Three methods for estimating a range of vehicular interactions

Milan Krbálek [a,*], Jiří Apeltauer [a,b], Tomáš Apeltauer [b], Zuzana Szabová [a]

[a] *Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Prague, Czech Republic*
[b] *Faculty of Civil Engineering, Brno University Of Technology, Brno, Czech Republic*

HIGHLIGHTS

- We present new approaches for estimating the number of cars influencing a decision-making procedure of drivers.
- Empirical data samples are subjected to advanced methods of statistical analysis.
- Consistency between the estimations used is surprisingly credible.
- We demonstrate that universally-accepted premise on short-ranged traffic interactions is not substantiated.
- All methods introduced have revealed that minimum number of actively-followed vehicles is two.

ARTICLE INFO

ABSTRACT

We present three different approaches how to estimate the number of preceding cars influencing a decision-making procedure of a given driver moving in saturated traffic flows. The first method is based on correlation analysis, the second one evaluates (quantitatively) deviations from the main assumption in the convolution theorem for probability, and the third one operates with advanced instruments of the theory of counting processes (statistical rigidity).

We demonstrate that universally-accepted premise on short-ranged traffic interactions may not be correct. All methods introduced have revealed that minimum number of actively-followed vehicles is two. It supports an actual idea that vehicular interactions are, in fact, middle-ranged. Furthermore, consistency between the estimations used is surprisingly credible. In all cases we have found that the interaction range (the number of actively-followed vehicles) drops with traffic density. Whereas drivers moving in congested regimes with lower density (around 30 vehicles per kilometer) react on four or five neighbors, drivers moving in high-density flows respond to two predecessors only.

© 2017 Elsevier B.V. All rights reserved.

1. Motivation

Nowadays, a deeper understanding of the principles of vehicular dynamics is becoming increasingly important because of many reasons. It is essential that numerical/theoretical models are capable to reproduce a traffic reality more and more authentically since elaborated simulators can then more-responsibly predict impending traffic congestions, manage control systems in autonomous vehicles, or help with planning of inter-vehicular communication networks. For these reasons we now ask fundamental questions on a nature/intensity/range of inter-vehicle interactions.

---

* Corresponding author.
E-mail address: milan.krbalek@fjfi.cvut.cz (M. Krbálek).

From a general viewpoint, any traffic system is an agent ensemble whose intelligent agents interact with a certain set of their neighbors. Although forces of such a kind are not directly measurable some recent works (see [1–9]) have revealed a way how to optimize a force-description to obtain a more realistic predictions of vehicular microstructure. In [2–5] authors have demonstrated that one-dimensional thermal gas, whose particles interact via a repulsion potential depending on reciprocal value of distance among succeeding cars, represents a suitable theoretical model that reproduces a microscopic structure of freeway traffic surprisingly good. However, this stochastic thermodynamic model as well as all the best-known microscopic traffic-models (follow-the-leader models, car-following models, intelligent driver model, optimal velocity model, Nagel–Schreckenberg model –see [10–12]) are based on the reductive premise that an "explicit force" exists between neighboring vehicles only. Such a property used to be referred to as a short-ranged interaction. To what extent does this premise correspond with traffic reality? Do there exist interactions among more neighbors as well? Is there any theoretically-substantiated methodology for estimating the number m of actively-followed vehicles (the so-called *interaction range*)? These are the issues that we will try to discuss in this paper.

Intuitively, one can expect (in contrast to basic principles of microscopic traffic modeling) that decision-making procedure of a driver (moving in congested traffic-regimes) is typically influenced by more circumjacent cars. This is referred to as a middle-ranged case. To conclude, the main objective of this paper is to estimate the interaction range in empirical traffic samples by means of several proven/innovative mathematical approaches. Furthermore, we plan to analyze how the interaction range evolves with changing values of basic macroscopic quantities like traffic flow or density.

2. Empirical data-sets and the segmentation procedure

Vehicle-by-vehicle data analyzed in this paper has been provided by the Road and Motorway Directorate of the Czech Republic (ŘSD ČR) and recorded at the Expressway R1 (also called the Prague Ring) in Prague, the Czech Republic. For intentions of this research we have used detectors located sufficiently far from any on/off-ramps where traffic flow is dense enough to generate congested traffic states. Here one can guarantee the conservation of the number of vehicles as well as a significant synchronization among moving vehicles. Only intensive inter-vehicular cooperation can hypothetically lead to a detection of an interaction range. Therefore, we have eliminated (by means of standard methods for statistical clustering) all the states associated to free traffic-phase where the intended detection loses its meaning. Above that, we analyze data from fast lanes only, because of a significant proportion of lorries, buses, tracks in main lanes. Such a reduction brings three important advantages. Firstly, in fast lanes there is an extremely low proportion of slow cars and no long vehicles. Secondly, fast-lane vehicles are coerced to more intensive interactions (due to higher speeds) and thirdly, fast-lane vehicles cannot be overtaken. These three factors strengthen target-ambitions of this paper.

The data records have been adapted into a set (or sets) having a form

$$\Omega = \left\{ (\tau_k^{(in)}, \tau_k^{(out)}, v_k, \xi_k) \in T^{(in)} \times T^{(out)} \times V \times \Xi : k = 1, 2, \ldots, N \right\}, \tag{1}$$

that is suitable for synoptical mathematical formulations. Here $T^{(in)}$ and $T^{(out)}$ are the sets of chronologically-ordered times when the kth vehicle entered/left the measuring device, respectively. V is the set of associated velocities and Ξ stands for the set of vehicular lengths. Denoting the sampling size by ℓ (which is here consistently considered equal to 50) and number of data-samples by K we acquire the data sub-samples

$$\Phi_j = \left\{ (\tau_k^{(in)}, \tau_k^{(out)}, v_k, \xi_k) \in \Omega : k = (j-1)\ell + 1, (j-1)\ell + 2, \ldots, j\ell \right\}. \tag{2}$$

For each sub-sample Φ_j ($j = 1, 2, \ldots, K$) one can calculate the local flux

$$J_j = \frac{\ell}{\tau_{j\ell}^{(in)} - \tau_{(j-1)\ell+1}^{(out)}} \tag{3}$$

and local average velocity $\bar{v}_j = \ell^{-1} \sum_{k=(j-1)\ell+1}^{j\ell} v_k$. Besides, the expression

$$\varrho_j = \frac{J_j}{\bar{v}_j} \tag{4}$$

is accepted as a plausible approximation of the local vehicular density (according to [5,10]). Individual (re-scaled) time-clearances are then calculated using the standard definition (e.g. [5])

$$t_k = \frac{\ell \left(\tau_k^{(in)} - \tau_{(k-1)}^{(out)} \right)}{\sum_{i=(\lceil k/\ell \rceil - 1)\ell + 1}^{\lceil k/\ell \rceil \ell} \tau_i^{(in)} - \sum_{i=(\lceil k/\ell \rceil - 1)\ell + 1}^{\lceil k/\ell \rceil \ell} \tau_{i-1}^{(out)}}, \tag{5}$$

which ensures that for all sample-means it holds $\bar{x}_j := \ell^{-1} \sum_{k=(j-1)\ell+1}^{j\ell} t_k = 1$. As it is well known (from [13–18]), such a re-scaling procedure brings a considerable profit when revealing general relations in many economic/physical/biologic/socio-physical/purely mathematical systems. In analogy, we define (re-scaled) space-clearances by

$$x_k = \frac{v_{k-1} \ell \left(\tau_k^{(in)} - \tau_{(k-1)}^{(out)} \right)}{\sum_{i=(\lceil k/\ell \rceil - 1)\ell + 1}^{\lceil k/\ell \rceil \ell} v_{i-1} \tau_i^{(in)} - \sum_{i=(\lceil k/\ell \rceil - 1)\ell + 1}^{\lceil k/\ell \rceil \ell} v_{i-1} \tau_{i-1}^{(out)}}. \tag{6}$$

Table 1
Basic information on empirical data.

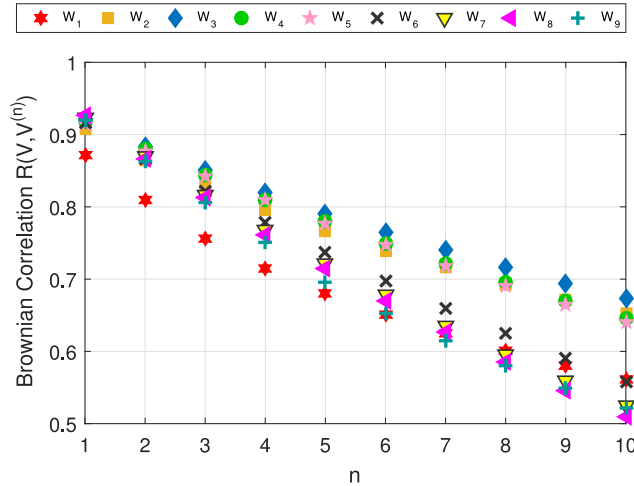| w	Window	ϱ [km^{-1}]	Δ_ϱ [km^{-1}]	J [h^{-1}]	Δ_J [h^{-1}]	$|I_w|$	$\hat{\beta}$
1	W_1	25	5	500	2500	201 200	0.72455
2	W_2	30	5	650	2350	163 600	1.0217
3	W_3	35	5	700	2500	121 750	1.2597
4	W_4	40	5	700	2500	80 600	1.4347
5	W_5	45	5	700	2500	52 850	1.6197
6	W_6	50	5	1000	2000	31 200	1.7382
7	W_7	55	5	1000	1800	15 850	1.8934
8	W_8	60	5	1000	1500	8 000	2.0526
9	W_9	65	5	1000	1500	2 850	2.1569

Fig. 1. Brownian distance-correlation of velocities. We plot a value of the distance-correlation coefficient $\mathcal{R}(V_w, V_w^{(n)})$ enumerated for chronologically-ordered sets (8) and (10) in various density regions (see legend for detail specifications).

Many recent investigations have shown that in regions of congested traffic the evolution of traffic quantities shows signs of chaotic systems (e.g. [10,19,20]). It means that a small change of initial conditions leads to sharp changes in evolution. Accompanying effects are, as is proven, significant divergences in probabilistic characteristics detected in various subspaces of the phasic diagram [21,22]. Indeed, similar effects have been revealed in the past also for empirical vehicular data. In [1,4,5,12,15,16] authors have shown that statistical distributions of traffic micro-quantities are intensively depending on actual density and flux. To eliminate undesirable mixing of different traffic states and to stabilize estimation-procedure aimed for this paper we proceed to the tried and tested stabilization-method (see [1,4,5,12,16]). Factually, such a method is based on a segmentation by density ϱ and flux J. Thus, for a fixed window-size $(\Delta_\varrho, \Delta_J)$ we define a flux–density window $W(\varrho, J) = [\varrho, \varrho + \Delta_\varrho] \times [J, J + \Delta_J]$ and analyze empirical data separately by various windows (indexed by w for brevity). Therefore, the segmentation

$$I_w \equiv I(W(\varrho, J)) = \{j : (\varrho_j, J_j) \in W(\varrho, J)\}$$

is a procedure selecting all sub-samples associated with the chosen wth flux–density window. This is a central strategy of all following considerations.

Consistently, in the entire text we use the flux–density windows summarized in Table 1.

3. Correlation analysis in subregions of congested phase

An elementary insight (slightly superficial) into the range of inter-vehicular interactions can be done with help of correlation analysis. It is evident that general correlations of various traffic micro-quantities are a result of mutual influence among several consecutive cars. Therefore many researchers focused their attention on studies of associated correlation coefficients. For example, the articles [5,23–26], review [10], or book [19] analyze statistical links between selected traffic quantities and discuss some consequences. However, frequently-used methodics applying the standard (Pearson) correlation coefficients suffers from many shortcomings. Firstly, correlation coefficient (see formula (3.19) in [19] and commentary below) detects a linear relationship between two random variables. It means that for non-linearly linked variables (which is a most probable situation in congested-traffic states) such a methodology fails. Secondly, zero value
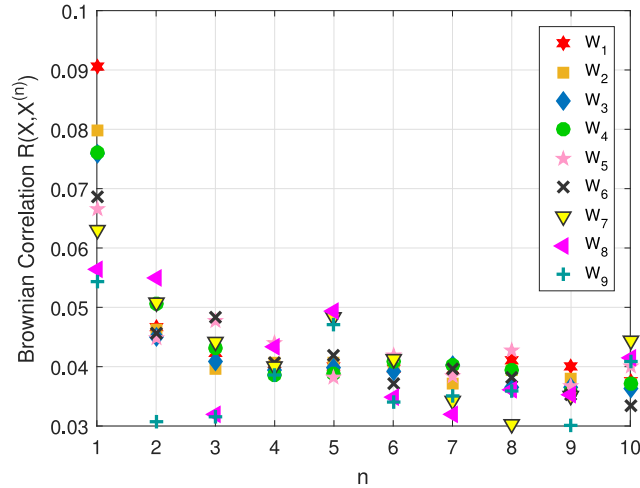
Fig. 2. Brownian distance-correlation of space-clearances. We plot a value of the distance-correlation coefficient $\mathcal{R}(X_w, X_w^{(n)})$ enumerated for chronologically-ordered sets (7) and (9) in various density regions (see legend for detail specifications).

of the classical Pearson correlation coefficient does not imply statistical independence because the applicable theorem is enunciated as a one-sided implication only (see [27]). This problem can be, however, eliminated by introducing Brownian distance-correlation (see Appendix A.1). Such a type of correlator relies on a stronger version of the above-cited theorem. In this case, zero value of Brownian correlation is, as proven in [27,28], a sufficient condition for statistical independence. Thirdly, non-zero value of correlation does not imply causality. Thus, it may occur that two strongly-correlated variables are not causatively linked. Despite these deficiencies, however, analyzing the Brownian correlations of empirical data can be illustrative for an approximate insight into the phase structure of congested traffic. Therefore, denoting

$$X_w := \{x_{k\ell+1}, x_{k\ell+2}, \ldots, x_{(k+1)\ell} : \ k \in I_w\}, \tag{7}$$

and

$$V_w := \{v_{k\ell+1}, v_{k\ell+2}, \ldots, v_{(k+1)\ell} : \ k \in I_w\} \tag{8}$$

the sets of individual clearances/velocities (chronologically ordered and associated to the wth flux–density window) we can proceed to evaluation of the distance-correlation coefficients $\mathcal{R}(X_w, X_w^{(n)})$ and $\mathcal{R}(V_w, V_w^{(n)})$, where (for natural number n)

$$X_w^{(n)} := \{x_{k\ell+n+1}, x_{k\ell+n+2}, \ldots, x_{(k+1)\ell+n} : \ k \in I_w\} \tag{9}$$

and

$$V_w^{(n)} := \{v_{k\ell+n+1}, v_{k\ell+n+2}, \ldots, v_{(k+1)\ell+n} : \ k \in I_w\}. \tag{10}$$

It illustratively follows from Fig. 1 that values of correlation coefficient $\mathcal{R}(V_w, V_w^{(n)})$ are decreasing with n in all density regions. If $n > 4$ is fixed the evolution of $\mathcal{R}(V_w, V_w^{(n)})$ clearly shows a separation of regions with different level of synchronization. If analyzing $\mathcal{R}(X_w, X_w^{(n)})$ in Fig. 2 one can see that the most intensive correlations are detected for $n = 1$, which means that free gap to a previous car is significantly more correlated with a previous clearance than with clearances between other cars. Such findings correspond to our intuitive conjectures as well as to the comparable results published in [23,24].

It directly follows from the previous text that an isolated analysis of correlations is (if intending to estimate an interaction range) ineffective. Therefore it is necessary to search for a supplementary instrument leading to a theoretical interpretation of given values of correlation coefficients. For those purposes we now choose the local thermodynamic model [1,12] whose legitimacy has been verified in many researches [3–6,15]. The main merits of that model are the facts that it is exactly solvable and furthermore, it authentically reproduces the phase structure of congested traffic. Here we introduce a generalized version of that model taking into account that interaction forces exist among many succeeding agents.

First of all, however, we consider the stochastic particle-system composed of M identical particles located in angular positions $a_1(\tau) < a_2(\tau) < \cdots < a_M(\tau) < a_1(\tau) + 2\pi =: a_{M+1}(\tau)$ and moving with positive speeds $v_1(\tau), v_2(\tau), \ldots, v_M(\tau)$. Level of stochasticity is here controlled by the coefficient of stochastic resistivity $\beta \geqslant 0$. A *short-ranged version* of this ensemble is described by the Hamiltonian

$$\mathcal{H}(\mathbf{a}(\tau), \mathbf{v}(\tau)) = \frac{1}{2} \sum_{k=1}^{M} (v_k(\tau) - v_d)^2 + \frac{2\pi}{M} \sum_{k=1}^{M} \frac{1}{a_{k+1}(\tau) - a_k(\tau)}, \tag{11}$$
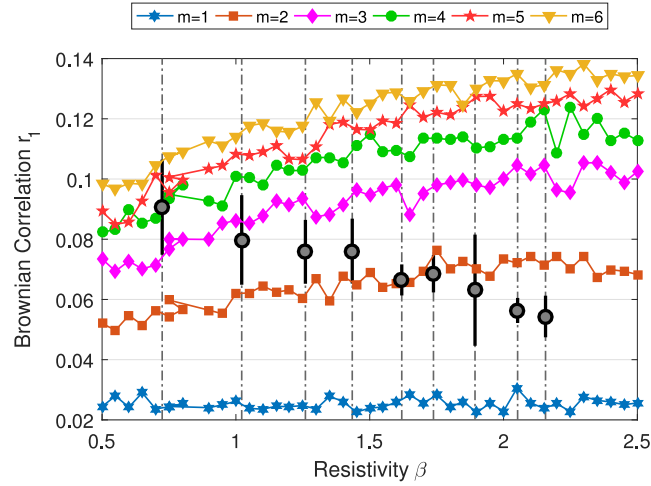
Fig. 3. A course of the theoretical correlation-function. We depict estimated evolution of the Brownian correlation coefficient r_1 induced by changing parameters: range m (see legend) and stochastic resistivity β (see the horizontal axis). Graphs have been constructed for a gas composed of $M = 100$ particles. Vertical lines represents specific values of resistivity detected in real-road data for various flux–density windows (see the last column in Table 1). Empirical values of Brownian correlation (see Fig. 2) are visualized by means of averages (black bullets) and associated deviation-intervals (black abscissae) demarcated by standard deviations.

where v_d is the desired velocity. This short-ranged variant has been successfully solved in [12]. The most important result of the cited article is a determination and validation of the probability density for distance x between neighboring particles (headway distribution). It reads

$$\wp(x) = A e^{-\frac{\beta}{x}} e^{-Dx}, \tag{12}$$

where

$$D = \beta + \frac{3 - e^{-\sqrt{\beta}}}{2}, \tag{13}$$

$$A^{-1} = 2\sqrt{\frac{\beta}{D}} \mathcal{K}_1(2\sqrt{\beta D}). \tag{14}$$

Here $\Theta(x)$ and $\mathcal{K}_\nu(x)$ stand for the Heaviside unit-step function and the Macdonald's function of the νth order, respectively.

Denoting now $a_{M+i}(\tau) := a_i(\tau)$ for all $i = 1, 2, \ldots, M$ one can define a *middle-ranged version* of the model. The corresponding Hamiltonian can be expressed as

$$\mathcal{H}_m(\mathbf{a}(\tau), \mathbf{v}(\tau)) = \frac{1}{2}\sum_{k=1}^{M}(v_k(\tau) - v_d)^2 + \frac{2\pi}{M}\sum_{k=1}^{M}\sum_{i=1}^{m}\frac{1}{a_{k+i}(\tau) - a_k(\tau)}, \tag{15}$$

where the natural number $m > 1$ to be referred to as an *interaction range*. At the moment, associated steady-states of the middle-ranged version are not analytically calculated, however, numerical representation of the model (based on principles of the simulated annealing [2]) allows obtaining steady-distributions for all micro-quantities. Steady-state locations $\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_M$ can be then used for enumerating steady-state headways $x_k := 2\pi(\tilde{a}_{k+1} - \tilde{a}_k)/M$, for which the average value is exactly equal to one. Fixing the interaction range m and stochastic resistivity β we can calculate the Brownian correlation coefficient $r_n := \mathcal{R}(\mathbf{X}, \mathbf{X}_n)$ for vectors $\mathbf{X} = (x_1, x_2, \ldots, x_M)$ and $\mathbf{X}_n = (x_{n+1}, x_{n+2}, \ldots, x_{n+M})$. In this way, in fact, we construct numerical estimations of the theoretical correlation-functions $r_n = r_n(\beta, m)$. A course of the most important correlation-function $r_1 = r_1(\beta, m)$ is sketched in Fig. 3.

If $m = 1$ one can see that correlation-function $r_1(\beta, 1)$ slightly fluctuates around a very low value, which represents, in fact, a statistical independence among neighboring headways. Situation dramatically changes for $m > 1$. Besides elevated values $r_1(\beta, m)$ one can also detect an upward trend, i.e. $\partial r_1/\partial \beta > 0$. Curves $r_1(\beta, m)$ and $r_1(\beta, m+1)$ are not intersecting which brings a stronger interpretation-potential. Another significant benefit of these results follows from the fact that numerical values and empirical values lie in the same region. Therefore, one can use the correlation-function $r_1 = r_1(\beta, m)$ for a rough estimation of the interaction range. As is apparent from Fig. 3 the estimated range m is depending on traffic density. Longer ranges are detected for lower densities, whereas correlation drops with increasing density. To be specific, Table 2 (the second column denoted by m_r) summarizes our preliminarily estimations for all flux–density windows.

Table 2
Detection of an interaction range.

Window	m_r (correlation analysis)	m_{mc} (multi-clearance analysis)	m_{sr} (rigidity analysis)
W_1	3–6	5	5
W_2	2–4	4	4
W_3	2–3	3	3
W_4	2–3	2	3
W_5	2	2	4
W_6	2	2	2
W_7	2	2	2
W_8	2	2	2
W_9	2	2	2

4. Multi-clearance as a random variable

An indispensable part of investigations intended is the statistical analysis of multi-clearances. Similarly to the articles [5, 7,26,29] we define *multi-clearances of the order* $\mu \in \mathbb{N}$ (μth multi-clearance) by a formula

$$x_k|\mu = x_k + x_{k-1} + \cdots + x_{k-\mu}. \tag{16}$$

Then the associated histogram-function $H(x|\mu)$ quantifies a statistical distribution of cumulated gaps measured among $\mu+2$ succeeding cars moving in clusters belonging to a chosen flux–density window. Note that due to the definition (16) the space occupied by individual vehicles is eliminated from our considerations. In agent's systems, where interaction rules are strictly short-ranged, one can reasonably expect that succeeding clearances are not correlated. Therefore random variables $x|\mu_1$ and $x|\mu_2$ (for $\mu_1 \neq \mu_2$) are independent which is resulting in the convolution formula

$$\wp(x|\mu) = \wp(x) \star \wp(x|\mu - 1) \equiv \int_{\mathbb{R}} \wp(s)\wp(x - s|\mu - 1)\,\mathrm{d}s. \tag{17}$$

Using (12) one finds (see [30])

$$\wp(x|\mu) = \Theta(x)\frac{(Ax)^\mu}{\mu!}\exp\left[-\frac{\beta(\mu+1)^2}{x} - Dx\right]. \tag{18}$$

Unfortunately, Eq. (18) represents a relevant prediction for ensembles with short-ranged forces only. If, in contrast, interactions among vehicles are middle-ranges, which is expectable for real-road traffic, the above-applied convolution rule is not further applicable and the prediction (18) therefore fails. However, one can hypothesize that the two-parametric family of probability densities

$$\wp(x|\mu) = A\Theta(x)x^\alpha \exp\left[-\frac{\beta}{x} - Dx\right], \tag{19}$$

where

$$D = \frac{\alpha}{\mu} + \frac{\beta}{\mu^2} + \frac{3 - \mathrm{e}^{-\sqrt{\beta/\mu}}}{2\mu}, \tag{20}$$

$$A^{-1} = 2\left(\frac{\beta}{D}\right)^{\frac{\alpha+1}{2}}\mathcal{K}_{\alpha+1}\left(2\sqrt{D\beta}\right), \tag{21}$$

can be used as a theoretical prognosis of real-road multi-clearance distributions. Analogous method has been successfully applied in [5].

The afore-mentioned hypothesis will now be tested by a special modification of the experienced MDE method. That method (referred to as Minimum Distance Estimation on equivariance curves) is based on estimating optimal parameters in distribution (19) while maintaining a value of the variance (to be equal to the squared deviation measured in data). This strategy (discussed in detail in Appendix A.2) stabilizes an estimation procedure and dramatically reduces the time required for numerical calculations. Values of statistical distance

$$\|\hat{\wp}(x|\mu) - H(x|\mu)\| = \left(\int_{\mathbb{R}} |\hat{\wp}(x|\mu) - H(x|\mu)|^2\,\mathrm{d}x\right)^{1/2} \tag{22}$$

measured between the empirical histogram-function $H(x|\mu)$ and the MDE-estimation $\hat{\wp}(x|\mu)$ (visualized in Fig. 4) confirm that the hypothesis on GIG-distributed multi-clearances is well-founded. Indeed, as is seen in Fig. 5 the analytical prediction (19) estimate empirical distributions very convincingly, which is important since the below-mentioned detection of a interaction range requires an increased accuracy of estimations used. To be complete, in Figs. 6, 7 we plot all estimated parameters $\hat{\alpha}_\mu$ and $\hat{\beta}_\mu$.
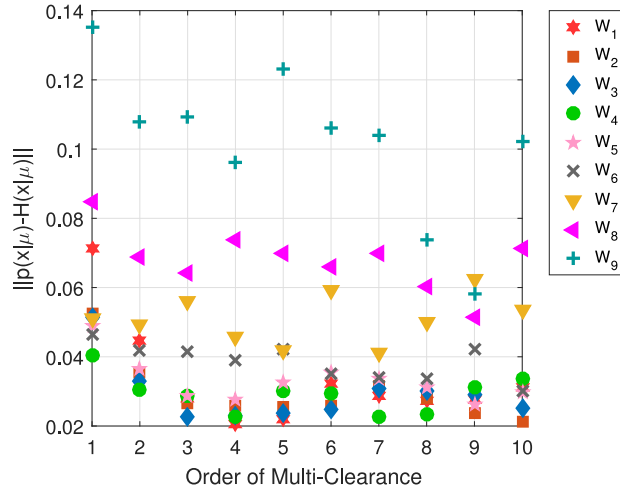
Fig. 4. Deviation between empirical multi-clearance distributions and relevant estimations. We plot the changes of the statistical distance $\|\hat{\wp}(x|\mu) - H(x|\mu)\|$ with changing order μ in various flux–density windows (see legend).
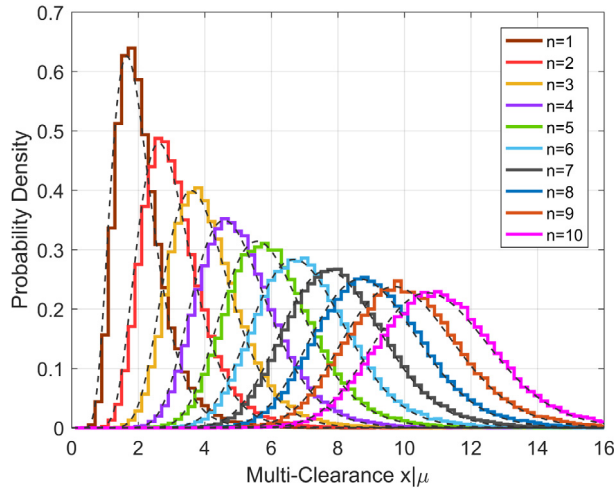

Fig. 5. Multi-headway distributions for the fourth window, i.e. for $w = 4$. Step-functions visualize multi-clearance distributions detected in empirical data, whereas dashed curves correspond to associate estimations (19) obtained by the modified Minimum Distance Estimation described in Appendix A.2.

5. Detection instrument based on a perturbation function

In the actual section we plan to introduce an analytical method for deciding how many immediately neighboring cars (in a driving direction) influence decision-making procedures of a given driver. This method is founded on testing statistical independence of different multi-clearances. It is well known that if random variables x and $x|\mu$ are uncorrelated then probability density $\wp(x|\mu + 1)$ can be computed with help of a convolution rule $\wp(x|\mu + 1) = \wp(x) \star \wp(x|\mu)$. Therefore a perturbation function

$$\varphi(x|\mu) = \int_{-\infty}^{x} \big(H(y|\mu) - H(y) \star H(y|\mu - 1) \big) \, \mathrm{d}y. \tag{23}$$

reflects a degree of independence between x and $x|(\mu - 1)$ in a following sense. If $\varphi(x|\mu)$ is insignificantly deflected from zero (in a statistical sense) then a presumption of negligible mutual interactions is not justified, which leads to suspicion that associated cars interact. Thus, courses of perturbation functions then help us to quantify an interaction intensity among farther vehicles. Advantageously, applying knowledge of the previous sections we define a smoothed perturbation function

$$\psi(x|\mu) = \int_{-\infty}^{x} \big(\hat{\wp}(y|\mu) - \hat{\wp}(y|0) \star \hat{\wp}(y|\mu - 1) \big) \, \mathrm{d}y, \tag{24}$$
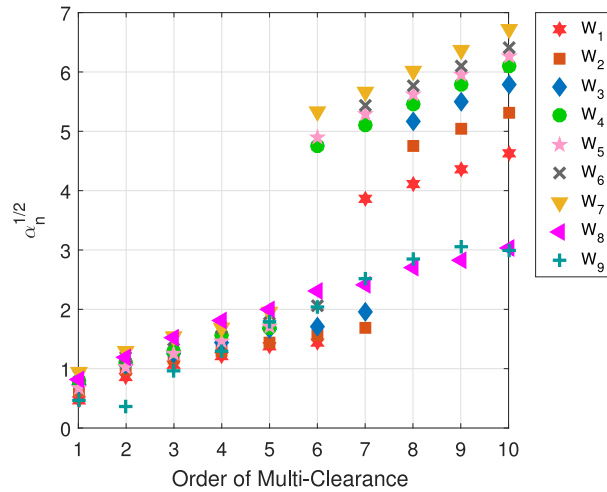
Fig. 6. Estimated values for parameter $\hat{\alpha}_\mu$. We plot the square-root of the first estimated parameter for various order (x-axis) and flux–density window (legend).
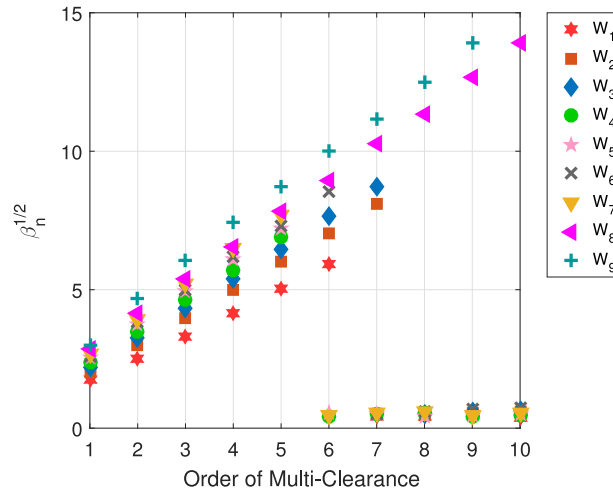

Fig. 7. Estimated values for parameter $\hat{\beta}_\mu$. We plot the square-root of the second estimated parameter for various order (x-axis) and flux–density window (legend).

that eliminates a step character of empirical histograms. Realistic courses of smoothed perturbation functions $\psi(x|\mu)$ (analyzed for samples of 5000 multi-clearances) are neatly shown in Fig. 8.

For quantifying a deviation of the smoothed perturbation function (24) from zero level we use the standard Kolmogorov distance

$$G(\mu) = \sup_{x \in \mathbb{R}} |\psi(x|\mu)| . \tag{25}$$

Hence, if the Kolmogorov distance $G(\mu)$ is greater than the critical value $G_{\text{crit}} = 1.36\sqrt{2/5000}$ (for the two-sample Kolmogorov–Smirnov test) then the null hypothesis (on negligible correlations among vehicular movements) is rejected (at significance level 0.05). In the opposite case the null hypothesis may not be rejected and therefore there exists a serious suspicion that those two cars are mutually affected. Observing Fig. 9 we can see that a rejection/acceptance of the null hypothesis depends on a specific location in flux–density diagram. Whereas in regions of lower densities (for $w = 1$) the null hypothesis is rejected for $\mu \in \{1, 2, 3, 4\}$, for regions of higher densities ($w \in \{4, 5, 6, 7, 8, 9\}$) the null hypothesis is rejected for $\mu = 1$ only. However, a rejection for $\mu = 1$ in all regions confirms our initial hypothesis that vehicular interactions are not short-ranged. Indeed, if vehicular influences were short-ranged then surely $\hat{\wp}(y|1) = \hat{\wp}(y|0) \star \hat{\wp}(y|0)$, which implies $G(\mu) = 0$. Therefore, in Table 2 (the third column denoted by m_{mc}) we present the second estimation of a interaction-range (determined by means of multi-clearance distribution). We note that Kolmogorov distance enumerated for eight and ninth windows (see crosses in Fig. 9) suffers from a lack of data.
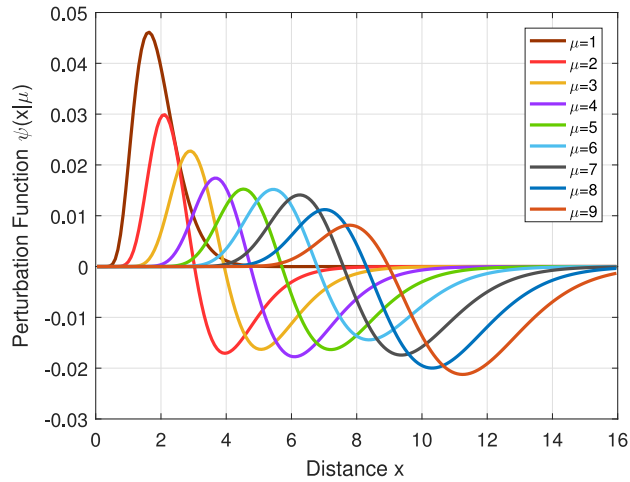
Fig. 8. Smoothed perturbation functions. We figure the functions $\psi(x|\mu)$ valid for the third flux–density window. Various curves correspond to various interaction ranges (see legend). Amplitude of those functions coincides with the Kolmogorov distance $G(\mu)$ defined in (25).
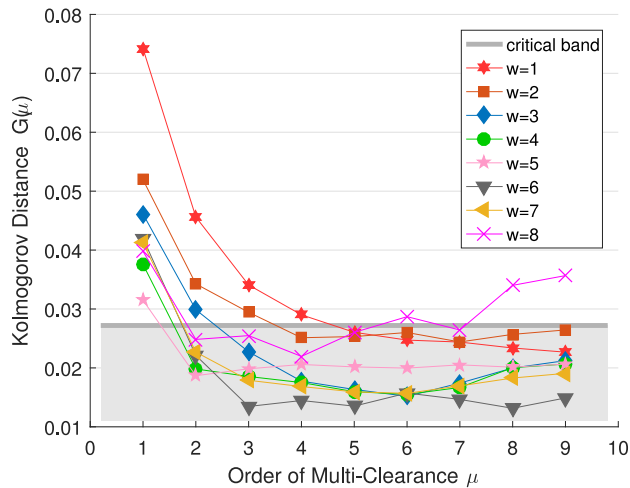

Fig. 9. Kolmogorov–Smirnov test. We visualize the Kolmogorov distance $G(\mu)$ as it changes with the window-index w and the order of multi-clearance μ. Gray zone represents a critical band in which the null hypothesis may not be rejected.

6. Detection instrument based on a statistical rigidity

It has been investigated in [2,4,15,29,33] that more effective tool for inspection of traffic microstructure, if comparing with classical approaches using clearance distributions, is the statistical rigidity. This advanced characteristics is understood as follows. Let us denote by U_L the number of vehicles occurring (at a fixed time) inside the road-segment of a length L. Statistically, $U_L(t)$ represents a random process (a counting process), whose average value is equal to L due to the fact that clearance statistics is studied in the re-scaled version. In general counting processes, however, may not be met an intuitive condition $E(U_L) = L$. Therefore variance $E(U_L - E(U_L))^2$ and generalized variance $E(U_L - L)^2$ do not necessarily represent the same values. A functional dependency of the generalized variance on L, here denoted by $\Delta(L)$, is the statistical rigidity.

It is proven by theoretical derivations in [4,15] that a course of the rigidity is uniquely determined by a probability density for gaps among succeeding agents/particles. However, this direct link is valid only if succeeding clearances are independent and identically distributed (i.i.d. property). If assuming that such a condition is fulfilled we can determine a linear asymptotics $\Delta(L) = \chi L + \gamma + \mathcal{O}(L^{-1})$ of the rigidity. For ensembles with GIG-distributed gaps (see formula (12)) it holds

$$\chi = \frac{\beta + 2}{D} - 1, \quad \gamma = \frac{6\sqrt{D\beta} + D\beta\left(21 + 4D\beta + 16\sqrt{D\beta}\right)}{24\left(1 + \sqrt{D\beta}\right)^4}. \tag{26}$$
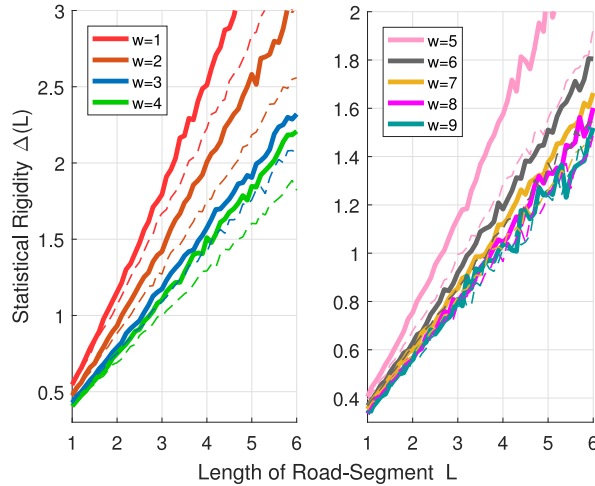
(See Fig. 10)

Fig. 10. Statistical Rigidity. We plot rigidities $\Delta(L)$, $\Delta_0(L)$ for not-shuffled/shuffled clearances respectively. Not-shuffled data are plotted by continuous curves whereas shuffled data are plotted by dashed curves.

Whereas i.d. condition can be, without any doubt, accepted in all traffic systems, independency is regarded as a disputable property. In other words, a deviation of empirical rigidity from analytical prognosis can be used as a indicator of influenced behavior of drivers. The last detection-method is based on that principle. We analyze the statistical rigidity for two sets: the set of empirical clearances in an original order and the set of clearances which have been repeatedly shuffled to break all correlation linkages among succeeding clearances. These two rigidities are here denoted by $\Delta(L)$ and $\Delta_0(L)$ respectively. (See Fig. 10.) Roughly speaking, angular deflection

$$\eta = \mathrm{arctg}(\chi) - \mathrm{arctg}(\chi_0) \tag{27}$$

between associated asymptotes $\chi L + \gamma$ and $\chi_0 L + \gamma_0$ will be here used for estimations intended. Similarly, the same procedure can be then applied for multi-clearances of order μ. It means that angular deflection

$$\eta(\mu) = \mathrm{arctg}(\chi(\mu)) - \mathrm{arctg}(\chi_0(\mu)) \tag{28}$$

reveals whether interactions exist between vehicles having μ cars among themselves. Range-detection can be then performed as follows. If a value η lies out of the fluctuation band (demarcated by pure statistical fluctuations – see a note below) then the statistical rigidities $\Delta(L)$ and $\Delta_0(L)$ are different, which means that succeeding clearances have to be correlated. It implies the fact that m cannot be equal to one. To be specific, if $\eta > \eta_{\mathrm{crit}}$ (see Fig. 11) then a corresponding interaction is not a short-ranged. Analogously, if $\eta(\mu) > \eta_{\mathrm{crit}}$ then the estimated range $m > \mu + 1$. Since $\eta(\mu)$ is a decreasing function (in all flux–density regions, basically) there always exists μ_0 so that $\eta(\mu) < \eta_{\mathrm{crit}}$ for $\mu \geqslant \mu_0$ and $\eta(\mu) > \eta_{\mathrm{crit}}$ for $\mu < \mu_0$. Then $m := \mu_0 + 1$ is the estimated range. Specific results of such an estimation procedure are summarized in Table 2 (the fourth column denoted by m_{sr}).

A note: For completeness, we add that a threshold-value $\eta_{\mathrm{crit}} \approx 0.0203$ is derived by an analysis of statistical fluctuations of regression coefficients measured for random uncorrelated data.

7. Summary, discussion and conclusion remarks

In the current paper we have presented three independent methods for estimating the so-called interaction range, i.e. number of preceding cars influencing a decision-making procedure of a given driver. The first method is based on a comparison of Brownian distance-correlation determined for ensembles of vehicles and agent-ensembles whose agents interact via two-body hyperbolical potential with a prescribed range. The second method imposes on recent knowledge on statistical distributions of vehicular random variables (like time clearances, time headways, spatial gaps, spatial headways, velocities). Theoretical predictions for a distribution of those variables or their compositions allow to investigate a statistical dependency among variables mentioned. Level of dependency is quantified by means of the perturbation function measuring distance between convolutional (i.e. calculated by a convolution) and empirical distributions of cumulative gaps. The third method is founded on a test of the statistical rigidity evaluated for clusters of several consecutive vehicles. Active interaction among vehicles leads to deviations between originally-ordered and randomly-shuffled clusters. Thus, a supra-threshold deflection between asymptotical regression-curves (estimating rigidities for not-shuffled and shuffled data) reveals mutual linkages among cars.
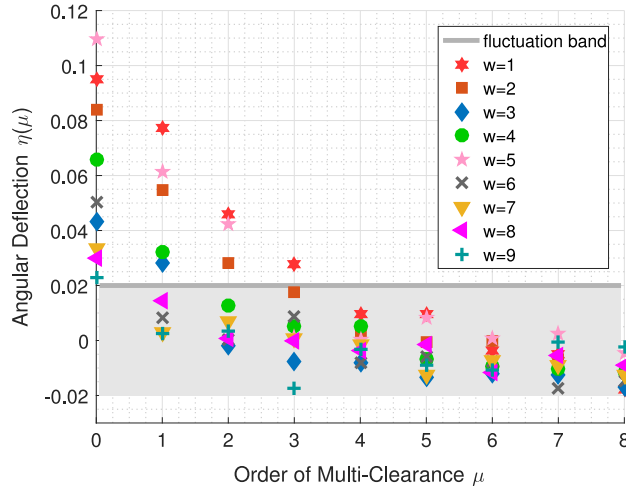
Fig. 11. Evolution of angular deflection. We figure changes of the deflection function (28) induced by changing multi-clearance order μ and changing location in the flux–density diagram.

The main result of this research is an irrefutable fact that vehicular interactions are definitely not short-ranged. It means that a generally-accepted premise (on short-ranged traffic interactions) do not correspond to traffic reality. In fact, in all congested-traffic regimes a minimum number of actively-followed vehicles is two. Moreover, for low-density congested traffic the interaction range reaches unexpectedly high values (four or five), which means that a driver tracks four of five cars moving in front of her/him.

Acknowledgments

Appendix A

A.1. Brownian distance-correlation

Consider a random sample $(\mathbf{X}, \mathbf{Y}) = \{(X_k, Y_k) : k = 1, 2, \ldots, n\}$ of n i.i.d. random vectors $\mathbf{X}$ and $\mathbf{Y}$ from Euclidean spaces $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively. Denote the Euclidean distances $a_{k\ell} := \|X_k - X_\ell\|_p$ and $b_{k\ell} := \|Y_k - Y_\ell\|_q$. For all $k, \ell = 1, 2, \ldots, N$ we define

$$A_{k\ell} = a_{k\ell} - \bar{a}_{k\bullet} - \bar{a}_{\bullet\ell} + \bar{a}_{\bullet\bullet}, \tag{29}$$

where

$$\bar{a}_{k\bullet} = \frac{1}{n} \sum_{\ell=1}^{n} a_{k\ell}; \quad \bar{a}_{\bullet\ell} = \frac{1}{n} \sum_{k=1}^{n} a_{k\ell}; \quad \bar{a}_{\bullet\bullet} = \frac{1}{n^2} \sum_{\ell=1}^{n} \sum_{k=1}^{n} a_{k\ell}. \tag{30}$$

Analogously one can define $B_{k\ell} = b_{k\ell} - \bar{b}_{k\bullet} - \bar{b}_{\bullet\ell} + \bar{b}_{\bullet\bullet}$. The Brownian sample distance-covariance $\mathcal{V}(\mathbf{X}, \mathbf{Y})$ is then defined as

$$\mathcal{V}^2(\mathbf{X}, \mathbf{Y}) = \frac{1}{n^2} \sum_{\ell=1}^{n} \sum_{k=1}^{n} A_{k\ell} B_{k\ell}. \tag{31}$$

Then the Brownian sample distance-variance $\mathcal{V}(\mathbf{X})$ reads

$$\mathcal{V}^2(\mathbf{X}) = \frac{1}{n^2} \sum_{\ell=1}^{n} \sum_{k=1}^{n} A_{k\ell}^2. \tag{32}$$

Finally, the Brownian sample distance-correlation $\mathcal{R}(\mathbf{X}, \mathbf{Y})$ is defined by

$$\mathcal{R}^2(\mathbf{X}, \mathbf{Y}) = \begin{cases} \dfrac{\mathcal{V}^2(\mathbf{X}, \mathbf{Y})}{\sqrt{\mathcal{V}^2(\mathbf{X})\mathcal{V}^2(\mathbf{Y})}}; & \mathcal{V}^2(\mathbf{X})\mathcal{V}^2(\mathbf{Y}) > 0; \\ 0; & \mathcal{V}^2(\mathbf{X})\mathcal{V}^2(\mathbf{Y}) = 0. \end{cases} \tag{33}$$

As is proven in [27,28] the Brownian correlation is zero if and only if the random vectors are independent.

A.2. Minimum distance estimation on equivariance curves

Let $H(x)$ be an estimated histogram-function for a random variable x located in Dom$(H) \subset [0, +\infty)$, for which the average value is 1 and the squared deviation is σ^2. Denote Ω a parametric space of real parameters α, β. Suppose that Ω is a compact subset of $\mathbb{R}^2$. Let $\wp(x|(\alpha, \beta))$ be a two-parametric family of probability-densities for which the first and the second moments exist and, moreover, $E(x) = 1$. Then the curve

$$C_\sigma := \left\{ (\alpha, \beta) \in \Omega : \int_0^\infty x^2 \wp(x|(\alpha, \beta)) \, dx = 1 + \sigma^2 \right\} \tag{34}$$

lying in the (α, β)−plane is called *an equivariance curve*. Let $\|\cdot\|$ be the arbitrary norm defined on a function space $L^2(0, +\infty)$ of quadratically integrable functions. Then a functional $g(\alpha, \beta)$ returning a norm $\|H(x) - \wp(x|(\alpha, \beta))\|$ represents a general metrics between two densities $H(x)$ and $\wp(x|(\alpha, \beta))$. The minimization

$$(\hat{\alpha}, \hat{\beta}) := \underset{(\alpha, \beta) \in C_\sigma}{\operatorname{argmin}} g(\alpha, \beta) = \underset{(\alpha, \beta) \in C_\sigma}{\operatorname{argmin}} \|H(x) - \wp(x|(\alpha, \beta))\| \tag{35}$$

will be referred to as a *minimum distance estimation* of $H(x)$ on an equivariance curve C_σ. Mathematically, this task can be identified with finding local/global extremes of a given function subject to equality constraints. For tasks of such a type the most effective solver is based on the method of Lagrange multipliers. Thus, introducing the Lagrangian

$$\mathcal{L}(\alpha, \beta, \lambda) := g(\alpha, \beta) - \lambda \int_0^\infty x^2 \wp(x|(\alpha, \beta)) \, dx - \lambda(\sigma^2 + 1) \tag{36}$$

one can calculate all critical points $(\alpha_j, \beta_j) \in C_\sigma$ of the Lagrangian by solving

$$\nabla_{\alpha, \beta, \lambda} \mathcal{L}(\alpha, \beta, \lambda) = (0, 0, 0). \tag{37}$$

If $\mathcal{L}(\alpha, \beta, \lambda)$ is continuously differentiable (which is an usual situation in most cases) then the global minimum of $g(\alpha, \beta)$ can be localized in a finite set $Q_\sigma := \{(\alpha_j, \beta_j) : j = 1, 2, \ldots\} \cup (\partial\Omega \cap C_\sigma)$. Thus,

$$(\hat{\alpha}, \hat{\beta}) = \underset{(\alpha, \beta) \in Q_\sigma}{\operatorname{argmin}} g(\alpha, \beta) \tag{38}$$

is the global minimum of the functional $g(\alpha, \beta)$ investigated under the condition $\text{VAR}(x) = \sigma^2$.

For intentions of this paper we specify the afore-mentioned methodology as follows. Distribution of random variable x is estimated by the generalized inverse Gaussian distribution [31–33]

$$\wp(x|(\alpha, \beta)) = Ax^\alpha \exp\left[-\frac{\beta}{x} - Dx \right] \tag{39}$$

with scaling and normalization constants

$$D = \alpha + \beta + \frac{3 - e^{-\sqrt{\beta}}}{2}, \tag{40}$$

$$A^{-1} = \begin{cases} 2\left(\dfrac{\beta}{D}\right)^{\frac{\alpha+1}{2}} \mathcal{K}_{\alpha+1}(2\sqrt{\beta D}); & \beta \neq 0; \\ \dfrac{\Gamma(\alpha + 1)}{(\alpha + 1)^{\alpha+1}}; & \beta = 0. \end{cases} \tag{41}$$

Function norm is suggested to be $\|f(x)\|^2 = \int_0^\infty |f(x)|^2 \, dx$ which is, in fact, the standard L^2−norm. Then the associated Lagrangian reads

$$\mathcal{L}(\alpha, \beta, \lambda) := \int_0^\infty |H(x) - \wp(x|(\alpha, \beta))|^2 \, dx - \lambda \left(\frac{\beta \, \mathcal{K}_{\alpha+3}(2\sqrt{\beta D})}{D \, \mathcal{K}_{\alpha+1}(2\sqrt{\beta D})} - \sigma^2 - 1 \right). \tag{42}$$

In this case the equivariance curve is determined by the equality

$$\frac{\beta \, \mathcal{K}_{\alpha+3}(2\sqrt{\beta D})}{D \, \mathcal{K}_{\alpha+1}(2\sqrt{\beta D})} = 1 + \sigma^2. \tag{43}$$
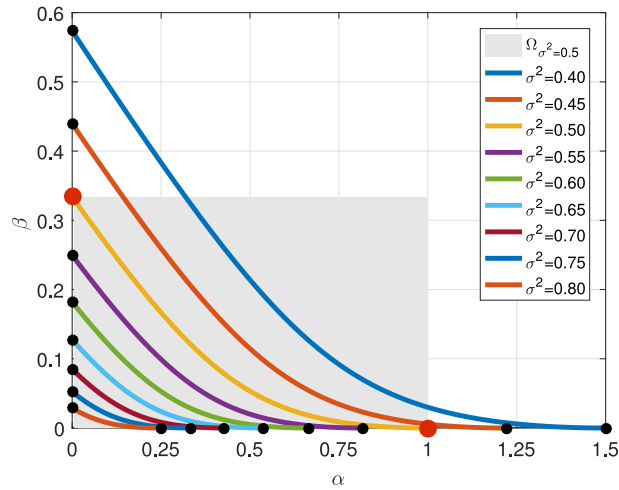
Fig. 12. Equivariance curves. We plot the curves C_σ for selected values of the squared deviation σ^2. Gray area corresponds to the parametric space Ω associated with $\sigma^2 = 0.5$.

For clarity, some of equivariance curves are visualized in Fig. 12. Implications

$$\beta = 0 \quad \Rightarrow \quad \int_0^\infty x^2 \, \wp(x|(\alpha, 0)) \, dx = \frac{2 + \alpha}{1 + \alpha}, \tag{44}$$

$$\alpha = 0 \quad \Rightarrow \quad \int_0^\infty x^2 \, \wp(x|(0, \beta)) \, dx = \frac{\beta \, \mathcal{K}_3(2\sqrt{\beta D})}{D \, \mathcal{K}_1(2\sqrt{\beta D})} \tag{45}$$

and a logical restriction $0 \leqslant \sigma^2 \leqslant 1$ lead to the demarcation of the parametric space Ω_σ (see Fig. 12).

A.3. Testing the stochastic stability for Brownian distance correlation

Here we discuss whether the stochastic process investigated (i.e. vehicular re-scaled clearance analyzed in a fixed flux–density window) is statistically stable with respect to the correlation instrument used. For these intentions we divide the set $X_w = \{x_1, x_2, \ldots, x_Q\}$ of all individual clearances (associated to the wth flux–density window) into subsets

$$Z_1 = \{x_1, x_2, \ldots, x_\delta\}, Z_2 = \{x_{\delta+1}, x_{\delta+2}, \ldots, x_{2\delta}\}, \ldots,$$
$$Z_q = \{x_{(q-1)\delta+1}, x_{(q-1)\delta+2}, \ldots, x_{q\delta}\} \tag{46}$$

where δ is the sampling size and $q = \lfloor Q/\delta \rfloor$. Denoting

$$Z_1^n = \{x_{n+1}, x_{n+2}, \ldots, x_{n+\delta}\}, \quad Z_2^n = \{x_{n+\delta+1}, x_{n+\delta+2}, \ldots, x_{n+2\delta}\}, \ldots,$$
$$Z_q^n = \{x_{n+(q-1)\delta+1}, x_{n+(q-1)\delta+2}, \ldots, x_{n+q\delta}\} \tag{47}$$

the subsets of shifted clearances we then calculate respective sample means

$$r_n(\delta) := \frac{1}{q} \sum_{i=1}^q \mathcal{R}(Z_i, Z_i^n), \tag{48}$$

which are depending on the sampling size. Subsequently, the stochastic stability is tested by means of the functional dependence $r_n = r_n(\delta)$. In Fig. 13 we plot the curves $r_n(\delta)$ enumerated for selected flux–density windows $w \in \{1, 2, \ldots, 7\}$ and shifts $n \in \{1, 3, 5, 7\}$. From all sub-figures it is quite obvious that all processes are stabilizing for increasing δ. To be specific, for $\delta > 2500$ all curves $r_n = r_n(\delta)$ are fluctuating around a certain constant value without any significant trend. Respective limiting values $\lim_{\delta \to +\infty} r_n(\delta)$ are then considered as estimates for correlations analyzed in the third section.
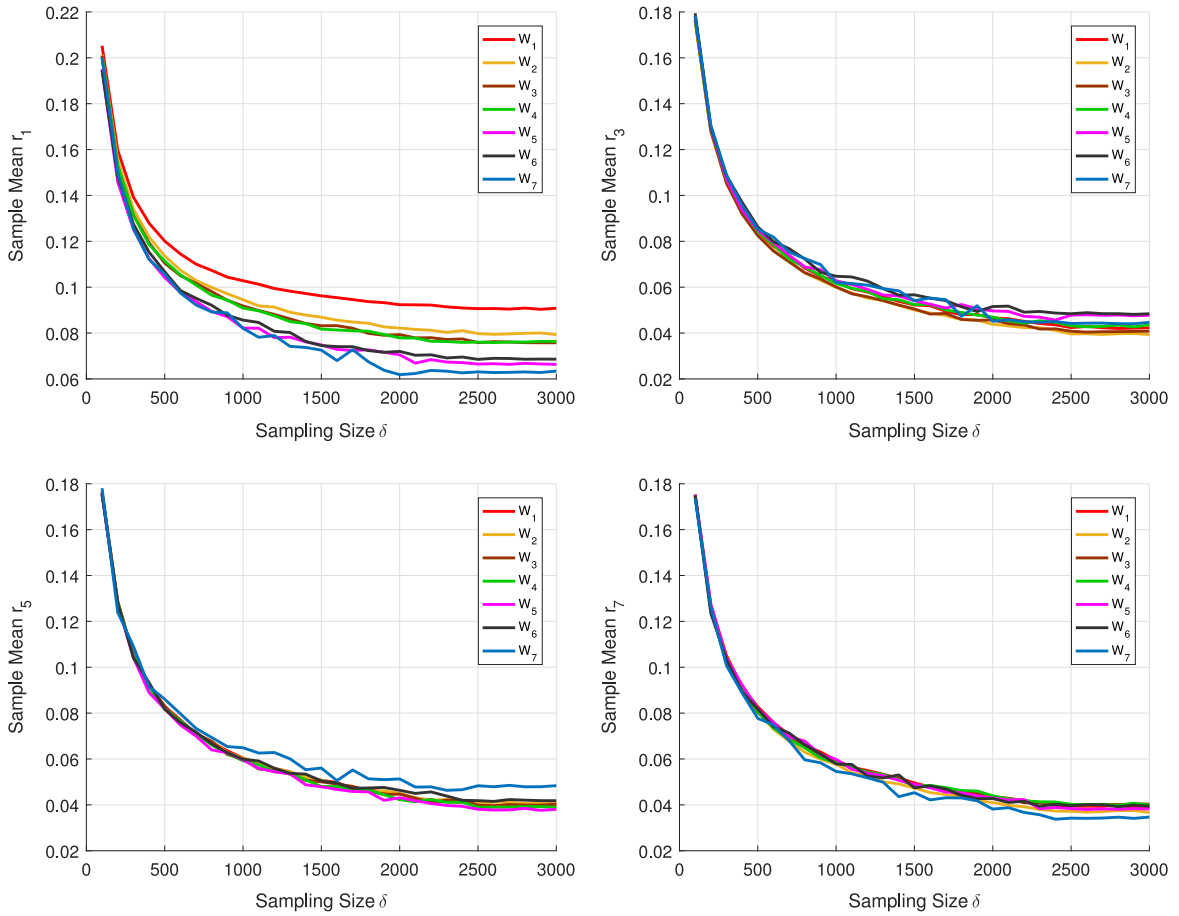
Fig. 13. Test of stability for the correlation instrument used. We show how the distance-correlations $r_n(\delta)$ (see the formula (48)) are changing with the sampling size δ. To illustrate such an evolution we plot four alternatives (for $n = 1$, $n = 3$, $n = 5$, and $n = 7$) analyzed in seven flux–density windows for which the number of headways is sufficient.

A.4. Table of indexing

Here (see Table 3) we summarize all important constants/indexes used in this paper.

Table 3
Indexing in the article.

Index/Constant	Meaning
N	Length of a data sample
$k = 1, 2, \ldots, N$	Index for individual data-item
$\ell = 50$	Size of all sub-samples
$K = \lfloor N/\ell \rfloor$	Number of all sub-samples
$j = 1, 2, \ldots, K$	Index of a sub-sample
$w = 1, 2, \ldots, 9$	Index of a flux–density window (a small sub-region in flux–density map)
$n = 1, 2, \ldots, 10$	Index of a forerunner, i.e. $n = 1$ means a nearest neighbor, $n = 2$ means a next-to-nearest neighbor
$M = 50$	Number of particles on a ring (in the local thermodynamic model)
$m \in \mathbb{N}$	Interaction range (number of actively-followed vehicles)
$\mu \in \mathbb{N}$	Index of a multi-clearance, i.e $\mu = 0$ means a clearance, $\mu = 1$ means a space-gap between three vehicles

References

  [1] M. Krbálek, D. Helbing, Physica A 333 (2004) 370.
  [2] M. Krbálek, J. Šleis, J. Phys. A 48 (2015) 015101.
  [3] M. Krbálek, J. Phys. A 41 (2008) 205004.
  [4] M. Krbálek, P. Šeba, J. Phys. A 42 (2009) 345001.
  [5] M. Krbálek, J. Phys. A 46 (2013) 445101.
  [6] M. Treiber, D. Helbing, Eur. Phys. J. B 68 (2009) 607.
  [7] X. Jin, Y. Zhang, F. Wang, L. Li, D. Yao, Y. Su, Z. Wei, Transp. Res. C 17 (3) (2009) 318.
  [8] D. Helbing, M. Treiber, A. Kesting, Physica A 363 (2006) 62.
  [9] X.Q. Chen, L. Li, R. Jiang, X.M. Yang, Chin. Phys. Lett. 27 (2010) 074501.
[10] D. Helbing, Rev. Modern Phys. 73 (2001) 1067.
[11] K. Nagel, M. Schreckenberg, J. Physique I 2 (12) (1992) 2221.
[12] M. Krbálek, J. Phys. A 40 (2007) 5813.
[13] M.L. Mehta, Random Matrices, third ed., Academic Press, New York, 2004.
[14] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral, Y. Guhr, H.E. Stanley, Phys. Rev. E 65 (2002) 066126.
[15] M. Krbálek, Kybernetika 46 (6) (2010) 1108.
[16] M. Krbálek, T. Hobza, Phys. Lett. A 380 (21) (2016) 1839.
[17] E. Bogomolny, O. Giraud, C. Schmit, Nonlinearity 24 (2011) 3179.
[18] S. Sarkar, S. Jalan, Europhys. Lett. 108 (2014) 48003.
[19] M. Treiber, A. Kesting, Traffic Flow Dynamics, Springer, Berlin, 2013.
[20] B.S. Kerner, The Physics of Traffic, Springer-Verlag, New York, 2004.
[21] M. Tabor, Chaos and Intergrability in Nonlinear Dynamics: An Introduction, Wiley, 1989.
[22] F. Haake, Quantum Chaos, Springer-Verlag, New York, 2001.
[23] D. Helbing, A. Hennecke, V. Shvetsov, M. Treiber, Math. Comput. Modelling 35 (5/6) (2002) 517.
[24] L. Neubert, L. Santen, A. Schadschneider, M. Schreckenberg, Phys. Rev. E 60 (1999) 6480.
[25] S.M. Abuelenin, A.Y. Abul-Magd, IEEE Trans. Intell. Transp. Syst. 16 (5) (2015) 2543.
[26] C. Appert-Rolland, Phys. Rev. E 80 (2009) 036102.
[27] G.J. Székely, M.L. Rizzo, Ann. Appl. Stat. 3/4 (2009) 1236.
[28] G.J. Székely, M.L. Rizzo, N.K. Barikov, Ann. Statist. 35 (2007) 2769.
[29] D. Helbing, M. Treiber, Phys. Rev. E 68 (2003) 067101.
[30] M. Krbálek, Quantitative analysis of interaction range in vehicular flows, Transp. Res. Procedia 25 (2017) 1268–1275.
[31] B. Jörgensen, Statistical Properties of the Generalized Inverse Gaussian Distribution, in: Lecture Notes in Statistics, vol. 9, Springer, Heidelberg, 1982.
[32] N.L. Johnson, S. Kotz, N. Balakrishnan, Continuous Univariate Distributions, Vol. 1, second ed., John Wiley & Sons, 1994.
[33] M. Krbálek, P. Hrabák, M. Bukáček, Physica A 490 (2018) 38–49.