 (21) (2013) 5431–5452. <http://dx.doi.org/10.1016/j.physa.2013.06.057>.
- [19] L.L. Chen, C. Beck, A superstatistical model of metastasis and cancer survival, Physica A 387 (13) (2008) 3162–3172. <http://dx.doi.org/10.1016/j.physa.2008.01.116>.
- [20] G.C. Yalcin, C. Beck, Currents in complex polymers: An example of superstatistics for short time series, Phys. Lett. A 376 (35) (2012) 2344–2347. <http://dx.doi.org/10.1016/j.physleta.2012.05.057>.
- [21] C. Nie, H. Lu, Y. Shi, A headway distribution model for urban expressway based on superstatistics, Highw. Eng. 01 (2010).
- [22] A.Y. Abul-Magd, Modeling highway-traffic headway distributions using superstatistics, Phys. Rev. E 76 (5) (2007) 057101. <http://dx.doi.org/10.1103/PhysRevE.76.057101>.