



## 3s-Unification for Vehicular Headway Modeling

Milan Krbálek<sup>a</sup> and Michaela Krbálková<sup>b,c</sup>

<sup>a</sup> Department of Mathematics, Faculty of Nuclear Science and Physical Engineering, Czech Technical University in Prague, 12000 Prague 2, Czech Republic

<sup>b</sup> Department of Physics, Faculty of Science, University of Hradec Králové, Hradec Králové, Czech Republic

<sup>c</sup> Department of Transport Technology and Control, Jan Perner Transport Faculty, University of Pardubice, Pardubice, Czech Republic

Email: milan.krbalek@fjfi.cvut.cz

**Abstract.** We explain why *a sampling* (division of data into homogenous sub-samples), *segmentation* (selection of sub-samples belonging to a small sub-area in ID plane – a segmentation zone), and *scaling* (a linear transformation of random variables representing a standard subroutine in a general scheme of an unfolding procedure) are necessary parts of any vehicular data investigations. We demonstrate how representative traffic micro-quantities (in an unified representation) are changing with a location of a segmentation zone. It is shown that these changes are non-trivial and correspond fully to some previously-published results. Furthermore, we present a simple mathematical technique for the unification of GIG-distributed random variables.

**Key words:** Vehicular Headway, Data Segmentation, Data Unification, Generalized Inverse Gaussian Distribution, GIG.

## 1 Introduction and motivation

Vehicular dynamics [1] is a complex scientific discipline having many interesting components. One of them is *Vehicular Headway Modeling (VHM)* [2] analyzing and predicting changes in a vehicular microstructure forced by external conditions (traffic density, intensity, or global velocity). Typical representatives of vehicular micro-quantities are individual velocities, time/space headways, or time/space clearances. Mathematically, all these quantities represent, in fact, random variables and associated probability densities (for headways and clearances) belonging to a specific family of distributions (see for example [3]). As is well known (see [4, 5, 6, 7, 8]), parameters of these distributions evolve rapidly over time and are therefore markedly dependent on actual values of density  $\rho$ , intensity  $I$ , and global velocity  $V$ . It excludes the possibility of applying standard statistical approaches to entire data structures and, on contrary, it enforces the use of more complex

procedures applied to partial data samples. The description of these methods is the main goal set for this article.

## 2 Typical sets of empirical traffic data

There exist many technologies suitable for vehicle-by-vehicle measurements. They are typically divided into two categories: *intrusive* (induction loops, piezo-electric cables, active infrared sensors, etc) and *non-intrusive* (passive infrared sensors, detection drones, ultrasonic sensors, video image processing, etc). However, if one aims to deal with Big Data sets the most of these methods are inappropriate since they provide a limited amount of data only. In contrast, commonly available magnetic induction double-loop detectors (as an example of intrusive measurement techniques) do not suffer from this inefficiency. Therefore, in the rest of the text we consider data gauged by magnetic loops or by similar technology.

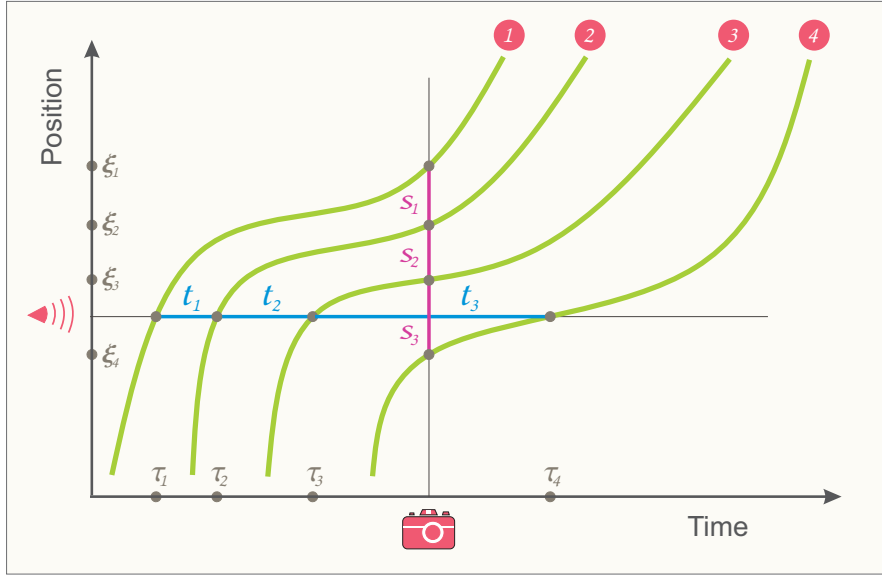


Figure 1: An illustration of time and space headways on the space-time plot. Each curve symbolizes a trajectory of a front car-bumper.

Typical attribute of those data is that they have been measured at the same time-fixed location (a so-called *detector line*). Considering now one-lane traffic flow (without loss of generality) it is easy to comprehend that characteristic outputs of a traffic measurement look like

$$T^{(\text{in})} = \{\tau_k^{(\text{in})} \in \mathbb{R} \mid k \in \hat{N} \wedge \tau_{k-1}^{(\text{in})} < \tau_k^{(\text{in})} \wedge \tau_0^{(\text{in})} := 0\}, \quad (1)$$

$$T^{(\text{out})} = \{\tau_k^{(\text{out})} \in \mathbb{R} \mid k \in \hat{N} \wedge \tau_{k-1}^{(\text{in})} < \tau_{k-1}^{(\text{out})} \leq \tau_k^{(\text{in})} < \tau_k^{(\text{out})} \wedge \tau_0^{(\text{out})} := 0\}, \quad (2)$$

$$\Upsilon = \{v_k \in \mathbb{R}_0^+ \mid k \in \hat{N}\}, \quad (3)$$

$$\Lambda = \{\ell_k \in \mathbb{R}^+ \mid k \in \hat{N}\}, \quad (4)$$

where  $T^{(\text{in})}$  and  $T^{(\text{out})}$  collect instants  $\tau_k$  when a front/rear bumper of a  $k$ th car has intersected a detector line and  $\Upsilon$  and  $\Lambda$  collect individual velocities  $v_k$  and lengths  $\ell_k$  of cars, respectively.

If loop measurements are accompanied by image processing technology then additional data sets

$$\Xi^{(\text{front})} = \{\xi_k^{(\text{front})} \in \mathbb{R} \mid k \in \hat{\mathbf{N}} \wedge \xi_k^{(\text{front})} < \xi_{k-1}^{(\text{front})} \wedge \xi_0^{(\text{front})} := +\infty\}, \quad (5)$$

$$\Xi^{(\text{rear})} = \{\xi_k^{(\text{rear})} \in \mathbb{R} \mid k \in \hat{\mathbf{N}} \wedge \xi_k^{(\text{rear})} < \xi_k^{(\text{front})} \leq \xi_{k-1}^{(\text{rear})} < \xi_{k-1}^{(\text{front})} \wedge \xi_0^{(\text{rear})} := +\infty\}, \quad (6)$$

(collecting positions of front/rear bumpers at the fixed time) are to disposal. However, such a doubled measurements are very rare, which results in the fact that locations  $\xi_k$  (in contrast to instants of time  $\tau_k$ ) belong to indirectly determined traffic quantities. Therefore, we refer instants of time  $\tau_k$ , velocities  $v_k$ , and lengths  $\ell_k$  as *primary quantities*, whereas locations  $\xi_k$  are referred to as *secondary quantities*.

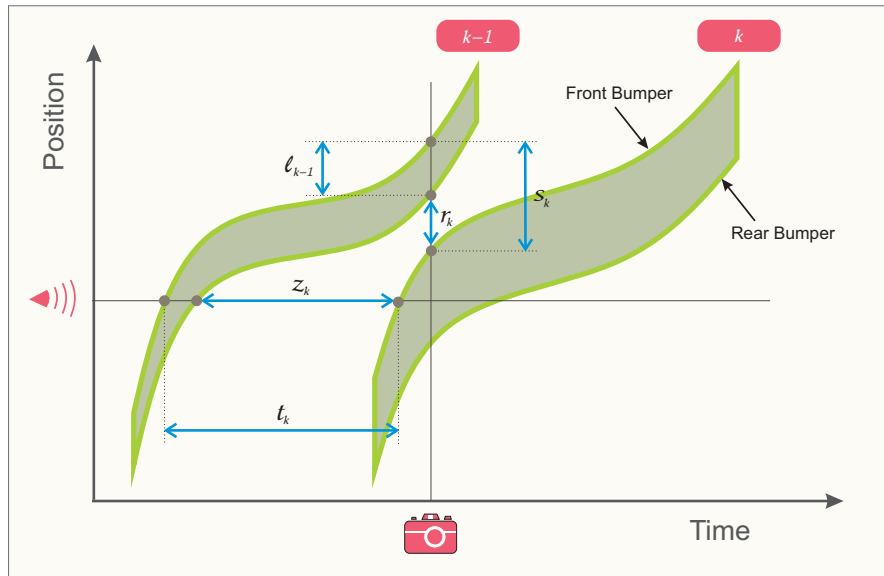


Figure 2: Individual vehicular characteristics obtained by cross-section detectors, e.g. induction double-loop detectors,  $(z_k, t_k)$  or by image processing technology  $(r_k, s_k)$ .

### 3 Random variables in VHM

Less formally (see for example [2]), in Vehicular Headway Modeling (VHM) a time headway is usually defined as the time between two successive vehicles as they pass a fixed point on the same lane. Analogously, a space headway is understood as distance between the same common feature (e.g., front/rear bumper) of two subsequent vehicles measured at a fixed time. Except headways, VHM often works with vehicular clearances. Time clearance represents the time between two following events: 1) rear bumper of a previous car intersects a detector line; 2) front bumper of a reference car intersects a detector line. Distance clearance is, in analogy, defined as a space gap between vehicles measured at

a fixed time. Illustratively, all these quantities are visualized in figures 1 and 2. More formally, they will be defined in the following subsection.

### 3.1 Standard typology of empirical headways

Now, knowing typical outputs of traffic measurements one can define empirical values of *time headways/clearances* using respective formulas

$$t_k := \tau_k^{(\text{in})} - \tau_{k-1}^{(\text{in})}, \quad z_k := \tau_k^{(\text{in})} - \tau_{k-1}^{(\text{out})}.$$

Distance headways are then defined by

$$s_k := \xi_{k-1}^{(\text{front})} - \xi_k^{(\text{front})}$$

and distance clearances read

$$r_k := \xi_{k-1}^{(\text{rear})} - \xi_k^{(\text{front})}.$$

Note that equalities  $s_k = v_k t_k$ ,  $r_k = v_k z_k$  are valid only provided that an individual velocity  $v_k$  is constant during time interval  $[\tau_{k-1}^{(\text{in})}, \tau_k^{(\text{in})}]$ , which is less probable assumption, especially when  $\tau_k^{(\text{in})} \gg \tau_{k-1}^{(\text{in})}$ . It means that loop measurements (without additional processing of video/photography) generate time clearances and time headways as primary variables, whereas distance headways/clearances (taking into account the method of data collection) represent secondary variables and their approximative values  $s_k \approx v_k t_k$ ,  $r_k \approx v_k z_k$  can be therefore burdened by systematic errors. Moreover, one finds

$$r_k = \xi_{k-1}^{(\text{rear})} - \xi_{k-1}^{(\text{front})} + \xi_{k-1}^{(\text{front})} - \xi_k^{(\text{front})} = s_k - \ell_{k-1}.$$

### 3.2 Stochastic representation of vehicular headways

Mathematically, all the quantities from the subsection 3.1 represent non-negative continuous random variables and are therefore characterized by a standard statistical description using associated probability densities and distribution functions (cumulated probability densities). Consider now a sequence  $(\mathcal{X}_k)_{k=1}^N$  of random variables of the same type (clearance, headway). From a statistical viewpoint the empirical headways  $x_k$  represent individual realizations of random variables  $\mathcal{X}_k$  and one can model respective distributions by standard statistical routines. Generally accepted premise in VHM says that  $\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_N$  are identically distributed provided that one analyzes homogeneous flows, where macroscopic quantities (*state variables*) are steady in time and a rate of long vehicles is low. To a certain degree of simplification it used to be sometimes speculated that  $\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_N$  are i.i.d. (e.g. [14]), which is definitely not a reasonable hypothesis. However, in many analytical studies this represent a useful (and simplifying) assumption. In fact, correlations among headways are significant [15] and reveal a more complex interaction rules among drivers. This is, unfortunately, beyond the scope of this article.

## 4 Macroscopic characteristics of traffic

Usually, three basic phase variables (density, intensity and mean speed), describing a macroscopic constellation of each vehicular ensemble, are defined somewhat vaguely (see

pages 15-17 in [1]). Intensity  $I(x, \tau) = \Delta N / \Delta \tau$  is understood as the number of vehicles  $\Delta N$  passing a given cross-section at locations  $x$  within a time interval  $\Delta \tau$ , whereas density  $\varrho(x, \tau) = \Delta N / \Delta x$  represents the number of vehicles  $\Delta N$  lying in an interval  $[x, x + \Delta x]$  during a fixed time  $\tau$ . Especially these two definitions are difficult to interpret since number of vehicles is a discrete quantity not allowing standard differentiations. Therefore, it is necessary to go to a mathematically correct formulation.

## 4.1 Theoretical definitions

Consider a population of  $M$  succeeding (dimensionless) vehicles located in a fixed time  $\tau$  at locations  $\alpha_1(\tau) > \alpha_2(\tau) > \dots > \alpha_M(\tau)$ . Let  $\wp(x|\sigma) \in \mathcal{C}^2(\mathbb{R})$  be an arbitrary probability density having 0 as an expected value and  $\sigma$  as a shape-parameter. Typical representatives of generating densities are summarized in Appendix 7.1. Then, after a selection of a suitable generating density, one can define *smoothed number of particles* as

$$N(\xi, \tau) := \int_{-\infty}^{\xi} \sum_{k=1}^M \wp(y - \alpha_k(\tau) | \sigma) \, dy. \quad (7)$$

Using such a function the *density* and *intensity* can be defined as

$$\varrho(\xi, \tau) := \frac{\partial N(\xi, \tau)}{\partial \xi}, \quad I(\xi, \tau) := -\frac{\partial N(\xi, \tau)}{\partial \tau}, \quad (8)$$

respectively. Moreover, both can be simplified into following forms:

$$\varrho(\xi, \tau) = \sum_{k=1}^M \wp(\xi - \alpha_k(\tau) | \sigma), \quad (9)$$

$$I(\xi, \tau) = \sum_{k=1}^M \frac{d\alpha_k(\tau)}{d\tau} \wp(\xi - \alpha_k(\tau) | \sigma) = \sum_{k=1}^M v_k(\tau) \wp(\xi - \alpha_k(\tau) | \sigma). \quad (10)$$

Owing to the prerequisite  $\wp(\xi | \sigma) \in \mathcal{C}^2(\mathbb{R})$  we have

$$\frac{\partial \varrho}{\partial \tau} = \frac{\partial^2 N(\xi, \tau)}{\partial \xi \partial \tau} = \frac{\partial^2 N(\xi, \tau)}{\partial \tau \partial \xi} = -\frac{\partial I}{\partial \xi}, \quad (11)$$

which corresponds to the well-known *equation of continuity*, here interpreted as the law of conservation for the number of vehicles. If we restrict our considerations to homogeneous flows only, where  $v_k(\tau) = V(\tau)$ , then we obtain a hydrodynamic equality

$$I(\xi, \tau) = V(\tau) \varrho(\xi, \tau), \quad (12)$$

that usually serves itself as a generally accepted approximation for a fundamental relationship among three basic traffic state variables. However, always it is necessary to keep in mind that equation  $I(\xi, \tau) = V(\xi, \tau) \varrho(\xi, \tau)$ , contrary to equation (11), does not represent a universally valid formula.

## 4.2 Fundamental phase relations in vehicular traffic

In physics of traffic, *fundamental diagrams* are understood as graphical visualizations of relationship among the three basic traffic macro-quantities (phase variables). In earlier works (see review [9]) authors started from the proposition that traffic intensity  $I$  and density  $\rho$  are interconnected by a certain functional dependence  $I = I(\rho)$ , whose graphical interpretation has a shape of a curve  $H(I, \rho) = 0$  lying within ID plane. Analogously, it was speculated that  $V = V(\rho)$  is a function and its graph is a curve. Kerner's three-phase theory [10] and his hypothesis about two-dimensional states of traffic flow, however, disproved this fact by the following claim:

*Consider a homogeneous synchronized flow, that is here understood as a hypothetical state of synchronized flow of identical vehicles and identical drivers in which all vehicles move with the same time-independent speed (see the yellow curve in figure 3) and have the same space gaps. Surprisingly, these homogeneous synchronized flows can occur anywhere in some two-dimensional sub-region of the ID plane. It means that at a given speed in congested flow, a driver does not use some unique gap. Quite the contrary, a driver can make an arbitrary choice for the space clearance to a preceding vehicle (see figure 4), which leads to the fact that (although a speed of the vehicles is constant) the set of admissible states (described by intensity and density) is infinite. Real-road traffic states with the same average speed therefore cover, in fact, a certain two-dimensional area lying inside the ID plane.*

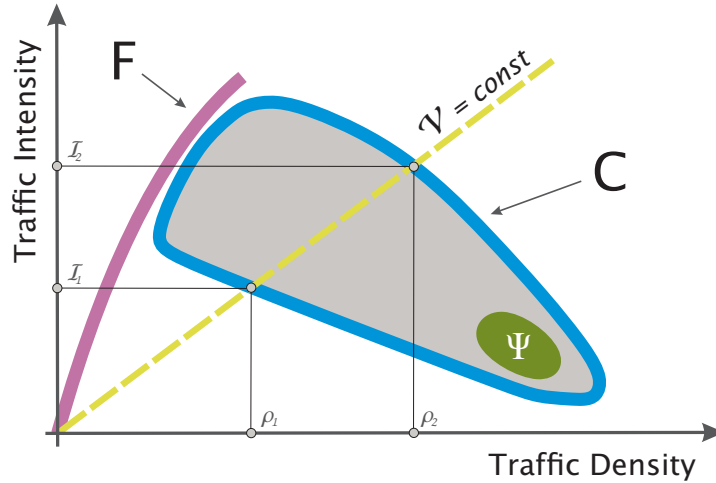


Figure 3: Schematic visualization of fundamental relations. Curve  $F$  represents *free-flow traffic*, whereas blue area  $C$  corresponds to *congested traffic*, where one can reveal (using an analysis of real-time trajectories) two sub-regimes: *synchronized flow* or *wide moving jam* (see [10] for details). Note that, even a velocity of cars is almost constant (i.e. near a yellow line), set of admissible densities covers the interval  $[\rho_1, \rho_2]$  and set of admissible intensities covers the interval  $[I_1, I_2]$ . Such behavior is in a direct contradiction with theories involving the fundamental diagram of traffic flow, which suppose a one-dimensional relationship between density and intensity. Green surface represents a selected segment (subset) of ID plane (a segmentation zone).

Mathematically, it means that, instead of a functional relationship  $I = I(\rho)$ , we need to switch to a mathematical description using a term of a *binary relation*, which, generally

speaking, represents an arbitrary set of ordered pairs. Thus, for vehicular applications a *binary ID relation*  $\Omega_{ID}$  is defined a set of all pairs  $(\varrho, I) \in (0, +\infty)^2$ , which can be gauged for a given vehicular stream. Then *the domain* of  $\Omega_{ID}$  is the set of all  $\varrho$  such that  $(\varrho, I) \in \Omega_{ID}$  for at least one  $I$ . *The range* is the set of all  $I$ , for which exists at least one  $\varrho$  such that  $(\varrho, I) \in \Omega_{ID}$ . Thus,  $\Omega_{ID} \subset \text{Dom}(\Omega_{ID}) \times \text{Ran}(\Omega_{ID})$ . In a similar way, the *a binary VD relation*  $\Omega_{VD}$  between mean speed and density is introduced.

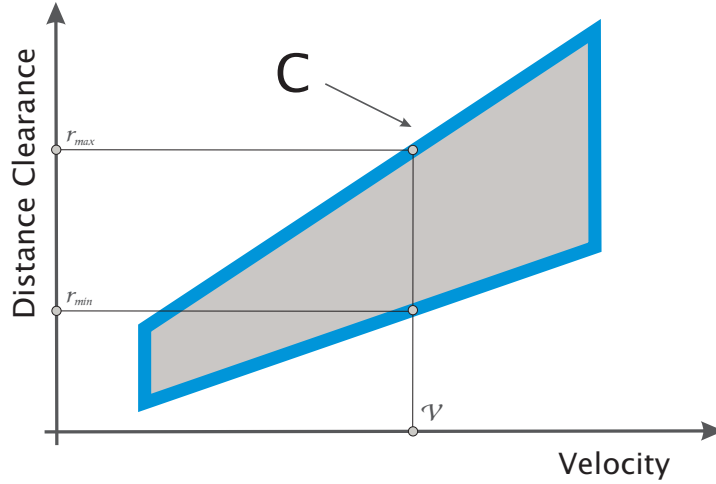


Figure 4: Schematic visualization of Kerner's hypothesis. Steady states of a homogeneous synchronized flow are spread on two-dimensional region.

### 4.3 Empirical extractions of fundamental binary relations

The way to get a graphical shape of the fundamental relations (*a phase diagram*) from empirical data is not difficult. For a given sample

$$\Xi = \left\{ (\tau_k^{(\text{in})}, \tau_k^{(\text{out})}, v_k, \ell_k) \in T^{(\text{in})} \times T^{(\text{out})} \times \Upsilon \times \Lambda : k = 1, 2, \dots, M \right\}. \quad (13)$$

of  $M$  succeeding cars one can calculate the local intensity  $I = M / (\tau_M^{(\text{out})} - \tau_1^{(\text{in})})$  and the local mean speed  $V = M^{-1} \sum_{k=1}^M v_k$ . Besides, the expression  $\varrho = I/V$  used to be (in accordance with the relationship (12)) usually accepted as a plausible approximation for the local density (according to [7, 9]). Then, applying the same computational routine for all available data samples  $i = 1, 2, \dots, m$  we acquire empirical binary relations  $\Omega_{ID} = \{(\varrho_i, I_i) : i = 1, 2, \dots, m\}$  and  $\Omega_{VD} = \{(\varrho_i, V_i) : i = 1, 2, \dots, m\}$ , whose graphs can be used for synoptic visualizations of macroscopic properties of traffic flow.

## 5 3s-Unification

Specific and well-known signs of traffic are (see [9]): a strong non-linearity of congested states, chaotic evolution of state variables, repetitive sharp increases of density, and a propagation of kinematic waves in a direction opposite to vectors of vehicular velocities.

All these effects cause that larger samples of succeeding vehicles show a significant non-homogeneities. Their microstructure is therefore also non-homogeneous which results in the fact that associated probability distributions are not one-component ones and produce, in contrast, courses typical for mixed systems composed from several different distributions.

To obtain homogeneous characteristics one has to apply the following 3s-unification procedure that prevents an undesirable mixing of states with different statistical properties (like resistivity, stochastic rigidity, or compressibility), different vehicular properties (headways, clearances, velocities) and psychological properties (vigilance of drivers, reaction times, decision-making strain).

## 5.1 Sampling

The first sub-routine in a three-phase unification procedure is *the sampling*, i.e. division of data into homogeneous samples of several neighboring cars.

Consider the data sets (1)–(4), a sampling size  $M$ , and number of samples  $m$ . Without loss of generality, we assume that  $mM = N$ . For each sample  $i \in \{1, 2, \dots, m\}$  we denote  $G_i = \{(i-1)M + 1, (i-1)M + 2, \dots, iM\}$  a respective index set and extract relations  $\Omega_{\text{TD}}^{(i)}$  and  $\Omega_{\text{VD}}^{(i)}$  as introduced in subsection 4.3. Moreover, we define sample-adjoint sets of individual headways (time and spatial)  $T_i := \{t_k : k \in G_i\}$  and  $S_i := \{s_k : k \in G_i\}$ , sets of individual clearances (time and spatial)  $Z_i := \{z_k : k \in G_i\}$  and  $R_i := \{r_k : k \in G_i\}$ , and sets of velocities  $\Upsilon_i := \{v_k : k \in G_i\}$  and lengths  $\Lambda_i := \{\ell_k : k \in G_i\}$ . From this approach it follows that the  $i$ th sample is described by the sample-adjoint values  $\varrho_i, I_i, V_i$  and by the random sets

$$T_i, S_i, Z_i, R_i, \Upsilon_i, \Lambda_i. \quad (14)$$

## 5.2 Scaling

As understandable, the mean values  $\langle T_i \rangle, \langle S_i \rangle, \dots, \langle \Lambda_i \rangle$  of the random data sets (14) can be easily enumerated by means of the values  $\varrho_i, I_i, V_i$ . Indeed, it holds

$$\langle T_i \rangle = 1/I_i; \quad \langle S_i \rangle = 1/\varrho_i; \quad \langle \Upsilon_i \rangle = V_i \quad (15)$$

and  $\langle R_i \rangle = \langle S_i \rangle - \langle \Lambda_i \rangle$ . It means that for traffic micro-quantities the statistical characteristics of the first order are hidden in relations  $\Omega_{\text{TD}}$  and  $\Omega_{\text{VD}}$ . Therefore, a scaling of individual micro-quantities represents no loss of information. Above that, the standard unfolding procedure (usually applied in many statistical studies aiming to reveal a non-trivial stochastic universality like in Random Matrix Theory [11, 12]) includes a scaling procedure as its integrated part. For example, by transition to the same expected value, two different random variables can be identified as identically distributed or as variables belonging to the same one-parametric distribution family where the one and only parameter rules a respective variance.

Mathematically, a scaling is understood as a special variant of a general *affine transformation* mapping a random variable  $\mathcal{X}$  (here understood as a non-negative and continuous random variable) into new random variable  $\mathcal{Y} = a\mathcal{X} + \mu$ , where  $a > 0$  and  $\mu = 0$ . As



simply follows from elementary chapters of theory of probability, if  $g(x)$  is a probability density associated with  $\mathcal{X}$  then

$$h(y) = \frac{1}{a} g\left(\frac{y - \mu}{a}\right) \quad (16)$$

is a density associated with  $\mathcal{Y} = a\mathcal{X} + \mu$ . Thus,

$$\mathbf{E}(\mathcal{Y}) = a\mathbf{E}(\mathcal{X}) + \mu, \quad \mathbf{VAR}(\mathcal{Y}) = a^2\mathbf{VAR}(\mathcal{X}). \quad (17)$$

Let  $\mathcal{X}$  be a non-negative and continuous variable with expected value  $\mathbf{E}(\mathcal{X}) > 0$ . Then we define a *scaling (scaling transformation)* as the affine transformation  $\mathcal{Y} = a\mathcal{X}$ , for which  $a = 1/\mathbf{E}(\mathcal{X})$ . Equalities (17) are resulting in  $\mathbf{E}(\mathcal{Y}) = 1$  and  $\mathbf{VAR}(\mathcal{Y}) = \mathbf{VAR}(\mathcal{X})/\mathbf{E}^2(\mathcal{X})$ . It means that the scaling transformation maps all random variables into variables having the unit expected value. For empirical/experimental data the scaling should be applied as follows. Traffic micro-quantities (clearances or headways) are converted to associate scaled alternatives. Such a conversion is here demonstrated on the example of the sample-adjoint set  $Z_i := \{z_k : k \in G_i\}$  of time clearances. Associate set  $Y_i := \{y_k : k \in G_i\}$  of scaled clearances is calculated using a definition

$$y_k = \frac{z_k \cdot \mathbf{M}}{\sum_{k \in G_i} z_k} = \frac{z_k}{\langle Z_i \rangle}; \quad (k \in G_i) \quad (18)$$

which ensures that  $\langle Y_i \rangle = 1$ . In analogy, we define scaled spatial clearances by

$$x_k = \frac{r_k \cdot \mathbf{M}}{\sum_{k \in G_i} r_k} = \frac{r_k}{\langle R_i \rangle}; \quad (k \in G_i). \quad (19)$$

### 5.3 Segmentation

The final step of a three-phase unification procedure is forced by the fact that most of traffic variables are significantly changing if the state variables vary. It means that random variable characteristics of the first, second, third, and fourth order (average, variance, skewness, and kurtosis, respectively) strongly depend on actual values of traffic macro-quantities. Naturally, a mixing of different traffic states (i.e. states with different values of phase variables) is significantly undesirable. For this reason, one has to analyze data from small sub-area of a phase diagram only. To be specific, denoting  $\Psi$  an arbitrary subset (a small, typically) of the  $\text{Dom}(\Omega_{\text{ID}}) \times \text{Ran}(\Omega_{\text{ID}})$  (see the green surface in figure 3 as an example) the *segmentation procedure* selects those samples having phase relations belonging to  $\Psi$ . Therefore, we introduce a  $\Psi$ -adjoint index set (referred to as *segmented index set*)

$$F_\Psi := \{i = 1, 2, \dots, m : (\varrho_i, I_i) \in \Psi\} = \{i = 1, 2, \dots, m : \Omega_{\text{ID}}^{(i)} \subset \Psi\}. \quad (20)$$

Then, statistical analysis intended is performed separately for scaled micro-quantities

$$Y_\Psi \times X_\Psi \times \Upsilon_\Psi := \{(y_k, x_k, v_k) \in Y_i \times X_i \times \Upsilon_i : i \in F_\Psi \wedge k \in G_i\} \quad (21)$$

extracted from a phase segment  $\Psi$ .

## 5.4 Unification procedure: less formally

The entire data file is divided into small samples of  $M$  successive vehicles (*a sampling stage*). Then, in all these samples, associate random variables (headways, clearances) are scaled so that in every sample the mean value is equal to one (*a scaling stage*). For each sample the phase variables (density, intensity, and mean speed) are calculated. A small phase segment  $\Psi$  in ID plane is chosen. Samples whose phase variables lie outside this segment are eliminated from further data processing. It means that statistical distributions of headways/clearances/individual velocities are analyzed (and estimated) for almost homogeneous data belonging to a small sub-region located within a graph of the fundamental phase relation  $\Omega_{ID}$ , or  $\Omega_{VD}$ , alternatively (*a segmentation stage*).

## 6 Statistical analysis of unified traffic data

In this section we analyze 3s-unified traffic sequences (21) obtained from induction double-loop measurements performed at the Expressway R1 (also called the Prague Ring) in Prague, the Czech Republic.

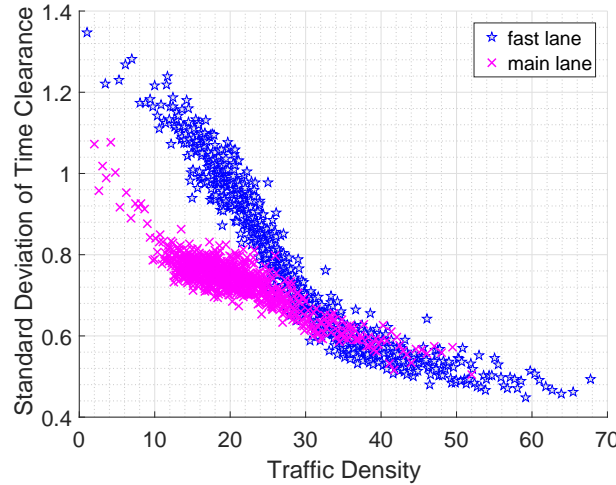


Figure 5: Standard deviation  $SD(Y_\Psi)$  of time clearance as a function of traffic density.

In order to illustrate typical outputs of statistical analysis applied to 3s-unified data we carry out the following procedure. For each sample  $Y_i$  (and  $V_i$ ) of  $M = 50$  succeeding values of scaled time clearances (and velocities) the pair  $(\varrho_i, I_i)$  is calculated. Consecutively, these samples are arranged by density in ascending order so that  $\varrho_i < \varrho_{i+1}$ . Then we split the ordered samples into 779 sets (each having  $m = 30$  samples, i.e. 1500 individual data). It means that  $\Psi_k = [\varrho_k, \varrho_k + \Delta_k) \times (0, +\infty)$ , where  $\Delta_k$  is determined so that  $|Y_{\Psi_k}| = |V_{\Psi_k}| = 1500$  and  $\varrho_{k+1} = \varrho_k + \Delta_k$ . This allows to enumerate (for each set) the standard deviation  $SD(Y_{\Psi_k})$ ,  $SD(V_{\Psi_k})$  of unified time clearances and non-scaled velocities. Results of the analysis are depicted in figures 5 and 6.

In the figure 5 we detect an extremely surprising behavior of standard deviation  $SD(Y_\Psi)$ . Although the theory of one-dimensional vehicular gases (driven by repulsive potentials and influenced by non-zero level of stochastic resistivity - see [4]) asserts that

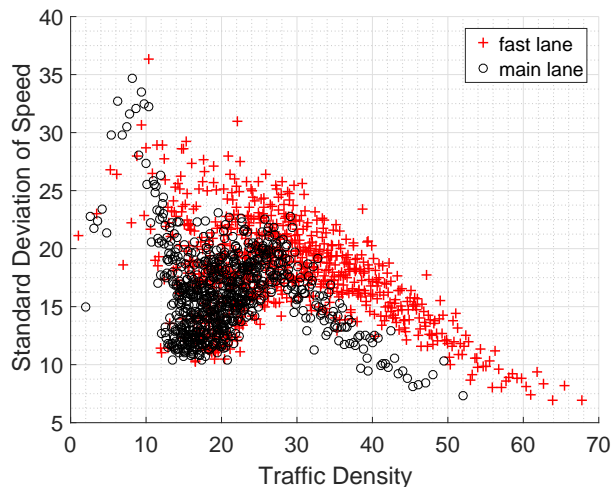


Figure 6: Standard deviation  $SD(V_\Psi)$  of speed as a function of traffic density.

scaled inter-vehicular gaps can show variances smaller than one (see [5]), the standard deviation  $SD(Y_\Psi)$  enumerated for vehicles moving in an overtaking lane of a two-lane freeway violates this theoretical restriction. It means that a simulation concept (using thermodynamical approaches applied to particle ensembles with a strict repulsion among particles) is not suitable for simulations of movements in an overtaking lane (for low traffic densities). Furthermore, differences between stars and crosses in figure 5 clearly confirms the general opinion that movements of vehicles in fast lane, in contrast to vehicles moving in main lane, are driven by different rules/forces/potentials. This knowledge represents a very interesting open problem in physics of traffic.

## 7 Conclusions

We introduce the 3s-unification procedure whose application is eligible if analyzing any vehicle-by-vehicle data by means of statistical methodology. We explain why all the three stages of this unifications (sampling, segmentation, and scaling) are requisite for correct understanding of statistical properties in vehicular microstructure. We mathematically formalize a structure and description of typical traffic data and provide a theoretically correct alternative for a definition of elementary macroscopic quantities in physics of traffic and precise/approximate relations among them.

Finally, we show how sample standard deviations  $SD(Y_\Psi)$ ,  $SD(V_\Psi)$  of unified time clearances  $Y_\Psi$  and non-scaled velocities  $V_\Psi$ , respectively, are changing with a location of a phase segment  $\Psi$ . Moreover, a thorough statistical analysis of standard deviations reveals a new open problem in the field VHM.

## Acknowledgements

Research presented in this work has been supported by the Grant SGS18/188/OHK4/-3T/14 provided by the Ministry of Education, Youth, and Sports of the Czech Republic (MŠMT ČR). The authors would also like to thank The Road and Motorway Directorate

of the Czech Republic (Ředitelství silnic a dálnic ČR) for providing traffic data analyzed in this paper.

## References

- [1] Treiber, M., Kesting, A., 2013. *Traffic Flow Dynamics*, Berlin: Springer.
- [2] Li, L., Chen, X.M., 2017. *Transportation Research Part C* 76, 170.
- [3] Krbálek, M., Hrabák, P., Bukáček, M., 2018, *Physica A* 490, 38.
- [4] Krbálek, M., 2007. *Equilibrium distributions in a thermodynamical traffic gas*, J. Phys. A: Math. Theor. 40, 5813.
- [5] Krbálek, M., Šeba, P., 2009. *Spectral rigidity of vehicular streams (Random Matrix Theory approach)*, J. Phys. A: Math. Theor. 42, 345001.
- [6] Krbálek, M., 2010. *Analytical derivation of time spectral rigidity for thermodynamic traffic gas*, *Kybernetika* 46(6), 1108.
- [7] Krbálek, M., 2013. *Theoretical predictions for vehicular headways and their clusters*, J. Phys. A: Math. Theor. 46, 445101.
- [8] Krbálek, M., Šleis, J., 2015. *Vehicular headways on signalized intersections: theory, models, and reality*, J. Phys. A: Math. Theor. 48, 015101.
- [9] Helbing, D., 2001. *Traffic and related self-driven many-particle systems*, *Rev. Mod. Phys.* 73, 1067.
- [10] Kerner, B.S., 2004. *The Physics of Traffic*, Springer-Verlag, New York.
- [11] Mehta, M.L., 2004. *Random matrices (Third Edition)*, New York: Academic Press.
- [12] Krbálek, M., Hobza, T., 2016. *Inner structure of vehicular ensembles and random matrix theory*, *Physics Letters A* 380 (21), 1839.
- [13] Borsalino is the oldest Italian company specializing in the manufacture of luxury hats.
- [14] Cowan, R.J., 1975. *Transp. Res.* 9, 371.
- [15] Krbálek, M., Apeltauer, J., Apeltauer, T., Szabová, Z., 2018. *Physica A* 491, 112.

## Appendix

### 7.1 Generating densities

To build the smoothed number of particles (7) it is necessary to select a suitable generating density. For these purposes it can be used the most usual generator, namely probability density

$$\wp(x|\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$$

of normal-distributed random variable. This Gaussian generator, however, has not a bounded support, which is somewhat unfavorable in vehicular applications. Instead, we can use more appropriate function family derived from the probability density

$$\omega(x) = \frac{1}{Z} \Theta(1 - |x|) e^{\frac{1}{x^2-1}} \quad (22)$$

where  $Z \approx 0.4439940$  and  $\Theta(x)$  stands for the Heaviside unit-step function. Then for  $\sigma > 0$  the function

$$B(x|\sigma) := \frac{1}{\sigma} \omega\left(\frac{x}{\sigma}\right)$$

is called *Borsalino function* (see figure 7 and a note [13]) centered to the point  $x = 0$  and shaped by the parameter  $\sigma$ . In general, a generating function  $\wp(x|\sigma)$  can be chosen arbitrarily from the class of functional densities centered to the origin, provided that  $\wp(x|\sigma) \in C^2(\mathbb{R})$ . For the both above-mentioned variants it holds

$$\lim_{\sigma \rightarrow 0+} \wp(x|\sigma) \stackrel{\mathcal{D}'}{=} \lim_{\sigma \rightarrow 0+} B(x|\sigma) \stackrel{\mathcal{D}'}{=} \delta(x),$$

where  $\delta(x)$  is the well-known Dirac delta pulse. By a transition to this zero-variance limit one can, in fact, obtain a discrete variant of the smoothed number of particles (7).

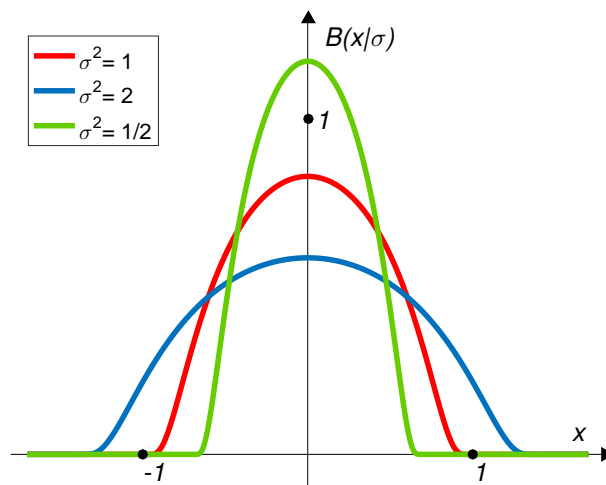


Figure 7: Borsalino functions. Center is  $\alpha = 0$  and the shape-parameter is  $\sigma \in \{1, \sqrt{2}, \sqrt{2}/2\}$ .

## 7.2 Summary of terminology used

For comfort and better understanding, in the following table we summarize a terminology used in this paper.

Table 1: Terminology.

Technical term	Symbol	Explanation
Vehicular Headway Modeling	VHM	scientific field dealing with inter-vehicular gaps
detector line		a static location of detector
primary traffic quantities		gauged by detector or stated by exact formulas
secondary traffic quantities		calculated from primary ones by approximations
phase (state) variables	$\varrho, I, V$	density, intensity, and mean speed
equation of continuity	$\frac{\partial \varrho}{\partial \tau} = -\frac{\partial I}{\partial \xi}$	law of conservation for traffic flow
phase segment (segmentation zone)	$\Psi$	small sub-area v ID plane
first phase relation	$\Omega_{ID}$	all existing intensity-density pairs
second phase relation	$\Omega_{VD}$	all existing velocity-density pairs
phase diagram		surface graph of phase relations
time headway	$z$	time between two cars as they pass a detector
time clearance	$t$	time gap between two successive vehicles
distance headway	$s$	space between two front bumpers in a fixed time
distance clearance	$r$	space gap between two successive vehicles
individual velocity	$v$	speed of a vehicle
scaled time clearance	$y$	time clearance after the unification applied
scaled distance clearance	$x$	space clearance after the unification applied
sample		set of several succeeding vehicles
sample of time headways ( $i$ th)	$Z_i$	set of time headways in one sample
sample of time clearances	$T_i$	set of time clearances in one sample
sample of space headways	$S_i$	set of space headways in one sample
sample of space clearances	$R_i$	set of space gaps in one sample
sample of velocities	$\Upsilon_i$	set of velocities related to a sample
sample of scaled time clearances	$Y_i$	set of scaled time clearances in one sample
sample of scaled space gaps	$X_i$	set of scaled space gaps in one sample
smoothed number of particles	$N(\xi, \tau)$	number of cars, mathematically optimized
segmented index set	$F_\Psi$	sample indices belonging to $\Psi$
segmented time clearances	$Y_\Psi$	set of unified time clearances belonging to $\Psi$
segmented space clearances	$X_\Psi$	set of unified space gaps belonging to $\Psi$
segmented velocities	$\Upsilon_\Psi$	set of velocities belonging to $\Psi$