



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FAKULTA STAVEBNÍ

ÚSTAV POZEMNÍCH KOMUNIKACÍ

STATISTICKÉ VLASTNOSTI
MIKROSTRUKTURY DOPRAVNÍHO
PROUDU

TEZE DISERTAČNÍ PRÁCE

Ing. Jiří Apeltauer

BRNO 2018

Obsah

1	Úvod	3
2	Teorie dopravního proudu	3
3	Data	4
3.1	Měření	4
3.2	Makro a mikrocharakteristiky datových souborů	5
4	Použité statistické nástroje	7
4.1	Shluková analýza K-means	7
4.2	Hustota pravděpodobnosti	8
4.3	Inverzní Gaussova hustota pravděpodobnosti (GIG)	9
4.4	Statistická vzdálenost χ	11
4.5	Konvoluce	12
4.6	Kolmogorovův-Smirnovův test	13
5	Fitovací procedura	14
5.1	Příprava dat	14
5.2	Separace ovlivněného dopravního proudu	15
5.3	Výpočet multi tc	16
5.4	Dvouparametrická estimační metoda	17
5.5	Dvouparametrická estimační metoda s využitím ekvivalentního rozptylu	19
5.6	Definiční obor estimace	21
5.7	Výsledek estimace	21
6	Odhad interakční vzdálenosti vozidel	24
6.1	Aplikace principu konvoluce	24
6.2	Detekce interakční vzdálenosti pomocí Kolmogorov-Smirnovova testu	25
7	Závěr	27
	Reference	29

1 Úvod

Současná teorie dopravního proudu vychází z předpokladu, že interakce probíhá pouze mezi sousedními, v čase po sobě následujícími, vozidly, respektive že vozidla mění svoji rychlost na základě rychlosti vozidla před sebou a vozidla vzdálenější nebere v potaz. Na tomto předpokladu byly založeny v 50. letech minulého století tzv. car-following modely dopravního proudu (modely následného vozidla), které od té doby prošly dlouhým vývojem a v současnosti jsou schopny poměrně přesně napodobit chování reálných vozidel. Příjmutí tohoto zjednodušení je zcela legitimní, i když je zřejmé, že reálný dopravní proud takto nefunguje. Chceme-li ovšem posunout teorii dopravního proudu dále, musíme modely dopravního proudu doplnit o širší interakci mezi vozidly tak, jak ji aplikují řidiči v reálném provozu. Tato práce předkládá postup založený na metodách matematické statistiky, který je schopen širší interakce mezi vozidly odhalit a do jisté míry i kvantifikovat.

2 Teorie dopravního proudu

Pravděpodobně jako úplně první se o dopravní proud začal zajímat Bruce D. Greenshields, který v roce 1933 na 13. výročním setkání Highway Research Board přednesl svoje výsledky. Greenshields navrhl velice efektivní a nadčasovou metodu měření. Použil totiž analýzu obrazu, metodu, která se v dnešní době díky pokroku elektroniky dostává do popředí. Greenshieldsův přístup byl samozřejmě manuální, nicméně geniální. Ve známé vzdálenosti od komunikace postavil fotoaparát a provedl něco, co bychom dnes pojmenovali časosběr (tedy v diskrétních časových krocích pořizoval fotografické snímky). Měření probíhalo na úseku dlouhém přibližně 125 stop (38 m) s časovým odstupem snímků 0,5–2s. Ze změny polohy automobilu na fotografiích poté mohl snadno spočítat základní veličiny teorie dopravního proudu, které byly právě v tento moment poprvé definovány. Jednalo se o rychlost, hustotu a intenzitu (kterou nejdříve Greenshields označoval jako hustotu druhého druhu neboli “Density-vehicles per Hour”) dopravního proudu. Rychlost je definována jako průměrná rychlost vozidel, která protnou sledovaný profil za daný čas, hustota je počet vozidel na jednotku délky komunikace a intenzita je počet vozidel, která protnou sledovaný profil za daný čas. Greenshields rovněž definoval vztahy těchto veličin.

Základy teorie dopravního proudu byly tedy položeny. Greenshieldsovy veličiny popisují dopravní proud jako celek neboli popisují jeho makroskopickou strukturu, protože agregují jednotlivá vozidla do větších celků. Chování jednotlivých vozidel nebo jejich interakce z nich tedy nevyčteme. S narůstajícími intenzitami dopravy se však začaly objevovat i první kolony a vyvstala otázka, kde leží ona hranice, kdy komunikace již není schopna dopravu pojmout a jak tuto hranici stanovit (nebo spíše odhadnout). Na řešení tohoto problému už makroskopický pohled nestačí, bylo potřeba pochopit vztahy mezi jednotlivými vozidly, tedy mikrostrukturu dopravního proudu.

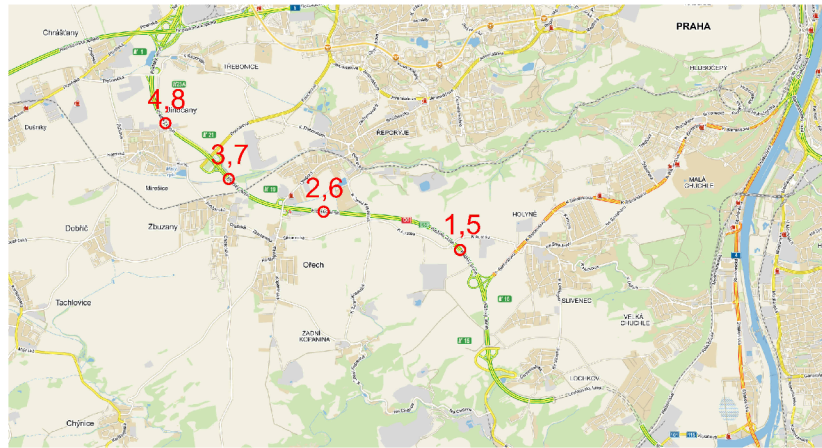
Jako jeden z prvních popsal tyto vztahy Luise A. Pipes ve své práci [16]. Brzy zjistil, že analytické řešení soustavy rovnic popisující chování dopravního proudu na základě interakcí mezi jednotlivými vozidly nebude vůbec jednoduché, pokud je vůbec možné. Nakonec uspěl pro případ skupiny vozidel za specifických počátečních podmínek a musel akceptovat fakt, že dopravní proud je natolik komplexní systém, že jeho modelování nebude bez využití nume-

rického přístupu a počítačů možné. Pipesův model měl nedostatky. Nezohledňoval například reakční dobu řidiče, měl pevně dané parametry odstupů, které se ve skutečnosti pro každé vozidlo liší atd., ale byl důležitým prvním krokem v oblasti tzv. car-following modelů, které jsou základem moderní teorie dopravního proudu.

3 Data

3.1 Měření

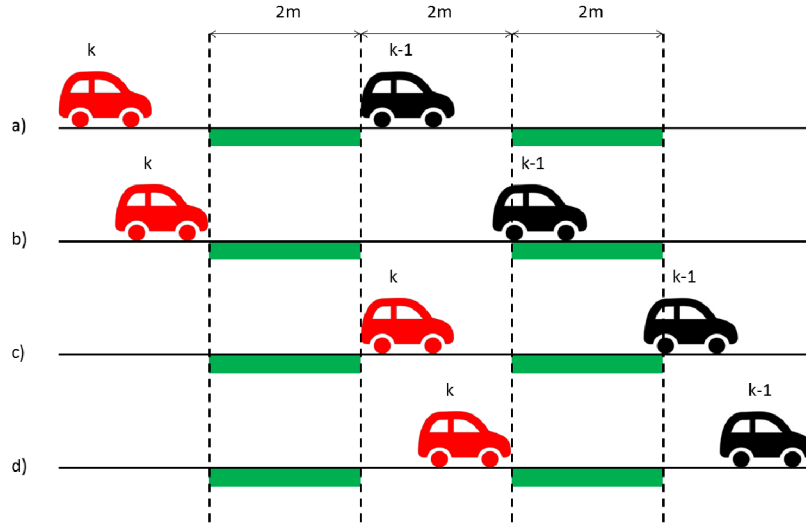
K získání statisticky významných výsledků je nutné zajistit dostatečný počet pozorování v řádech milionů záznamů, a aby bylo možné pozorovat mikrostrukturu dopravního proudu, musí se jednat o záznamy jednotlivých vozidel. Tuto podmínku splňují detekční profily umístěné na Pražském velkém městském okruhu (obr. 1), jimiž naměřená data se podařilo získat.



Obrázek 1: Přibližné umístění detekčních profilů na pražském okruhu mezi 16. a 22. kilometrem. Profily 1-4 jsou ve směru severo-západ (po směru staničení komunikace), profily 5-8 poté v protisměru. V každém směru jsou vedeny dva jízdní pruhy.

Vzhledem k tomu, že se jedná o permanentní instalaci sloužící k liniovému řízení jsou data zaznamenávána 24 hodin, 7 dní v týdnu v námi požadovaném rozlišení vozidlo co vozidlo. Data v surové podobě jsou zaznamenávána na lokální server, odkud jsou po agregaci odeslána na centrálu liniového řízení. Bylo tedy nutné fyzicky navštívit jednotlivé sčítací profily a datové soubory stáhnout přímo z lokálních serverů (v dalších částech práce tedy budeme mluvit o datových souborech, jejichž čísla jsou totožná s čísly profilů na obr. 1). Rozsah těchto souborů je značný. Pokrývají období od listopadu 2013 do května 2014, celkový počet záznamů v jednotlivých souborech se pohybuje kolem 6 milionů.

Konstrukce detektoru je dvojitá indukční smyčka, která umožňuje měřit kromě četností i rychlosti vozidel. Smyčka měla délku 2 metry a rozestup náběžných hran smyček byl 4 m. Měřeny jsou celkem 4 události, ze kterých jsou následně dopočítávány charakteristiky dopravního proudu a vozidel (podrobně viz. obr. 2).



Obrázek 2: Schéma detektoru. Zaznamenávané události jsou a) $\tau_{k-1}^{(a)}$ čas, kdy zadní nárazník $(k-1)$ -ho vozidla protne výstupní hranu první smyčky, b) $\tau_k^{(b)}$ čas, kdy přední nárazník k -tého vozidla protne vstupní hranu první smyčky, c) $\tau_k^{(c)}$ čas, kdy zadní nárazník k -tého vozidla protne výstupní hranu první smyčky a d) $\tau_k^{(d)}$ čas, kdy přední nárazník k -tého vozidla protne vstupní hranu druhé smyčky.

3.2 Makro a mikrocharakteristiky datových souborů

Časové události uvedené na obr. 2 jsou na úrovni detektoru zpracovány do podoby základních mikrocharakteristik dopravního proudu. Jako první je dle chronologie měření zjištěna časová mezera (gap) mezi $(k-1)$ a k -tým vozidlem

$$g_k = \tau_k^{(b)} - \tau_{k-1}^{(a)}. \quad (1)$$

Z pohledu datového řádku tedy mezera před vozidlem. Vzhledem k tomu, že vozidlo má nenulovou délku, nejedná se o čistou časovou mezera, tu je nutné dopočítat (7). Následně je změřena obsazenost (occupancy), když k -té vozidlo opustí vstupní smyčku

$$o_k = \tau_k^{(c)} - \tau_k^{(b)} \quad (2)$$

a jako poslední je změřena cestovní doba (travel time), když přední nárazník k -tého vozidla protne vstupní hranu druhé smyčky

$$tt_k = \tau_k^{(d)} - \tau_k^{(b)}. \quad (3)$$

Tím je ukončeno měření k -tého vozidla a díky známým rozměrům detektoru je možné spočítat zbývající charakteristiky, které jsou však zatíženy neznámou chybou. Ta je způsobena faktem, že rychlost vozidla mezi náběžnými hranami smyček není konstantní. Zejména u malých rychlostí a vysokých hustot může být tato chyba pro jednotlivá vozidla nezanedbatelná.

Můžeme však rovněž předpokládat, že tato chyba bude mít normální rozdělení a na statistiky z velkého množství záznamů nebude mít významný vliv. Rychlost

$$v_k = l_{det}/tt_k \quad (4)$$

a délka

$$l_k = (v_k \cdot o_k) - l_{ic} \quad (5)$$

vozidla je počítána na úrovni detektoru, kdy l_{det} je vzdálenost vstupních hran indukčních smyček a l_{ic} je délka indukční smyčky. Dle dokumentace dodavatele je $l_{det} = 4$ m a $l_{ic} = 2$ m. Analýza dat ale ukázala, že v praxi se mohou tyto délky lišit (například pro datový soubor 1 je vypočtená střední hodnota $l_{ic} = 1,958$ m při zachování předpokladu že $l_{det} = 4$ m) a jsou pravděpodobně výsledkem kalibrace konkrétního detektoru po jeho instalaci. Pro potřeby práce bylo nutné dopočítat řadu dalších charakteristik. První veličina, která je přímo odvozena z předcházejících, je časový odstup th (time headway). Jedná se o časovou mezeru mezi protnutím stejného místa stjenou částí (například zadní nárazník) $(k - 1)$ -ho a k -tého vozidla

$$th_k = g_k + o_k = \tau_k^{(c)} - \tau_{k-1}^{(a)}. \quad (6)$$

Následně můžeme odvodit aproximativní hodnotu tc (time clearance), což je čistá časová mezera neboli odstup mezi protnutím stejného místa zadním nárazníkem $(k - 1)$ -ho vozidla a předním nárazníkem k -tého vozidla

$$tc_k = th_k - \frac{l_k \cdot 3,6}{v_k}. \quad (7)$$

Tímto máme definovány všechny potřebné charakteristiky mikrostruktury. Abychom mohli definovat charakteristiky makrostruktury dopravního proudu, je nutné zvolit počet vozidel nebo časový krok, pro který budeme data agregovat. Pro účely práce byla zvolena agregace podle počtu po sobě jdoucích vozidel, kdy $m = 50$ voz (optimální počet pro stanovení střední hodnoty časových odstupů [2]). Pro každou takovouto m -tici vozidel můžeme poté definovat intenzitu

$$I_j = \frac{m}{\sum_{k=1}^m th_k} 3600 \quad [\text{voz/h}], \quad (8)$$

hustotu

$$P_j = \frac{I_j}{\frac{1}{m} \sum_{k=1}^m v_k \cdot 3,6} \quad [\text{voz/km}], \quad (9)$$

a rychlost dopravního proudu

$$V_j = \frac{1}{m} \sum_{k=1}^m v_k \cdot 3,6 \quad [\text{km/h}], \quad (10)$$

kdy j značí pořadí dané m -tice v datovém souboru.

Pro každou m -tici je ještě vypočten normovaný odstup vozidel s , který je nezbytný pro účely matematické statistiky. Jedná se o normovanou hodnotu čistého časového odstupu tc ,

kdy střední hodnota normovaných odstupů v dané m -tici je rovna 1 (respektive $n + 1$, viz. 5.3)

$$s_k = \frac{tc_k}{\frac{1}{m} \sum_{k=1}^m tc_k}, \quad (11)$$

Abychom měli představu o struktuře dat kompletní, uveďme pro příklad souhrnnou statistiku pro datový soubor 1 (tab. 1, jednotlivé soubory jsou si ze statistického hlediska velice podobné). Tato statistika shrnuje celou dobu měření a byla provedena po očištění od řádků ve kterých rychlost nebo čistý časový odstup překročil hodnotu quantilu 99 % nebo klesl pod hodnotu quantilu 1 % a jednalo se tedy o chybová nebo odlehlá měření.

Tabulka 1: Statistické charakteristiky základních veličin, datový soubor 1.

	Levý jízdní pruh -0	Pravý jízdní pruh -1
Rychlost, průměr [km/h]	108,9	92,5
Rychlost, s. odchylka [km/h]	15,1	12,5
Rychlost, min. [km/h]	33,0	45,0
Rychlost, median [km/h]	110,0	91,0
Rychlost, max [km/h]	144,0	130,0
tc , průměr [s]	3,4	3,8
tc , s. odchylka [s]	4,4	3,9
tc , min. [s]	0,3	0,4
tc , median [s]	1,7	2,5
tc , max [s]	29,5	27,1
Podíl O	88,9	62,3
Dodávky	5,9	6,2
Nákladní	4,2	26,2
Ostatní	1,3	5,3

4 Použité statistické nástroje

4.1 Shluková analýza K-means

Pojem *k-means* poprvé použil ve své práci James MacQueen v roce 1967 [14]. Základy metody však sahají dále do historie. Otcem původní myšlenky je Hugo Steinhaus a formuloval ji v roce 1957 [18]. První algoritmus pro její řešení navrhl ve stejném roce Stuart Lloyd [13]. Matematická formulace je následující [20]: Mějme pozorování (x_1, x_2, \dots, x_n) , kde každé pozorování je d -dimenzionální reálný vektor. Cílem procesu je rozdělit n pozorování do k ($k \leq n$) shluků $S = \{S_1, S_2, \dots, S_k\}$ tak, abychom v každém shluku minimalizovaly součet všech čtverců (vzdálenost jednotlivých pozorování od středu shluku) (12)

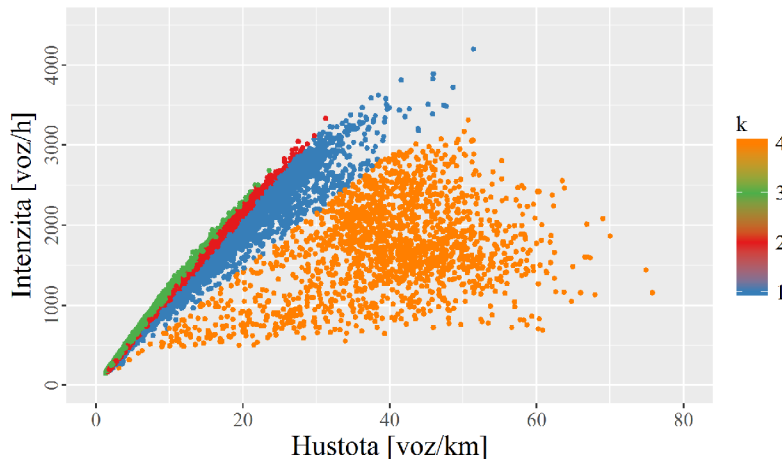
$$\arg \min \sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|^2, \quad (12)$$

kde μ_i je střední hodnota pozorování v S_i . Výpočet probíhá v krocích. V prvním kroku jsou (zpravidla náhodně) vybrány středy shluků (k náhodně vybraných pozorování), tak zvané centroidy. Pomocí Voroniovovy teselace je prostor rozdělen na k oblastí a do nich jsou přiřazena jednotlivá pozorování. Z těchto pozorování jsou spočteny nové centroidy pro danou oblast,

které nově leží v jejím těžišti a proces teselace se opakuje. Tyto dva kroky běží ve smyčce, dokud není dosažena podmínka (12) nebo maximální počet iterací.

Na obr. 3 je ukázka aplikace na soubor dat 1. Parametrem pro hledání shluků je rychlost a byly hledány 4 shluky (kvůli náhodné volbě počátečních centroidů neodpovídá jejich číslování uspořádání v grafu).

Fundamentální diagram po aplikaci K-means, soubor dat 1, pruh 0



Obrázek 3: Výsledky metody K-means při separaci fundamentálního diagramu podle rychlosti do $k = 4$ shluků: S_4 ($\mu_1 = 45$ km/h, 87 350 m -tic), S_1 ($\mu_i = 99,3$ km/h, 423 350 m -tic), S_2 ($\mu_3 = 109,2$ km/h, 1 356 800 m -tic) a S_3 ($\mu_2 = 117$ km/h, 1 032 050 m -tic).

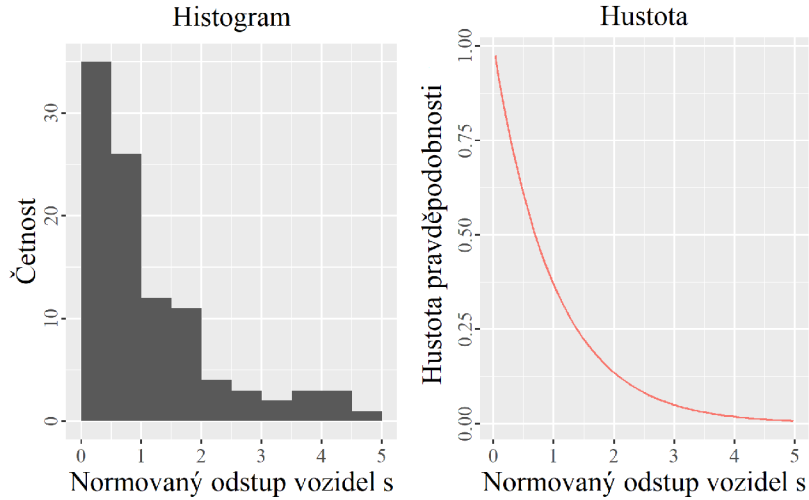
4.2 Hustota pravděpodobnosti

K popisu pravděpodobnosti výskytu náhodného diskrétního jevu používáme tak zvanou pravděpodobnostní funkci, která každému výsledku předepisuje pravděpodobnost výskytu. Součet pravděpodobností výskytu jednotlivých jevů je roven jedné. V případě spojité veličiny je pravděpodobnostní funkce nahrazena hustotou pravděpodobnosti $f(x)$, která má jistou podobnost s histogramem dané veličiny, například normovaným odstupem vozidel, jak je vidět na obrázku 4. Narozdíl od pravděpodobnostní funkce, respektive diskrétní veličiny, nemá smysl určovat pravděpodobnost výskytu konkrétního jednoho jevu, například časové mezery o délce 1 s, protože tato pravděpodobnost by byla vždy rovna nule. V případě spojité veličiny je pravděpodobnost dána plochou pod křivkou hustoty pravděpodobnosti, čili celá tato plocha musí být rovna jedné a zároveň plocha pod jedním bodem je rovna nule. Můžeme určit pouze pravděpodobnost výskytu jevu v určitém rozsahu (například výskyt časové mezery v rozsahu 0,5-1,5 s) jakožto integrál plochy pod křivkou.

Pro hustotu pravděpodobnosti výskytu časové mezery určité délky se v dopravním inženýrství používá například exponenciální rozdělení hustoty pravděpodobnosti [15]

$$g(s) = \Theta(s)\lambda e^{-\lambda s}, \quad (13)$$

kde $\lambda > 0$ je parametr související s danou veličinou, v našem případě normovaným časovým odstupem vozidel s a Θ je Heavisidova funkce



Obrázek 4: Podobnost histogramu a hustoty pravděpodobnosti (aproximace odstupů pomocí exponenciálního rozdělení (13)).

$$\Theta(s) = \begin{cases} 1 & \text{pro } s > 0 \\ 0 & \text{pro } s \leq 0 \end{cases} \quad (14)$$

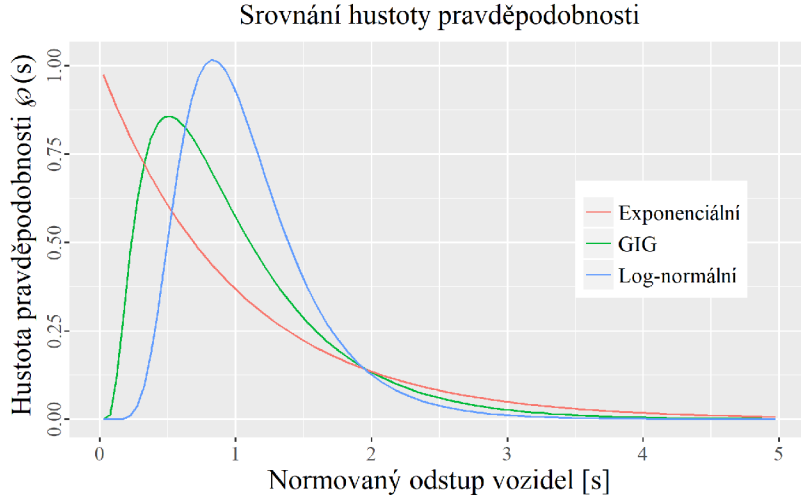
Tato aproximace je poměrně dobře použitelná při dopravních stavech s nižší hustotou dopravy. Při vysokých hustotách však začne selhávat (jak je vidět na obr. 5). Pravděpodobnost pro malé odstupy tak predikuje extrémně nepřesně. Lepší aproximaci empirických dat získáme s takzvaným dvouparametrickým log-normálním rozdělením (15)[17], které již respektuje minimální velikost časového odstupů mezi vozidly při vysokých hustotách

$$g(s) = \Theta(s) \frac{1}{s\sigma\sqrt{2\pi}} e^{-\frac{(\ln s - \mu)^2}{2\sigma^2}}, \quad (15)$$

kde σ a μ jsou kalibrační parametry. Nicméně ani tato aproximace však není ideální, a to z důvodu vyšších matematických principů, které jsou již nad rámec této práce a znalostí autora. V následující kapitole (4.3) je tedy popsána třetí hustota pravděpodobnosti, u které bylo v několika publikacích prokázáno, že je schopna velice dobře aproximovat hustotu pravděpodobnosti časových odstupů vozidel za všech zkoumaných podmínek.

4.3 Inverzní Gaussova hustota pravděpodobnosti (GIG)

Existuje jistá a poměrně překvapivá podobnost mezi chováním termodynamického plynu a volného dopravního proudu (překvapivá je proto, že termodynamický plyn je rovnovážná soustava). Stejně jako v dopravě, tak i u termodynamického plynu není možné přímo měřit interakce mezi jednotlivými vozidly nebo částicemi. Avšak pro termodynamický plyn se podařilo M. Krbálkovi odvodit statistické metody [2], které tyto interakce popisují nepřímou na základě sledování rozdělení hustoty pravděpodobnosti jistých mikroskopických veličin. Tuto metodu Krbálek úspěšně aplikoval na dopravní proud [9]. Bylo však potřeba splnit některé předpoklady. Rovnice použité pro termodynamický plyn jsou platné vždy pouze pro konkrétní



Obrázek 5: Srovnání exponenciální ($\lambda = 1$, (13)), log-normální ($\sigma = 0,43$; $\mu = 0$; (15)) a GIG ($\alpha = 0$; $\beta = 0,43$; (16)) hustoty pravděpodobnosti.

system o určité celkové energii. Pokud chceme zkoumat systém o jiné energii, je potřeba parametry rovnice upravit. Tento předpoklad je tedy nutné aplikovat i v případech, že chceme stejnými metodami zkoumat dopravní proud a je nezbytné jej rozdělit na úzké hustotní intervaly.

V těchto intervalech je následně sledováno rozdělení hustoty pravděpodobnosti časových odstupů vozidel. Studie prokázala, že tento postup lze k popisu interakcí mezi vozidly úspěšně použít a navíc autoři prokázali, že chování řidičů není ve všech stavech dopravního proudu stejné a mění se v závislosti na hustotě. Toto zjištění je poměrně zásadní, jelikož v současnosti používané dopravní modely a teorie užívají zjednodušující předpoklad, že chování řidiče je při různých stavech dopravního proudu konstantní.

Krbálek tento postup dále rozvíjí pro případ rozdělení hustoty pravděpodobnosti časových odstupů vozidel čekajících na červený signál na křižovatce [6] a potvrdil předpoklad, že chování kondenzovaného dopravního proudu může být lokálně (pro úzké hustotní pásmo) chápáno jako chování termodynamického systému (plynu). V následujících pracích [6] [12] [7] [11] Krbálek postupně formuluje jednoparametrické rozdělení hustoty pravděpodobnosti (nazve jej GIG - generalized inverse gaussian distribution), které velice dobře aproximuje hustotu pravděpodobnosti časových odstupů vozidel. V jedné ze svých posledních prací [10] poté doplňuje i druhý parametr. Právě toto dvouparametrické rozdělení hustoty pravděpodobnosti je použito v této práci (obr. 5) a je definováno rovnicí

$$\varphi(s) = A\Theta(s)s^\alpha e^{-\beta/s - Ds}, \quad (16)$$

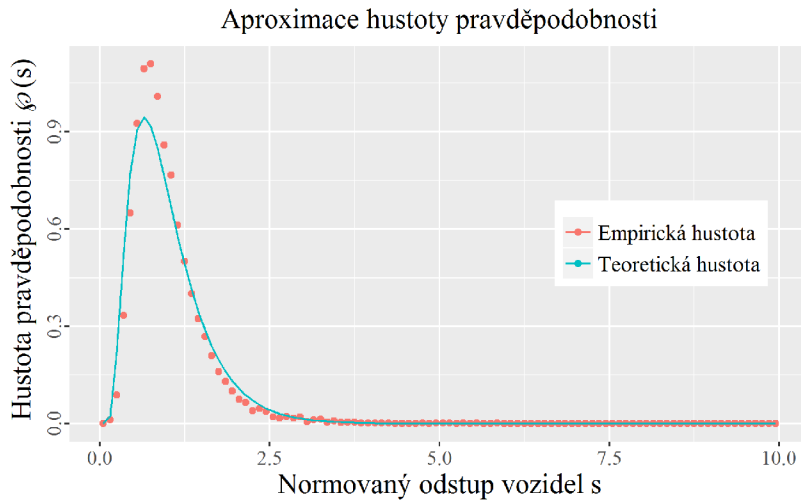
kde

$$D = \alpha + \beta + \frac{3 - e^{-\sqrt{\beta}}}{2} \quad (17)$$

a

$$A^{-1} = 2\left(\frac{\beta}{D}\right)^{\frac{\alpha+1}{2}} \mathcal{K}_{\alpha+1}(2\sqrt{\beta D}). \quad (18)$$

Význam jednotlivých členů rovnic je následující: α a β jsou parametry funkce odhadnuté estimačním procesem (tzv. fitováním), s je sledovaná veličina (v našem případě normovaný časový odstup vozidel) a \mathcal{K}_α je Mac-Donaldova funkce řádu α . Člen $e^{-\beta/s}$ řídí funkci v oblastech $s \rightarrow 0$ (obr. 6, $s < 0,5$). Jelikož v reálném dopravním proudu nebude nikdy časová mezera (respektive vzdálenost) mezi vozidly nulová, je i hustota pravděpodobnosti výskytu nulové časové mezery (a blízkých hodnot) rovna nule. Člen e^{-Ds} pak řídí exponenciální chvost funkce (obr. 6, $s \gg 2$), kdy se teoretická časová mezera mezi vozidly blíží nekonečnu (v praxi je maximální možná mezera dána nastavením detektoru).



Obrázek 6: Výsledek fitovacího procesu pro datový soubor 1; pruh 0; $\rho = 55$ [voz/h]; $\alpha = 0$; $\beta = 0,96$.

Jak je naznačeno v předchozím textu, odhad parametrů α a β je platný vždy jen pro úzké rozpětí hustoty (v této práci bylo zvoleno $\Delta\rho = 5$ voz/km a pro zjednodušení zápisu je dále v práci vždy uváděna pouze horní hranice rozpětí), která do jisté míry časové odstupy s determinuje a je zřejmé, že absolutní hodnoty s se budou pro různé hustoty a pozorování lišit. Abychom tedy aproximační funkci (16) mohli použít pro všechna hustotní pásma (případně různá měření) a zároveň mohli jednotlivé výsledky srovnávat, musíme časové odstupy s normovat na střední hodnotu 1.

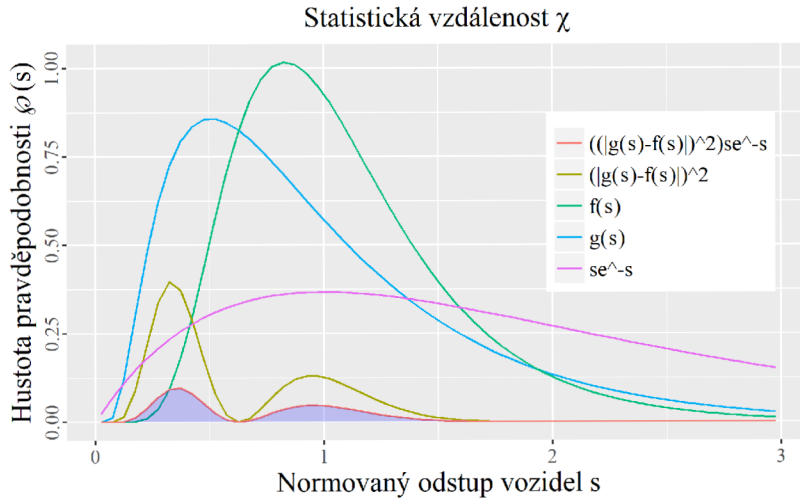
4.4 Statistická vzdálenost χ

Jak bylo zmíněno v předchozí kapitole, parametry α a β funkce (16) jsou získány estimačním procesem, neboli fitováním. Při tomto procesu jsou do funkce (16) postupně dosazovány různé hodnoty α a β . Získané teoretické hustoty pravděpodobnosti jsou srovnávány s hustotami empirickými (obr. 6). Srovnávací kvantitou je statistická vzdálenost χ [11] a cílem je pro dané hustotní pásmo a definiční obor α a β najít minimum funkce

$$\chi_{(\alpha,\beta)} = \int_0^\infty |\wp_{\alpha,\beta}(s) - h(s)|^2 se^{-s} ds, \quad (19)$$

kde $\wp_{\alpha,\beta}(s)$ představuje teoretickou hustotu pravděpodobnosti podle (16), $h(s)$ je empirická hustota pravděpodobnosti a se^{-s} představuje váhu, pomocí které určíme jaká fitovaná oblast

je pro nás „důležitější“. Grafická reprezentace χ je zobrazena na obr. 7. Podrobněji je využití statistické vzdálenosti v procesu iterace popsáno v kapitole 5.4.

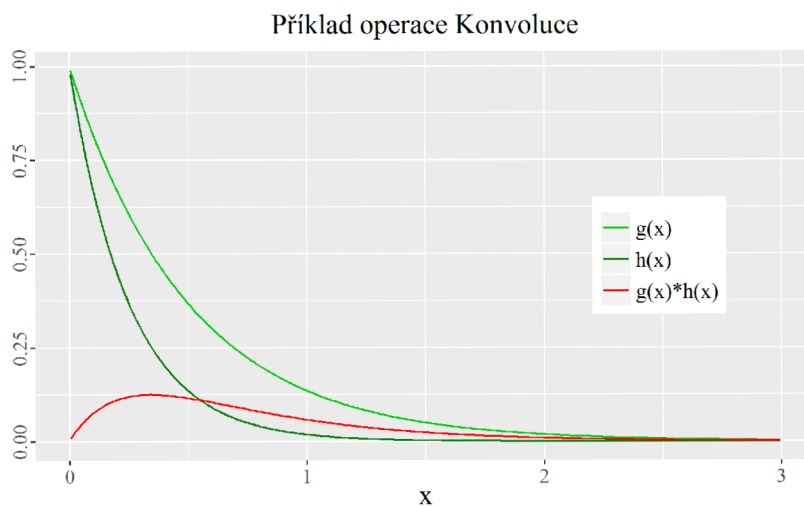


Obrázek 7: Statistické vzdálenosti χ (19) (modrá plocha). Aplikace váhy se^{-s} zvýší význam exponenciální části funkce.

4.5 Konvoluce

Konvoluce je matematická operace, která zpracovává dvě funkce a je definována předpisem [21]

$$g(x) * h(x) = \int_0^x g(y) \cdot h(x - y) dy. \quad (20)$$



Obrázek 8: Výsledek konvoluce (20) funkcí $g(x) = \Theta(x)e^{-xa}$ a $h(x) = \Theta(x)e^{-xb}$ pro $a = 2$ a $b = 4$.

Algoritmicky probíhá proces tak, že funkci $h(x)$ otočíme zrcadlově kolem osy y a ve zvolených krocích ji posunujeme po ose x zleva doprava. V každém kroku vypočteme součin funkcí a plocha kterou nám tento součin vytvoří je hodnotou konvoluce pro dané x . Konvoluce je využívána hojně při procesech zpracování signálu, kdy je používána jako filtr nebo maska. Výsledek procesu je vidět na obr. 8. Pro tuto práci je klíčové statistické využití konvoluce, kdy lze s její pomocí vypočítat kombinovanou hustotu pravděpodobnosti dvou nezávislých veličin (kap. 6.1).

4.6 Kolmogorovův-Smirnovův test

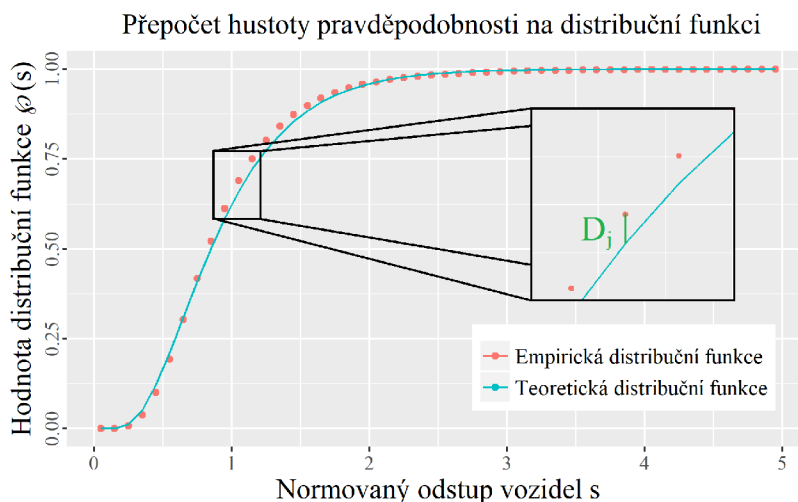
Základní jednovýběrový Kolmogorovův-Smirnovův test se používá k ověření hypotézy, zda náhodný výběr pochází z daného rozdělení pravděpodobnosti (tzv. nulová hypotéza H_0). Základem testu je určení maximální odchylky spojitě (teoretické) distribuční funkce $F_0(x)$ a výběrové (empirické) distribuční funkce $F_n(x)$ (21)

$$D_j = \sup_{-\infty < x < \infty} |F_j(x) - F_0(x)| = \max(D_1^*, D_2^*, \dots, D_j^*), \quad (21)$$

kde

$$D_i^* = |F_j(x_i) - F_0(x_i)|, \quad i = 1, 2, \dots, j. \quad (22)$$

Graficky je hodnota D_j znázorněna na obr. 9. Je-li hodnota D_j větší než kritická, nulová hypotéza se zamítá. Pro $j < 40$ je kritická hodnota tabelována, pro $j > 40$ a hladinu významnosti 5 % se vypočte ze vztahu $D_{krit} = 1,36\sqrt{2/j}$. Pokud teoretickou distribuční funkci $F_0(x)$ nahradíme druhým náhodným výběrem, můžeme provést dvouvýběrový Kolmogorovův-Smirnovův test, při kterém testujeme hypotézu, zdali mají dva náhodné výběry stejné rozdělení pravděpodobnosti (H_0).



Obrázek 9: Grafická reprezentace D_j (21). Na rozdíl od statistické vzdálenosti χ (obr. 7) Kolmogorovská vzdálenost D_j zohledňuje pouze maximální odchylku mezi funkcemi.

5 Fitovací procedura

5.1 Příprava dat

K dispozici bylo celkem 8 datových souborů o velikosti 3,2 GB. Měření bylo provedeno stejnou technologií ve všech případech a na jedné komunikaci (viz. 3.1). Struktura všech datových souborů je tedy stejná, což jejich zpracování značně ulehčilo. Jako příklad bude uvedeno zpracování datového souboru 1, formát surových dat je v tabulce 2. Význam jednotlivých sloupců je následující:

Date Time Datum a časová známka záznamu průjezdu vozidla.

L Jízdní pruh (Levý jízdní pruh–0, Pravý jízdní pruh–1).

Occ Obsazenost detektoru (2.)

Gap Časová mezera (1).

Ttime Cestovní čas (3).

Speed Rychlost vozidla (4).

Length Délka vozidla (5).

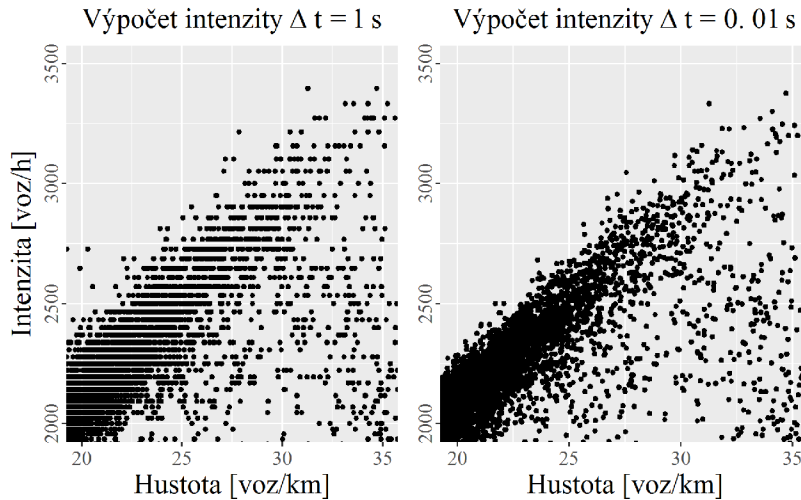
Category Kategorie vozidla (přiřazena podle délky).

Tabulka 2: Ukázka surových dat, soubor 1.

Date Time	L	Occ[s]	Gap[s]	Ttime[s]	Speed[km/h]	Length[m]	Category
2013-11-20 08:24:29	0	0,37	0,63	0,23	61	4,20	car
2013-11-20 08:24:30	0	0,34	0,48	0,23	63	4,10	car
2013-11-20 08:24:30	1	0,46	3,99	0,22	66	6,00	lor
2013-11-20 08:24:31	0	0,35	0,24	0,23	62	4,20	car
2013-11-20 08:24:31	0	0,37	0,15	0,23	62	4,30	car
2013-11-20 08:24:31	1	0,35	0,55	0,23	63	4,20	car
2013-11-20 08:24:32	0	0,37	0,65	0,25	57	3,90	car
2013-11-20 08:24:33	1	0,29	2,05	0,18	80	4,10	car
2013-11-20 08:24:34	0	0,33	1,47	0,23	63	3,90	car

Datový soubor splňuje veškeré požadavky potřebné pro tuto práci, vyjma přesnosti časové známky záznamu. Při vyšších intenzitách a rychlostech dochází k diskredizaci intenzit (obr. 10 vlevo). Naštěstí je možné kontinuální čas dopočítat s přesností 0,01 s ze sloupců Gap a Occ (6) a intenzity spočítat s potřebnou přesností (obr. 10 vpravo).

Jakmile je datový soubor doplněn o kontinuální čas, čistý časový odstup tc (7) a časový odstup th (6), můžeme jej rozdělit na jednotlivé m -tice a vypočítat lokální charakteristiky dopravního proudu (pro účely práce bylo zvoleno $m = 50voz$, výpočty jsou uvedeny v kapitole 3.2). Jednotlivé m -tice je nutné zařadit do hustotních pásem $Pf_j = \{5,10,15,\dots,80\}voz/h$ (intervaly jsou zprava otevřené, např. pro $P_j \in \langle 5,10 \rangle \Rightarrow Pf_j = 10$), pro které jsou následně počítány odhady parametrů α a β . Iterační proces rovněž vyžaduje normování tc a th na střední hodnotu 1 (viz. 4.3).



Obrázek 10: Ukázka diskredizace fundamentálního diagramu při nedostatečné přesnosti měření času průjezdu vozidel.

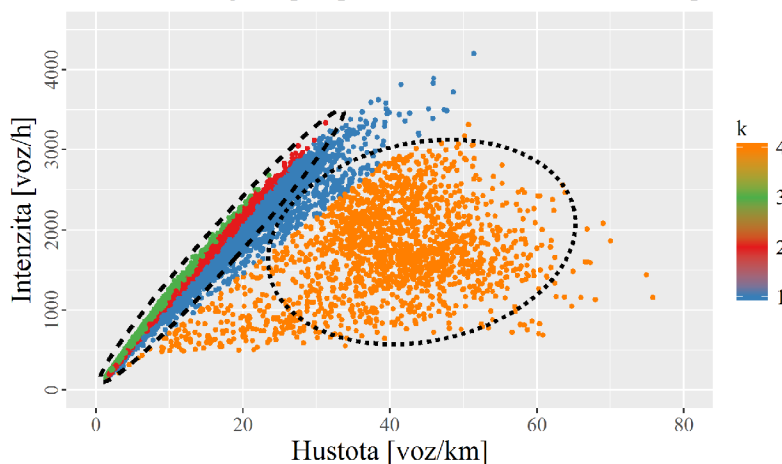
5.2 Separace ovlivněného dopravního proudu

Vzhledem k tomu, že předmětem práce je zkoumat interakce mezi vozidly a tyto interakce má význam sledovat pouze v oblasti ovlivněného dopravního proudu, je výhodné volný dopravní proud z dat vyseparovat a významně tak snížit výpočetní náročnost úlohy. Bohužel není zatím zřejmé, kde přesně hranice mezi volným a ovlivněným dopravním proudem leží. Existuje mnoho názorů. Například Kerner ve své práci [5] předkládá rozdělení dopravního proudu na volný a kondenzovaný, ten dále na synchronizovaný dopravní proud a rozprostřenou pohybující se kolonu. Bohužel při použití profilového měření není jednoznačná separace možná.

Možné řešení je využití fundamentální hypotézy třířákové teorie dopravy [4], která se pokouší jednotlivé fáze dopravního proudu označit v dvoudimenzionálním prostoru fundamentálního diagramu. Na obr. 11 je podle této teorie čárkovaně naznačena oblast volného dopravního proudu, tečkovaně potom oblast synchronizovaného dopravního proudu. Přesnou definici, podle které by se dal fundamentální diagram rozdělit, Helbing (ani ostatní) nepředkládá. To je však zásadní problém, chceme-li proces separace dopravního proudu opakovat u různých měření a výsledky srovnávat. Je tedy nutné najít metodu, která možná nebude zcela přesná z pohledu teorie dopravního proudu, avšak poskytne matematicky přesný a replikovatelný proces.

Jako vhodný nástroj se jeví metoda K-means (kap. 4.1). Bylo provedeno několik pokusů, na jaký parametr (nebo jejich kombinace) ji aplikovat (například klasický korelační koeficient nebo Brownovský korelační koeficient [19]). Vzhledem k tomu, že však zatím neexistuje uznaná a ověřená metoda, jak dopravní proud dělit, není možné říci, který parametr je pro účely separace vhodnější. Aby se tedy do procesu separace nevnášely další nejasnosti, byla jako parametr pro metodu K-means zvolena rychlost. Výsledky metody jsou vidět na obr. 11 a v tab. 3. Datový soubor byl vždy rozdělen na 4 shluky, přičemž první dva (sestupně dle rychlosti) byly z datového souboru vyřazeny (na obr. 11 červený(2) a zelený(3) shluk). Takto byly datové soubory upraveny pro další postup.

Fundamentální diagram po aplikaci K-means, soubor dat 1, pruh 0



Obrázek 11: Naznačení aplikace fundamentální hypotézy třífázové teorie dopravy. Čárkovane volný, tečkovaně synchronizovaný dopravní proud, k jsou jednotlivé shluky získané metodou K-means (kap. 4.1).

Tabulka 3: Výsledek aplikace metody K-means na datové soubory (vždy dva nejrychlejší shluky byly z datového souboru odstraněny).

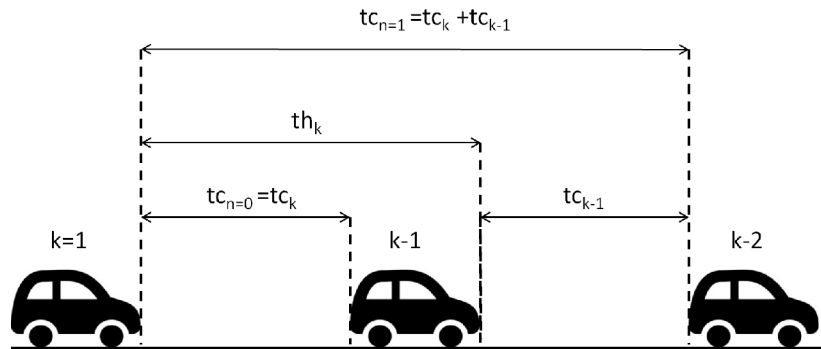
Soubor dat (s-soubor, p-pruh)	Celkem pozorování [mil.]	Pozorování po separaci [mil.]	Úbytek	Průměrná rychlost shluků [km/h]			
				1	2	3	4
s 1, p 0	2,93	1,80	39 %	117,0	109,3	99,3	45,0
s 1, p 1	3,45	2,53	27 %	100,2	93,1	86,6	42,3
s 2, p 0	3,17	1,61	49 %	124,6	114,8	102,3	62,3
s 2, p 1	3,21	1,69	47 %	106,7	98,1	90,8	56,6
s 3, p 0	2,89	1,80	38 %	120,1	110,9	98,7	41,3
s 3, p 1	3,49	1,70	51 %	102,9	94,4	86,4	38,5
s 4, p 0	0,27	0,14	51 %	118,1	109,0	98,1	48,7
s 4, p 1	0,34	0,16	54 %	101,0	92,8	85,9	73,2
s 5, p 0	2,70	2,25	17 %	120,1	108,7	93,7	51,2
s 5, p 1	3,68	1,83	50 %	102,3	91,8	80,3	46,3
s 6, p 0	3,27	1,64	50 %	123,8	113,0	99,2	60,6
s 6, p 1	3,12	1,39	55 %	106,2	96,6	88,8	55,9
s 7, p 0	3,07	1,25	59 %	121,4	111,4	97,6	45,0
s 7, p 1	3,29	0,97	70 %	103,8	94,7	87,1	43,2
s 8, p 0	2,98	2,41	19 %	119,7	111,7	102,2	44,3
s 8, p 1	3,40	0,93	73 %	103,1	95,1	88,4	41,7

5.3 Výpočet multi tc

V kapitole 3.2 jsou uvedeny výpočty charakteristik pro sousední vozidla. Tuto „vzdálenost“ označme jako $n = 0$ pro interakce vozidel k a $k - 1$ (čili $n + 2$ je počet vozidel a $n + 1$ počet mezer). Chceme-li však pozorovat interakce vzdálenější, například mezi vozidly k a $k - 2$, kterou označíme jako $n = 1$ (viz obr. 12), musíme zvolit vhodnou veličinu. V duchu kapitoly 4.3 budeme sledovat čistý časový odstup vozidel, normovaný na střední hodnotu 1. Jak je však

zřejmě z rovnice (7), nezahrnuje čistý časový odstup samotné vozidlo. Naše výpočty se tak posouvají do vyšší roviny abstrakce, kdy sledujeme dopravní proud prostřednictvím posloupnosti tc a přítomnost vozidel jako takových prakticky ignorujeme. Tímto krokem naplníme předpoklady uvedené v kapitole 4.3 a na dopravní proud se budeme dívat jako na termodynamický plyn, přičemž vlastnosti dopravního proudu budeme popisovat hustotou pravděpodobnosti tc_n , kdy

$$tc_n = \sum_{k=1}^{n+2} tc_k. \quad (23)$$



Obrázek 12: Princip výpočtu multi tc .

5.4 Dvouparametrická estimační metoda

Parametry α a β funkce (16) jsou vždy hledány pro úzké hustotní pásmo (v našem případě je šířka pásma 5 voz/km) a estimační proces tedy pracuje pouze s m -ticemi, které do daného hustotního pásma spadají. Cílem je najít minimum funkce (19) s přesností určení α a β na dvě desetinná místa. Jako prostředí pro implementaci iterační metody byl zvolen jazyk R [3] s nadstavbou Rstudio [1], který je pro podobné účely vhodný především díky své jednoduchosti a implementaci matematických a statistických nástrojů.

První iterační algoritmus je primitivní a spoléhá na hrubou výpočetní sílu. Pro parametry α a β v rozmezí $\langle 0,20 \rangle$ jsou spočítány všechny hodnoty χ (19) a ve výsledné matici je hledána minimální hodnota χ s odpovídajícími hodnotami α a β . Tím je stanoven optimální fit pro dané hustotní pásmo. Přesto že je metoda časově náročná (matice má 4 mil. buněk), její hlavní výhoda spočívá v robustnosti, jelikož neznáme tvar plochy matice χ . Kdyby totiž tato plocha nebyla monotónní a obsahovala lokální maxima a minima, mohly by optimalizované metody hledání minima χ vést k mylným výsledkům. Výsledky získané tímto algoritmem však potvrdily monotónnost matice (obr. 13 nahoře). Víme-li, že je matice monotónní, můžeme použít optimalizované metody. Vzhledem k charakteru úlohy a již vytvořenému kódu byla zvolena velice jednoduchá, ale efektivní metoda dílčí matice. Postup algoritmu je následující:

1. Z definičního oboru α a β je vytvořena dílčí matice 20×20 s počátkem v bodě $\alpha = \beta \cong 0$ (z praktického hlediska výpočtu nejsou voleny nulové hodnoty).
2. Pro dílčí matici je proveden výpočet χ .

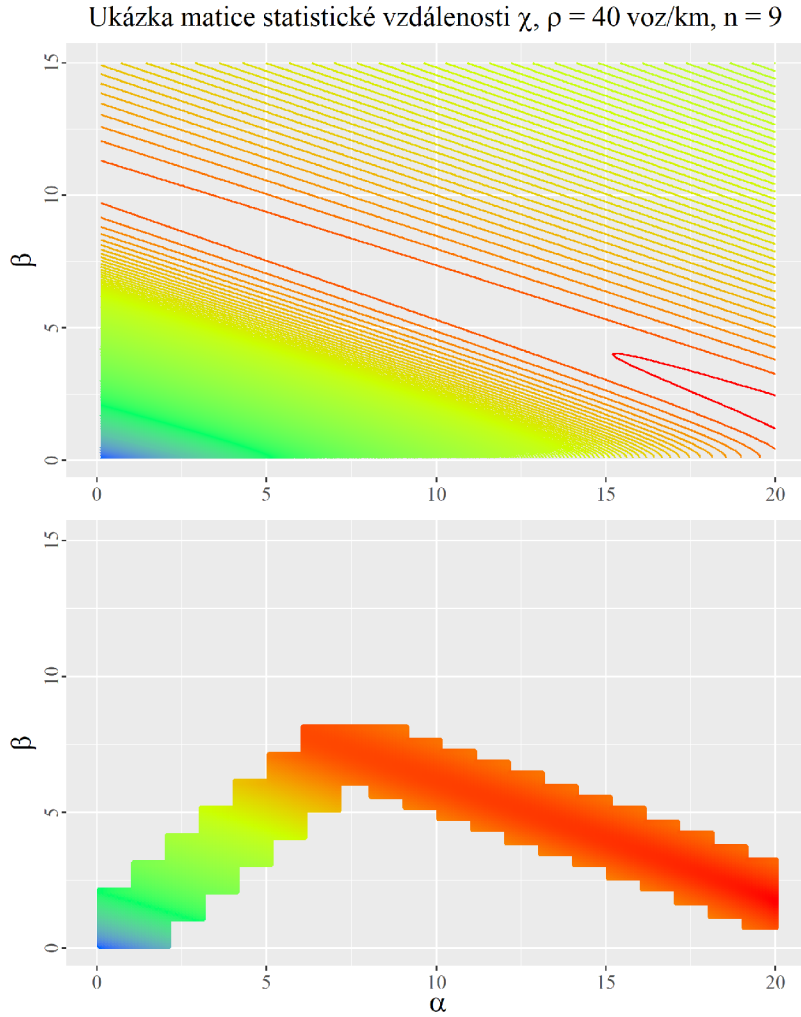
3. Je nalezeno minimum $\chi[\alpha, \beta]$, tato buňka je novým počátkem pro další dílčí matici 20×20 (z praktického hlediska se matice překrývají) a výpočet se opakuje.

Výsledek tohoto procesu je vidět na obr. 13 dole. V tomto případě je ještě vidět, že minimum matice je omezené definičním oborem α a β . Algoritmus byl tedy doplněn o funkci, která definiční obor α a β rozšiřuje do doby, než je minimum matice nalezeno. Toto minimum často leží na kraji matice, kdy se α nebo $\beta = 0$ (v případě z obr. 13 se po rozšíření matice iterace zastaví na hodnotách $\alpha = 23,2$; $\beta \cong 0$; $\chi = 1,10e^{-06}$).

Po zpracování všech datových souborů se ukázalo, že fitovací procedura vykazuje tyto pravidelnosti:

- Minima matice χ leží prakticky na přímce (v závislosti na přesnosti výpočtu) $\alpha = \beta$ až do kritického bodu (na obr. 13 $\alpha \cong \beta \cong 7,5$), ve kterém převládne jeden z parametrů a minimum matice „vybočí“ směrem k jedné z os.
- Matice vždy obsahuje pruh hodnot s velice malými rozdíly, který je kolmý k diagonále a leží v něm minimum matice (na obr. 13 červená oblast dat).

Tyto vlastnosti naznačují, že mezi parametry α a β existuje závislost, kterou je možné využít pro další optimalizaci iteračního algoritmu.



Obrázek 13: Znázornění matice χ pro datový soubor 1, rychlý pruh (barevné měřítko je logaritmické, $\chi_{max} = 4,10e^{-03}$, $\chi_{min} = 1,42e^{-06}$). Na horním obrázku je zobrazen průběh izolinií kompletní matice. Patrná je monotónnost plochy a její globální minimum (vpravo dole). V oblasti globálního minima jsou rozdíly χ již velice malé. Na spodním obrázku je pak zobrazen výsledek optimalizované metody pro stejná data, kdy jsou pro iteraci použity dílčí matice.

5.5 Dvouparametrická estimační metoda s využitím ekvivalentního rozptylu

Zjevná závislost mezi parametry α a β poukazuje na skutečnost, že empirická a teoretická hustota pravděpodobnosti musejí mít stejný (v rámci dané přesnosti odhadu α a β na dvě desetinná místa) rozptyl [8]. Známe-li tedy rozptyl empirické hustoty pravděpodobnosti $V_{h(s)}$, můžeme spočítat všechny kombinace α a β rozptylu teoretické hustoty pravděpodobnosti $V_{\varphi(s)}(\alpha, \beta)$, aby platil vztah

$$V_{h(s)} = V_{\varphi(s)}(\alpha, \beta) (\alpha \geq 0) (\beta \geq 0). \quad (24)$$

Hodnoty α a β pro ekvivalentní rozptyly jsou stanoveny pouze pro $n = 0$. Pro $n > 0$ je nutné fitovací algoritmus následovně upravit:

1. Mějme data pro $n = 9$. Tyto data je nutné přeškálovat na $n = 0$, což provedeme podle vzorce

$$tc_0 = \frac{tc_n}{n+1}. \quad (25)$$

2. Následně musíme rovněž upravit rozptyl

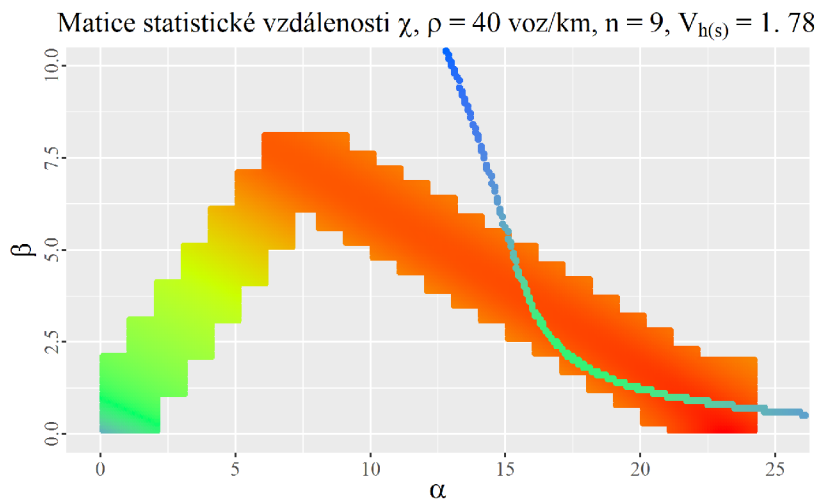
$$V_{h(s)}^{n=0} = \frac{V_{h(s)}^n}{(n+1)^2}. \quad (26)$$

3. Odhadneme hodnoty α a β pro $n = 0$ a splňující podmínku (24).

4. Převedeme hodnoty α a β pro $n = 9$

$$\begin{aligned} \alpha_n &= \alpha_0, \\ \beta_n &= \beta_0 \cdot (n+1). \end{aligned} \quad (27)$$

Podmínka (24) nám v matici χ definuje odpovídající dvojce α a β , pro které má smysl iteraci provádět. Na obr. 14 jsou tyto dvojce znázorněny řadou bodů a srovnány s výsledkem získaným optimalizovanou metodou z kap. 5.4. Z barevného měřítka je zřejmé, že hodnoty χ získané s využitím ekvivalentního rozptylu ($\chi_{min} = 1,03e^{-03}$, v obrázku znázorněné řadou bodů) jsou o několik řádů vyšší než při použití optimalizované metody ($\chi_{min} = 1,10e^{-06}$, v obrázku znázorněné plochou) a rovněž se liší odhady parametru $\alpha = 17,2$, $\beta = 2,2$ respektive $\alpha = 23,2$, $\beta \cong 0$. To je zapříčiněno tím, že výběr možných α a β je podmínkou (24) značně omezen a není proto dosaženo tak přesného fitu jako v případě optimalizované metody, která však pracuje s nereálnými hodnotami α a β , a že pro fitování jsou data předvedena na $n = 0$ což vede ke změně tvaru plochy χ .



Obrázek 14: Znázornění matice χ pro datový soubor 1, rychlý pruh (barevné měřítka je logaritmické), $\chi_{max} = 2,16e^{-02}$, $\chi_{min} = 1,10e^{-06}$). Plocha znázorňuje výsledek optimalizované iterační metody z kap. 5.4. Řada bodů představuje kombinace α a β , pro které je splněna podmínka ekvivalentního rozptylu (24).

5.6 Definiční obor estimace

Původním záměrem bylo provést fitovací proceduru zvlášť pro každý datový soubor a každý jízdní pruh. S postupem práce se však ukázalo, že i tak velký soubor dat, jaký byl k dispozici, má svá omezení. Jak je uvedeno v předchozích kapitolách, fitování parametrů α a β je prováděno pro úzká hustotní pásma, přičemž hustota je definována pro jednotlivé m -tice. Tyto m -tice jsou tedy sdruženy do dílčích souborů dat se stejnou hustotou (respektive s hustotou ve stejném intervalu). Pro okrajová hustotní pásma se počet měření pro jednotlivá hustotní pásma rychle snižuje až na úroveň desítek měření, což je pro věrohodnou fitovací proceduru nedostatečné. Rovněž je velká difference mezi levým (0) a pravým (1) jízdním pruhem, kdy v levém jízdním pruhu pozorujeme větší rozsah hustot (5–80 voz/km) než v pruhu pravém (5–65 voz/km) a výrazně se liší distribuce měření mezi hustotní pásma. Pro rozmezí 10–20 voz/km je v levém pruhu cca 155 tis. měření a v pravém 620 tis., přičemž celkový počet měření je 1,8 mil respektive 2,53 mil (tab. 3).

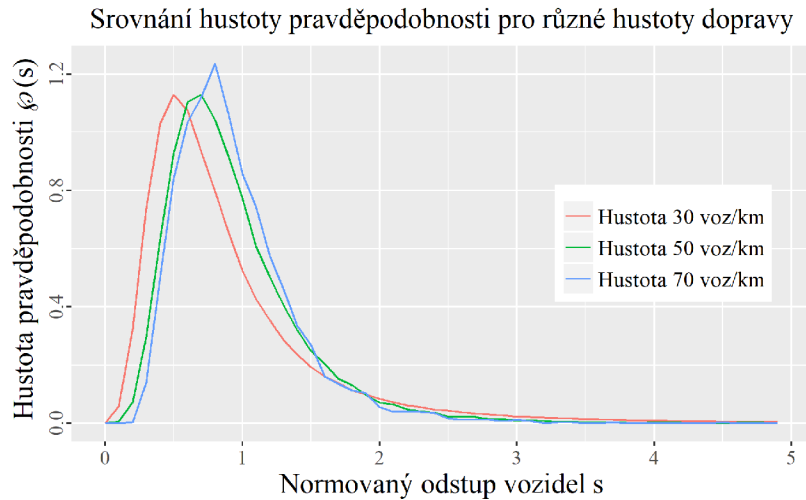
Na další úskalí v množství dat narazíme v situaci, kdy chceme na data uplatnit Kolmogorovův - Smirnovův test (kap. 4.6), jehož kritická hodnota vychází z počtu měření a je mu nepřímě úměrná. Pokud tedy bude j příliš velké, nulová hypotéza bude vždy zamítnuta. Byla tedy zvolena hodnota $j = 5000$, při které je zachována rozumná hodnota D_{krit} ale zároveň je zachována statistická smysluplnost testu. Rozdělení datových souborů na další dílčí celky (po 5000 měřeních) by však v praxi vedlo k tomu, že bychom mohli pracovat pouze s velice omezeným rozsahem hustot, ve kterých jsme schopni získat alespoň jeden takovýto blok dat (bloky musely být stejně velké z důvodu požadavku stejné hodnoty D_{krit}).

Pro další práce bylo tedy nutné datové soubory sloučit. Rovněž bylo rozhodnuto, že konečná analýza bude provedena pouze na levém jízdním pruhu, ve kterém je k dispozici mnohem větší rozsah hustot a dopravní proud je více homogenní (viz tab. 1). Dále bylo rozhodnuto, že pro další vyhodnocení budou použity hustoty v rozsahu „pouze“ 25–65 voz/km (tab. 5). Spodní hranice je dána předpokladem, že při nižších hustotách se už dostáváme do oblasti volného dopravního proudu. Horní hranice je dána počtem měření, které při hustotě 65 voz/km a sloučení všech datových souborů vystačí pouze na jeden soubor 5000 měření.

5.7 Výsledek estimace

V prvním kroku byly hodnoty α a β iterovány dle původního záměru pro všechny datové soubory, oba jízdní pruhy a $n = \{1, 2, \dots, 9\}$. Obdobně jako u ostatních charakteristik byly i zjištěné hodnoty α a β pro jednotlivé datové soubory velice podobné (viz tab. 4). Jako příklad jsou uvedeny hodnoty z datového souboru 1. Na obr. 16 jsou výsledné hodnoty α a β pro různé hustoty a n (pro sousední vozidla $n = 0$) znázorněny graficky. Ve spodní části obrázku je dobře patrné, že β má rostoucí tendenci směrem k vyšším hustotám a vyšším n . To je dáno tím, že parametr β řídí špičatost funkce (16).

Stejným způsobem se chová i empirická hustota pravděpodobnosti časových odstupů tc (obr. 15), jelikož při vyšších hustotách dopravního proudu se zmenšují rozdíly mezi odstupy jednotlivých vozidel (přibližujeme se k saturovanému dopravnímu proudu). U hodnoty α celoplošný trend nepozorujeme, ale jistou pravidelnost lze nalézt u extrémních případů hustot. To je zapříčiněno pravděpodobně tím, že parametr α řídí tvar funkce (16) zejména v exponenciální oblasti a při minimálních hustotách dopravního proudu získáváme vlivem extrémně velkých odstupů tc dlouhý exponenciální chvost hustoty pravděpodobnosti s ostrým zlomem. V oblasti maximálních hustot jsou vysoké hodnoty α zapříčiněny spíše nestabilitou výpočtu,



Obrázek 15: Empirické hustoty pravděpodobnosti pro různá hustotní pásma dopravního proudu. Je patrné, že s narůstající hustotou roste i špičatost funkce a maximum se posouvá směrem ke střední hodnotě 1 pro $n = 0$, což naznačuje zmenšování rozptylu hodnot tc .

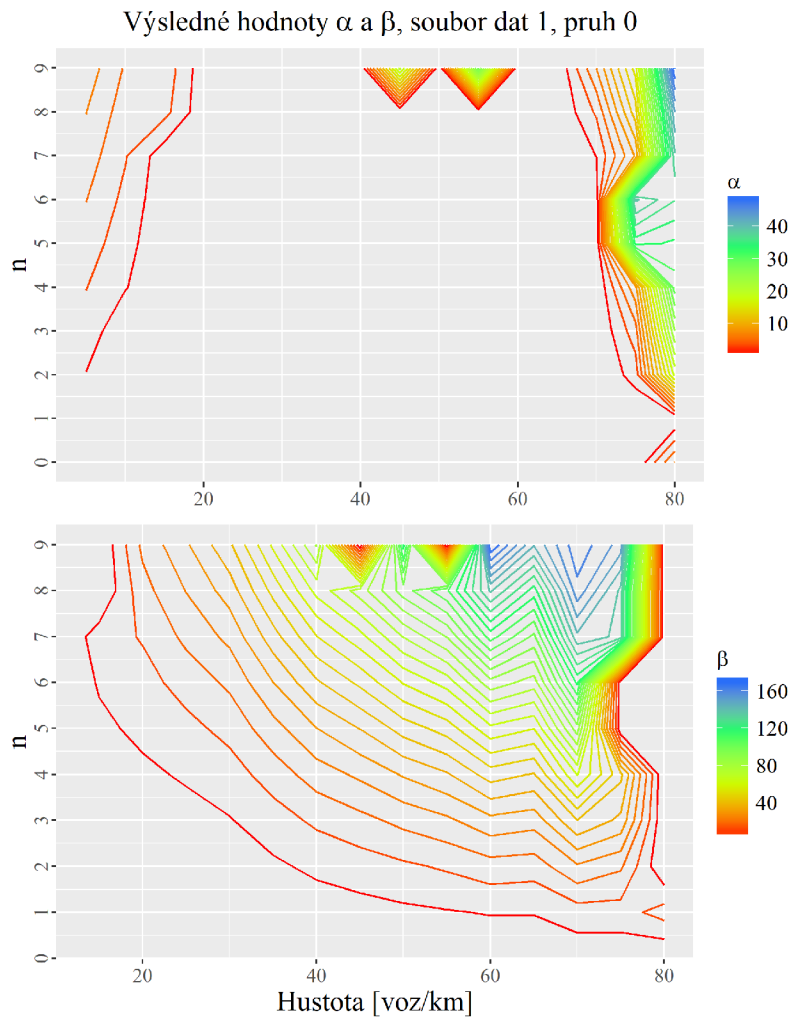
kteřá je způsobena malým počtem měření. Tato nestabilita se projeví rovněž skokovým nárůstem směrodatné odchylky odhadu α a β při vysokých hustotách v tabulce 4. Výsledky z prvního kroku nám poslouží jako kontrolní pro další postup, kdy jsme z důvodů uvedených v předchozí kapitole (5.6) sloučili všechny datové soubory do jednoho a pracovali v rozsahu hustot 25–65 voz/km.

Tabulka 4: Pruh 0, $n = 0$, datové soubory zpracované zvlášť.

[voz/km]	α		β	
	Průměr	Směrodatná odchylka	Průměr	Směrodatná odchylka
5	0,00	0,00	0,00	0,00
10	0,00	0,00	0,00	0,00
15	0,00	0,00	0,00	0,00
20	0,00	0,00	0,00	0,00
25	0,00	0,00	0,00	0,00
30	0,16	0,15	0,05	0,07
35	0,00	0,00	0,33	0,12
40	0,00	0,00	0,57	0,18
45	0,00	0,00	0,80	0,19
50	0,04	0,12	0,97	0,22
55	0,00	0,00	1,06	0,23
60	0,01	0,04	1,14	0,47
65	0,00	0,00	1,58	0,43
70	0,00	0,00	1,87	0,30
75	1,15	2,81	1,51	0,78
80	3,29	4,65	1,53	2,17

Sloučení datových souborů do jednoho velkého datového souboru bylo klíčovým krokem práce, který výrazně zjednodušil zpracování dat, ale hlavně umožnil provést analýzu inter-

akcí pomocí konvoluce a Kolmogorovova-Smirnovova testu. Jak je vidět v tabulce 5, ani po sloučení není dostupných dat příliš. Pro nejvyšší hustotu provozu 65 voz/km je k dispozici právě jedna sada 5000 měření. V tabulce 6 je vidět srovnání obou přístupů. Rozdíly jsou malé a jsou způsobené především vyhlazením empirických hustot pravděpodobnosti, protože při sloučení 8 datových souborů do jednoho se nám úměrně zvedl počet měření ve jednotlivých hustotních pásmech průměrně cca 8krát. Takto získané hodnoty parametrů α a β nám potvrdily stabilitu fitovacího procesu a zároveň nám posloužily jako výchozí hodnoty pro Kolmogorovův-Smirnovův test.



Obrázek 16: Srovnání globálního průběhu parametrů α a β (průběh je obdobný pro všechny datové soubory). U parametru β můžeme pozorovat jasný rostoucí trend směrem k vysokým hodnotám hustoty dopravního proudu a vzdálenosti n .

Tabulka 5: Pruh 0, datové soubory sloučené.

	[voz/km]	30	35	40	45	50	55	60	65
$n = 1$	α	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	β	0,23	0,60	0,85	1,05	1,28	1,47	1,54	2,02
$n = 2$	α	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	β	1,95	3,47	4,44	5,24	6,13	6,93	7,29	9,34
$n = 3$	α	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,69
	β	5,06	8,41	10,51	12,31	14,19	16,19	17,13	19,77
$n = 4$	α	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,01
	β	9,51	15,54	18,92	22,20	25,40	28,54	30,68	34,42
$n = 5$	α	0,00	0,00	0,00	0,00	0,00	0,32	0,45	1,72
	β	15,32	24,57	29,85	35,06	39,08	43,41	45,86	51,54
$n = 6$	α	0,24	0,17	0,00	0,00	0,14	0,78	0,50	1,70
	β	21,24	34,88	43,58	50,72	55,62	60,42	65,49	73,57
$n = 7$	α	0,47	0,31	0,00	0,00	0,56	1,02	1,25	1,84
	β	28,09	46,98	59,17	68,87	73,96	82,22	86,11	99,02
$n = 8$	α	0,75	0,55	0,00	0,00	0,75	28,80	1,16	2,40
	β	35,71	60,68	78,71	91,34	96,11	0,12	114,41	124,06
$n = 9$	α	0,94	0,66	0,00	0,00	1,06	32,25	1,48	1,83
	β	44,78	76,88	100,09	114,90	120,50	0,15	142,07	159,18
	Počet měření	435050	218900	122450	74150	49500	28200	14550	5650

Tabulka 6: Srovnání variant fitování, pruh 0, $n = 0$.

[voz/km]	α		β	
	Zvlášť (průměr)	Sloučené	Zvlášť (průměr)	Sloučené
30	0,16	0,00	0,05	0,23
35	0,00	0,00	0,33	0,60
40	0,00	0,00	0,57	0,85
45	0,00	0,00	0,80	1,05
50	0,04	0,00	0,97	1,28
55	0,00	0,00	1,06	1,47
60	0,01	0,00	1,14	1,54
65	0,00	0,00	1,58	2,02

6 Odhad interakční vzdálenosti vozidel

6.1 Aplikace principu konvoluce

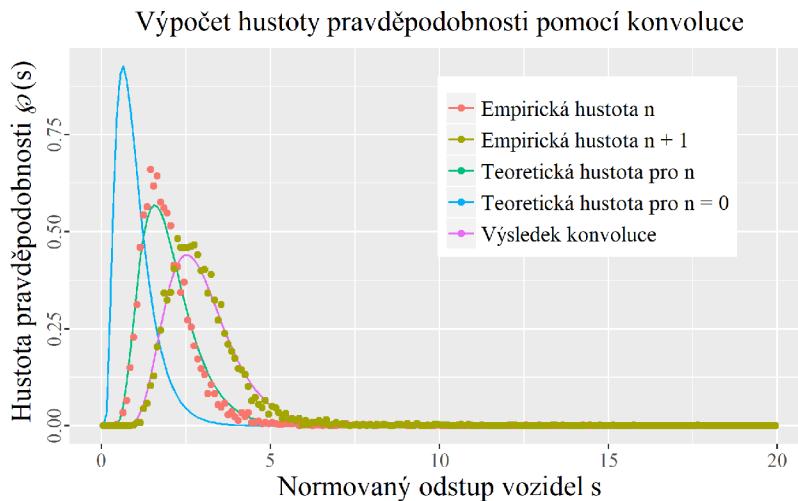
Stanovení interakční závislosti vozidel pomocí přímých metod je značně komplikované, protože často ani sám řidič si není vědom, kolik vozidel před sebou bere v potaz. Je tedy nutné hledat metody nepřímé. V předchozích kapitolách jsme detailně popsali proces, kterým jsme schopni poměrně přesně aproximovat hustotu pravděpodobnosti časových odstupů mezi vozidly tc_n pro sousední $n = 0$ a vzdálenější vozidla $n = \{1, 2, \dots\}$ (obr. 12). K aproximaci hustoty pravděpodobnosti jsme použili funkci (16) definovanou pomocí parametrů α a β , které jsme získali fitováním pro vybraná hustotní pásma a hodnoty n (viz tab. 5).

Na takto definované teoretické hustoty pravděpodobnosti můžeme použít princip konvoluce, který nám říká, že pokud jsou náhodné veličiny x a $x|n$ nekorelované, můžeme hustotu pravděpodobnosti $\varphi(s)(x|n+1)$ spočítat jako konvoluci hustot pravděpodobností těchto náhodných veličin

$$\varphi(x|n+1) = \varphi(x) * \varphi(x|n), \quad (28)$$

kde v našem případě je x teoretická hustota pravděpodobnosti dle funkce (16) pro $n = 0$ a $x|n$ je teoretická hustota pravděpodobnosti pro $n = \{0,1,\dots,9\}$. Výsledek tohoto procesu je vidět na obrázku 17, kde byla vypočtena konvoluce pro $n+1 = 2$.

Přijmeme tedy nulovou hypotézu, že interakce vozidel na vzdálenost $n+1 = 2$ (tedy 4 vozidla a 3 mezery, viz. kap. 5.3) je statisticky nevýznamná, pokud je statisticky nevýznamný i rozdíl mezi výsledkem konvoluce a empirickou hustotou pravděpodobnosti pro dané $n+1$, protože v tomto případě nejsou tyto dvě veličiny korelované a související vozidla mezi sebou nevyvíjí žádnou interakci. V opačném případě, kdy se výsledek konvoluce a příslušné empirické hustoty pravděpodobnosti významně liší, je nutné nulovou hypotézu zamítnout, což je důsledek interakce příslušných vozidel a vzniku korelačního vztahu mezi x a $x|n$ [8].



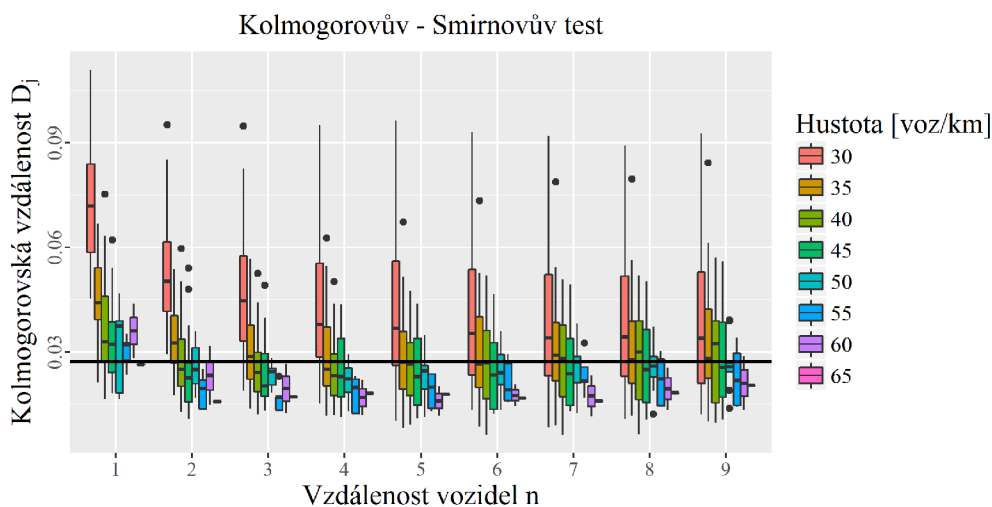
Obrázek 17: Ukázka výpočtu konvoluce, $n = 1$ a 40 [voz/km]

6.2 Detekce interakční vzdálenosti pomocí Kolmogorov-Smirnovova testu

Abychom mohli určit, zdali je nulová hypotéza z předchozí kapitoly splněna, musíme použít vhodný statistický test. Jako ideální se jeví Kolmogorovská vzdálenost (4.6) D_j , která pracuje s distribuční funkcí. To je pro nás výhodné, protože teoretická hustota pravděpodobnosti se může od empirické hustoty pravděpodobnosti lokálně významně lišit, avšak přepočtení obou hustot na distribuční funkci tyto lokální odchylky částečně eliminuje. Kolmogorův-Smirnovův test má rovněž větší sílu testu než, známější χ^2 . Jak bylo zmíněno v předchozích kapitolách, výsledek testu je závislý na počtu pozorování. Čím je počet pozorování větší, tím je test „přísnější“, jelikož se hodnota D_{krit} snižuje. Zároveň je zřejmé, že pokud se v hustotních pásmech bude lišit počet pozorování, což se nám pro různá hustotní pásma liší, budou se lišit i výsledky testu a nebude možné je vzájemně porovnat. Datový soubor byl tedy rozdělen

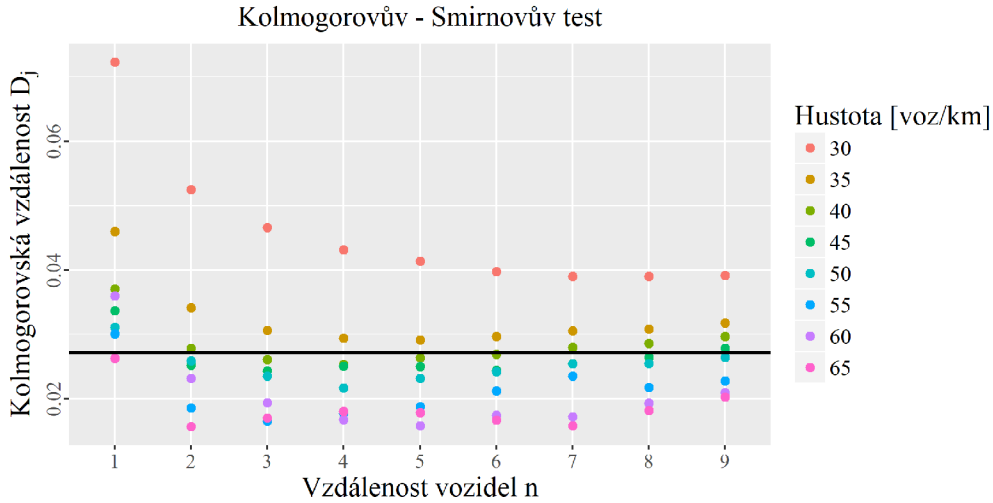
na bloky po 5000 měřeních, pro které $D_{krit} = 0,0272$ (na hladině významnosti 0,05). Díky tomuto kroku můžeme jednotlivé datové soubory mezi sebou srovnávat.

Samotné testování poté probíhá tak, že pro každý blok dat a každé n je vypočtena konvoluce, která je testována s empirickou hustotou pravděpodobnosti tohoto bloku dat, přičemž oboje je pro účel testu převedeno na distribuční funkci. Pro každý blok dat je tedy vyhotoven graf jako na obr. 17, empirická hustota $n + 1$ a výsledek konvoluce je převeden na distribuční funkci a tyto dvě distribuční funkce jsou podrobeny Kolmogorovu-Smirnovovu testu, čímž získáme hodnotu D_j pro daný blok dat. Pro každé hustotní pásmo tak získáme několik hodnot D_j , jak je vidět na obrázku 18, kde jsou hodnoty D_j pro jednotlivá hustotní pásma zobrazeny pomocí tzv. boxplotu. Pro poslední hustotní pásmo 65 voz/km máme k dispozici pouze jeden blok dat a je tedy zobrazena pouze střední hodnota.



Obrázek 18: Výsledek Kolmogorov-Smirnovova testu. Počet výsledků se pro jednotlivá hustotní pásma liší v závislosti na počtu měření. Černá čára představuje hodnotu G_{krit} , při jejímž překročení zamítáme nulovou hypotézu.

Poměrně velký rozptyl hodnot D_j v rámci jednoho hustotního pásma je dán postupem výpočtu, kdy empirické hustoty pravděpodobnosti pro jednotlivé 5000 bloky dat srovnáváme s teoretickými hustotami pravděpodobnosti dle tab. 5 kde jsou hodnoty α a β stanoveny pro hustotní pásmo jako celek. Bylo by sice možné hodnoty α a β určit znovu pro každý blok 5000 dat, což by však mělo negativní dopad na interpretovatelnost výsledků a ztratili bychom informaci o variabilitě dat a stabilitě testu. Abychom však mohli z výsledků testu učinit konkrétní závěr, musíme přijmout jisté zjednodušení. Jako reprezentativní hodnota pro každé hustotní pásmo byl tedy zvolen aritmetický průměr D_j .



Obrázek 19: Výsledek Kolmogorov-Smirnovova testu - průměr hodnot D_j . Černá čára představuje hodnotu D_{krit} , při jejímž překročení zamítáme nulovou hypotézu.

Při pohledu na obrázek 19 můžeme vidět, pro které hodnoty n a hustoty dopravního proudu byla nulová hypotéza zamítnuta (nad černou čarou, není možné vyloučit interakce mezi vozidly) a pro které hodnoty nebyla zamítnuta (pod černou čarou, interakce mezi vozidly je možné vyloučit). Výsledky naplnily očekávání a potvrdily závislost intenzity interakcí mezi vozidly na hustotě dopravního proudu a vzdálenosti vozidel, kdy je jasně patrný trend klesající míry interakce pro vyšší hustoty a vzdálenosti. Pod kritickou hodnotu se nedostanou pouze hustotní pásma 30 a 35 voz/hod, což samozřejmě není reálné, nicméně při pohledu na obr. 18 je vidět, že některé bloky dat těchto hustotních pásem pod kritickou hodnotu klesly. Pro vyšší hustoty se již výsledek chová dle očekávání a rostoucí tendence D_j pro vysoká $n > 6$ lze přičíst nestabilitě výpočtu v těchto oblastech, a to téměř ve všech krocích výpočtu, odhadem α a β počínaje.

Klíčové jsou však hodnoty pro $n = 1$ (tedy interakce přes dvě mezery, neboli mezi k -tým a $k - 2$ vozidlem na obr. 12), kde pro hustotní pásma 30-60 voz/hod nulovou hypotézu zamítáme a hustotní pásmo 65 voz/hod je hraniční. To potvrzuje naši úvodní hypotézu, že interakce neprobíhá pouze mezi sousedními vozidly, ale sahá nejméně o jedno vozidlo dále.

7 Závěr

Podarilo se prokázat, že předpoklady použité v současné teorii dopravního proudu a v tzv. car-following modelech nejsou přesné. Chování řidiče není ovlivněno pouze nejbližším předcházejícím vozidlem, ale je ovlivněno nejméně dvěma předcházejícími vozidly (v rozsahu vyšetřovaných hustot dopravního proudu), v některých případech však i čtyřmi až pěti vozidly (u nižších hustot). Tento výsledek podporují i další metody uvedené v [8] a rovněž je toto chování povědomé samotným řidičům, kteří přirozeně sledují provoz i dále, než jen po nejbližší před nimi jedoucí vozidlo. Hraniční nález v hustotním pásmu 65 voz/hod naznačuje, že toto zjištění nemusí být zcela platné v případech extrémně vysokých hustot, bohužel však nebylo možné tyto stavy dopravního proudu vyšetřit z důvodu nedostatku dat v této oblasti. Uve-

dená zjištění jsou pro oblast teorie dopravního proudu poměrně významná a v jejich výzkumu je nutno pokračovat.

Pokud přijmeme (respektive potvrdíme) předpoklad, že chování řidiče, respektive vozidla, není funkcí pouze $k - 1$ vozidla, ale je funkcí množiny vozidel $\{k - 1, k - 2, \dots, k - n\}$, musíme v první řadě stanovit míru, s jakou se jednotlivá vozidla na chování k -tého vozidla podílejí. Problematika je analogická se situací, kdy vyšetřujeme sílu, kterou působí na těleso jedna pružina a silou, kterou působí soustava pružin (je například nutné určit, jestli je toto působení paralelní, sériové, nebo kombinované). Pokud se podaří tyto závislosti nalézt, bude je možné implementovat do nového modelu dopravního proudu (car-following 2.0) a následně je promítnout například do normativů popisujících kapacitní výpočty nebo modelovacích softwarů. To by v konečném důsledku mělo vést ke zvýšení přesnosti dopravně inženýrských výpočtů a analýz, které jsou důležitým podkladem pro plánování rozvoje dopravní infrastruktury.

Reference

- [1] RStudio.
URL <https://www.rstudio.com/home/>
- [2] Dereziński Jan, G. J.: *Scattering Theory of Classical and Quantum N-Particle Systems*. Springer-Verlag Berlin Heidelberg, 1997.
- [3] Foundation, R.: The R Project for Statistical Computing.
URL <https://www.r-project.org/>
- [4] Helbing, D.: Traffic and related self-driven many-particle systems. *Rev. Mod. Phys.*, ročník 73, Dec 2001: s. 1067–1141, doi:10.1103/RevModPhys.73.1067.
URL <http://link.aps.org/doi/10.1103/RevModPhys.73.1067>
- [5] Kerner, B.: *The Physics of Traffic: Empirical Freeway Pattern Features, Engineering Applications, and Theory*. Understanding Complex Systems, Springer, první vydání, 2004, ISBN 978-3-540-20716-0.
- [6] Krbalek, M.: Inter-vehicle gap statistics on signal-controlled crossroads. *Journal of Physics A: Mathematical and Theoretical*, ročník 41, č. 20, 2008: str. 205004.
- [7] Krbalek, M.: Theoretical predictions for vehicular headways and their clusters. *Journal of Physics A: Mathematical and Theoretical*, ročník 46, č. 44, 2013: str. 445101.
- [8] Krbalek, M.; Apeltauer, J.; Apeltauer, T.; aj.: Three methods for estimating a range of vehicular interactions. *Physica A: Statistical Mechanics and its Applications*, ročník 491, č. Supplement C, 2018: s. 112 – 126, ISSN 0378-4371, doi: <https://doi.org/10.1016/j.physa.2017.09.008>.
URL <http://www.sciencedirect.com/science/article/pii/S0378437117309019>
- [9] Krbalek, M.; Helbing, D.: Determination of interaction potentials in freeway traffic from steady-state statistics. *Physica A: Statistical Mechanics and its Applications*, ročník 333, č. 1-4, 2004: s. 370–378, cited By 32.
- [10] Krbalek, M.; Hobza, T.: Inner structure of vehicular ensembles and random matrix theory. *Physics Letters A*, ročník 380, č. 21, 2016: s. 1839 – 1847, ISSN 0375-9601.
- [11] Krbalek, M.; Jiri Sleis: Vehicular headways on signalized intersections: theory, models, and reality. *Journal of Physics A: Mathematical and Theoretical*, ročník 48, č. 1, 2015: str. 015101.
- [12] Krbalek, M.; Seba, P.: Spectral rigidity of vehicular streams (random matrix theory approach). *Journal of Physics A: Mathematical and Theoretical*, ročník 42, č. 34, 2009: str. 345001.
- [13] Lloyd, S.: Least squares quantization in PCM. *IEEE Transactions on Information Theory*, ročník 28, č. 2, Mar 1982: s. 129–137, ISSN 0018-9448, doi: 10.1109/TIT.1982.1056489.

- [14] MacQueen, J.: Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, Berkeley, Calif.: University of California Press, 1967, s. 281–297.
URL <http://projecteuclid.org/euclid.bsmsp/1200512992>
- [15] Medelská, D. V.: *Dopravné Inžinierstvo*. Alfa, 1991.
- [16] Pipes, L. A.: An Operational Analysis of Traffic Dynamics. *Journal of Applied Physics*, ročník 24, 1953: str. 274.
- [17] Prof. Ing. Pavel Příbyl, C.: *Řídicí systémy silniční dopravy*. Fakulta dopravní (FD), 2003.
- [18] Steinhaus, H.: Sur la division des corps matériels en parties. *Bulletin de l'Académie Polonaise des Sciences*, ročník Cl. III — Vol. IV, č. 12, 1956: s. 801–804.
- [19] Szabová, Z.: *Correlation analysis in ensembles of empirical vehicular data*. České vysoké učení technické v Praze, 2016.
- [20] Wikipedia: k-means clustering.
URL https://en.wikipedia.org/wiki/K-means_clustering
- [21] Wikipedia: Konvoluce.
URL <https://cs.wikipedia.org/wiki/Konvoluce>