



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA STAVEBNÍ
FACULTY OF CIVIL ENGINEERING

ÚSTAV POZEMNÍCH KOMUNIKACÍ
INSTITUTE OF ROAD STRUCTURES

STATISTICKÉ VLASTNOSTI MIKROSTRUKTURY DOPRAVNÍHO PROUDU

STATISTICAL CHARACTERISTICS OF THE TRAFFIC FLOW MICROSTRUCTURE

DISERTAČNÍ PRÁCE
DOCTORAL THESIS

AUTOR PRÁCE
AUTHOR

Ing. Jiří Apeltauer

VEDOUCÍ PRÁCE
SUPERVISOR

doc. Ing. PETR HOLCNER, Ph.D.

BRNO 2018

ABSTRAKT

Současná teorie dopravního proudu předpokládá, že zde interagují pouze sousední vozidla. Toto východisko je legitimní, avšak vychází z možností vědy a techniky 50. let minulého století, které jsou v současnosti již překonány. Dnes je již zřejmé, že obecně probíhá interakce mezi vozidly na vzdálenosti větší (neboli mezi více vozidly), avšak zatím nebyl předložen postup, který by dokázal vzdálenost této interakce kvantifikovat. V této práci je předložena metoda, která s využitím nástrojů matematické statistiky a přesného měření časových odstupů jednotlivých vozidel umožňuje tyto interakční vzdálenosti mezi více vozidly určit. Současně byla ověřena její platnost pro úzká hustotní pásma dopravního proudu. Podařilo se prokázat, že při vysokých hustotách dopravního proudu probíhá interakce minimálně mezi třemi po sobě jedoucími vozidly, u nižších hustot mezi čtyřmi až pěti vozidly. Výsledky práce mají dopad na vývoj nových mikroskopických modelů dopravního proudu a jejich verifikaci.

KLÍČOVÁ SLOVA

Dopravní proud, Model následného vozidla, Interakce, K-means, Hustota pravděpodobnosti, GIG, Konvoluce, Kolmogorov–Smirnov test

ABSTRACT

The actual traffic flow theory assumes interactions only between neighbouring vehicles within the traffic. This assumption is reasonable, but it is based on the possibilities of science and technology available decades ago, which are currently overcome. Obviously, in general, there is an interaction between vehicles at greater distances (or between multiple vehicles), but at the time, no procedure has been put forward to quantify the distance of this interaction. This work introduces a method, which uses mathematical statistics and precise measurement of time distances of individual vehicles, which allows to determine these interacting distances (between several vehicles) and its validation for narrow densities of traffic flow. It has been revealed that at high traffic flow densities there is an interaction between at least three consecutive vehicles and four and five vehicles at lower densities. Results could be applied in the development of new traffic flow models and its verification.

KEYWORDS

Traffic flow, Car following model, Interaction, K-means, Probability density, GIG, Convolution, Kolmogorov–Smirnov test

BIBLIOGRAFICKÁ CITACE VŠKP

Ing. Jiří Apeltauer Statistické vlastnosti mikrostruktury dopravního proudu. Brno, 2018. 57 s., 27 s. příl. Disertační práce. Vysoké učení technické v Brně, Fakulta stavební, Ústav pozemních komunikací. Vedoucí práce doc. Ing. Petr Holcner, Ph.D.

PROHLÁŠENÍ

Prohlašuji, že jsem disertační práci zpracoval samostatně a že jsem uvedl všechny použité informační zdroje.

V Brně dne

Ing. Jiří Apeltauer
autor práce

Poděkování

Poděkování patří především doc. Mgr. Milanu Krbálkovi, Ph.D., který se podílel výrazně na formování tématu práce a jeho rady v oblasti matematiky a statistiky. Velký dík patří také mým blízkým, a to zejména mojí ženě, za poskytnutí klidného a pozitivního zázemí, které je pro tvůrčí práci nezbytné.

Statistika je jediná věda, která při použití stejných vzorců umožňuje různým expertům dostat různé výsledky.

Evan Esar

Obsah

1	Úvod	15
2	Teorie dopravního proudu	16
2.1	Úvod do teorie dopravního proudu	16
2.2	Makroskopické modely dopravního proudu	17
2.2.1	Greenshieldsův lineární model	18
2.2.2	Greenbergův logaritmický model	18
2.2.3	Underwoodův exponenciální model	19
2.2.4	Pipesův a Munjalův zobecněný model	20
2.3	Mikroskopické „car-following“ modely dopravního proudu	21
2.3.1	Model Luise A. Pipese	21
2.3.2	Model Gazis – Herman – Rothery (GHR)	21
2.3.3	Gippsův model	22
3	Data	23
3.1	Měření	23
3.2	Makro a mikrocharakteristiky datových souborů	25
4	Použité statistické nástroje	29
4.1	Shluková analýza K-means	29
4.2	Hustota pravděpodobnosti	30
4.3	Inverzní Gaussova hustota pravděpodobnosti (GIG)	32
4.4	Statistická vzdálenost χ	34
4.5	Konvoluce	34
4.6	Kolmogorovův-Smirnovův test	35
5	Fitovací procedura	37
5.1	Příprava dat	37
5.2	Separace ovlivněného dopravního proudu	38
5.3	Výpočet multi tc	39
5.4	Dvouparametrická estimační metoda	40
5.5	Dvouparametrická estimační metoda s využitím ekvivalentního rozptylu	42
5.6	Definiční obor estimace	44
5.7	Výsledek estimace	44
6	Odhad interakční vzdálenosti vozidel	51
6.1	Aplikace principu konvoluce	51
6.2	Detekce interakční vzdálenosti pomocí Kolmogorov-Smirnovova testu	52
7	Závěr	54
	Reference	55
	Seznam obrázků	57

Seznam tabulek	59
8 Přílohy	60
8.1 R code	60
8.1.1 Úvod	60
8.1.2 Načtení a příprava dat	60
8.1.3 Generování <i>m-tic</i>	61
8.1.4 Separace volného dopravního proudu metodou K-means	61
8.1.5 Výpočet multi <i>tc</i>	62
8.1.6 Vytvoření podsouboru <i>tc</i> a rozdělení na hustotní pásma	63
8.1.7 Výpočet hustoty pravděpodobnosti	63
8.1.8 Fitování	63
8.1.9 Vykreslení grafu	65
8.1.10 Estimační procedura	66
8.1.11 Konvoluce	67
8.1.12 Kolmogorův-Smirnovův test	68
8.1.13 Postprocesing dat	72
8.2 Intenzity a rychlosti	74
8.3 Fundamentální diagramy	82

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. Přijetí 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.

Práce je dělena celkem do 5 celků. V první části se seznámíme s teorií dopravního proudu a modely dopravního proudu tak, jak se vyvíjely v čase. Z pohledu řešení problematiky je klíčová především kapitola 2.3, ve které jsou uvedeny příklady modelů následného vozidla. V další kapitole jsou popsány datové soubory, které byly pro práci využity, včetně základních makro a mikro charakteristik zkoumaného dopravního proudu. V kapitole 4 se čtenář seznámí s nástroji matematické statistiky, které byly pro vyšetřování závislostí mezi jednotlivými vozidly použity. Jednotlivé nástroje jsou seřazeny tak, jak byly na datové soubory postupně aplikovány a podrobný postup jejich použití je popsán v kapitole 5. Klíčová hypotéza této práce (tedy existence širších interakcí mezi vozidly) je pak testována v kapitole 6.

V rámci práce se pod pojmem „Ovlivněný dopravní proud“ rozumí fáze dopravního proudu, kdy řidiči začínají svoji rychlost přizpůsobovat vozidlům před sebou a to do takové míry, že je tuto interakci možné statisticky sledovat. Zpravidla se jedná o situaci, kdy je vzdálenost vozidel v řádech desítek metrů nebo menší a rychlosti jsou výrazně pod průměrem volného dopravního proudu. Pod pojmem „Volný dopravní proud“ se rozumí fáze dopravního proudu, kdy není možné statisticky sledovat interakce mezi vozidly, která se pohybují na hranici nebo za hranicí dohledu a rychlosti se blíží rychlostem dovoleným (nebo možným) na daném úseku.

2 Teorie dopravního proudu

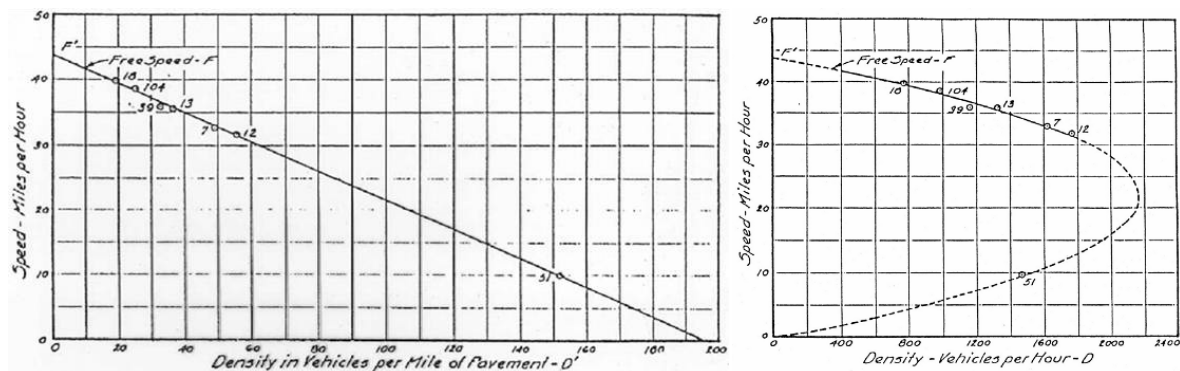
2.1 Úvod do teorie dopravního proudu

Z historie vyplývá, že čím byla daná civilizace mobilnější, tím měla větší kulturní a vojenskou moc. Úspěch prvních civilizací byl tedy do značné míry ovlivněn přístupem ke koním a zvládnutím jejich chovu a šlechtění tak, aby co nejlépe zvládali specifické úlohy, ať už vojenské nebo civilní. Významným mezníkem byl vynález kola, který vedl k rozvoji cest a v podstatě definoval jejich podobu, jak ji známe v současnosti. Kůň má však svá omezení daná přírodou, a proto rozvoj silniční dopravy dlouhá staletí stagnoval a omezoval se pouze na zlepšování konstrukce vozovky, která byla zkoušena stále těžšími povozy. Tento stav trval až do 18. století. V té době už byly ulice velkých měst plné kočárů, avšak s hustotami dnešní dopravy se nemohly měřit ani zdaleka. K velké změně došlo až na přelomu 19. a 20. století. Automobil byl již obecně známý dopravní prostředek a po komunikacích se jich pohybovala celá řada. Nicméně stále šlo o luxusní zboží, které si běžní lidé nemohli dovolit. Tato situace však měla velice záhy skončit, protože v roce 1908 byl na trh uveden automobil Ford model T dostupný široké populaci. Automobilismus, jak jej známe v dnešní době, se v západní Evropě rozvinul masivně během 60.–70. let 20. století, ve střední Evropě poté v 80.–90. letech.

Netrvalo dlouho a o tento nový fenomén se začali zajímat zvědaví vědci. 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. Jde o lineární závislost rychlosti a hustoty, ze které vyplývá parabolická závislost mezi rychlostí a intenzitou (obr. 1).

Tento první model byl samozřejmě značně zjednodušený, avšak odhalil základní vlastnosti dopravního proudu. Ukázal, že existuje maximální intenzita při optimální hustotě dopravy. Tento bod rozděluje graf závislosti rychlosti na intenzitě (obr. 1 vpravo) na dvě oblasti, ve kterých totožná hodnota intenzity odpovídá vždy jiné rychlosti. Například pro hodnotu intenzity 1200 voz/h můžeme v Greenshieldsově grafu najít rychlost 7,5 mil/h a 36 mil/h. Nižší hodnota se nachází v nestabilní oblasti dopravního proudu (vysoká hustota), vyšší potom v oblasti stabilní (nízká hustota).

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



Obrázek 1: Původní Greenshieldsovy diagramy závislostí základních veličin dopravního proudu. Vlevo lineární závislost rychlosti a hustoty, v pravo parabolická závislost mezi rychlostí a intenzitou (v diagramu ještě označena jako “Density-vehicles per Hour”) [8].

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 [22]. Jeho model předpokládal, že vzdálenost mezi vozidly je rovna jedné délce vozidla na každých 10 mil/h plus minimální vzdálenost mezi vozidly, pokud stojí. Tento předpoklad měl velice prostý základ. Vycházel ze soudobé příručky pro řidiče, kde bylo uvedeno: „A good rule for following another vehicle at a safe distance is to allow yourself the length of a car (about 15 ft) for every 10 miles per hour you are travelling.“ (ve volném překladu: „Dobрым pravidlem je při jízdě za vozidlem dodržet bezpečnou vzdálenost v délce jednoho vozidla (asi 15 stop) na každých 10 mil/h vaší cestovní rychlosti.“). Zda-li byl tento předpoklad správný není z dnešního pohledu podstatné. Mnohem závažnějším problémům čelil Pipes při aplikaci svých rovnic do praxe. 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 numerického 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.

2.2 Makroskopické modely dopravního proudu

Jak už bylo řečeno v předchozí kapitole, makroskopický pohled na dopravní proud agreguje informace o jednotlivých vozidlech jak v čase, tak v prostoru. Na dopravní proud je tedy nahlíženo podobně jako na tok kapaliny a na dopravní síť jako na potrubí nebo koryto, kterým tato kapalina proudí. Dopravní proud je zpravidla definován intenzitou (průtokem) a rychlostí, jelikož tyto veličiny jsou přímo měřitelné. Hustota je ze své podstaty veličina, která je měřitelná jen obtížně, a proto se zpravidla nepoužívá. Sice je možné ji vypočítat z rychlosti

a intenzity, avšak za předpokladu, že rychlost vozidla se v čase mění, je tento výpočet vždy zatížen chybou.

Omezená přesnost makroskopických modelů je vyvážena jejich jednoduchostí, díky které mohly být používány při analytických (a často ručních) výpočtech a v počátcích výpočetní techniky s jejich pomocí bylo možné modelovat i rozsáhlé dopravní sítě v rozumném čase. Postupně jsou nahrazovány takzvanými mesoskopickými modely, které více či méně zohledňují i jednotlivá vozidla, ale parametry těchto vozidel jsou zobecněny.

2.2.1 Greenshieldsův lineární model

Nejstarší a zároveň nejjednodušší model byl již zmíněn v předchozí kapitole. Jeho lineární část je definována závislostí mezi rychlostí a hustotou

$$v = v_f \left(1 - \frac{\rho}{\rho_j}\right), \quad (1)$$

kde v je aktuální rychlost vozidla, v_f je rychlost na nezatížené komunikaci (volná nebo také požadovaná rychlost), ρ je aktuální hustota [voz/km], ρ_j je hustota při kongesci. Pokud do této rovnice dosadíme základní Greenshieldsův postulát, že intenzita (q) je rovna součinu rychlosti a hustoty $q = v\rho$, dostaneme druhou parabolickou část Greenshieldsova modelu, která popisuje vztah intenzity a rychlosti

$$q = v_f \left(\rho - \frac{\rho^2}{\rho_j}\right). \quad (2)$$

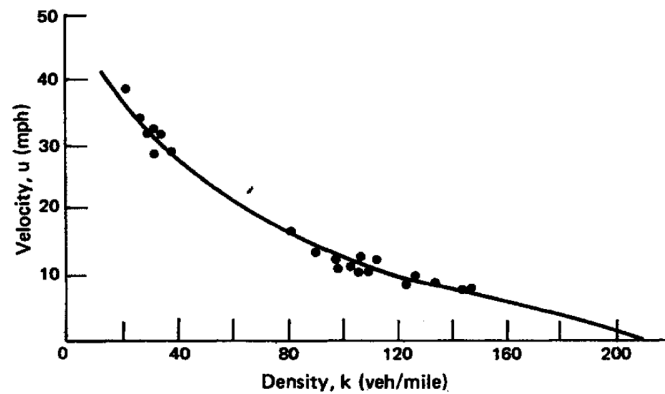
Grafy obou závislostí jsou vykresleny na obrázku 1. Pro svou jednoduchost byl model oblíben především v polovině 20. století. Bohužel v kombinaci se soudobými statistickými metodami a metodami dopravních průzkumů nedosahoval kýžené přesnosti. V té době totiž jako vstup do modelu sloužila hustota, která se dopočítala z přesně změřené rychlosti a intenzity. Dnes už však víme, že takto zjištěná hustota je zatížena chybou, která odpovídá změně rychlosti vozidla v čase a ta může být především v nestabilní fázi dopravního proudu nezanedbatelná.

2.2.2 Greenbergův logaritmický model

Dalším používaným modelem byl Greenbergův logaritmický model, který více odpovídal empirickým datům (obr. 2 a 4) a byl definován rovnicí

$$v = v_0 \cdot \ln \left(\frac{\rho_j}{\rho}\right). \quad (3)$$

I na tento model lze aplikovat základní Greenshieldsovy rovnice a odvodit tak další vztahy. Hlavní nevýhodou je matematický princip logaritmu, kdy pro hustoty blízké nule bude výpočet nestabilní, a obdobně jako Greenshieldsův model trpěl nepřesným výpočtem hustoty.



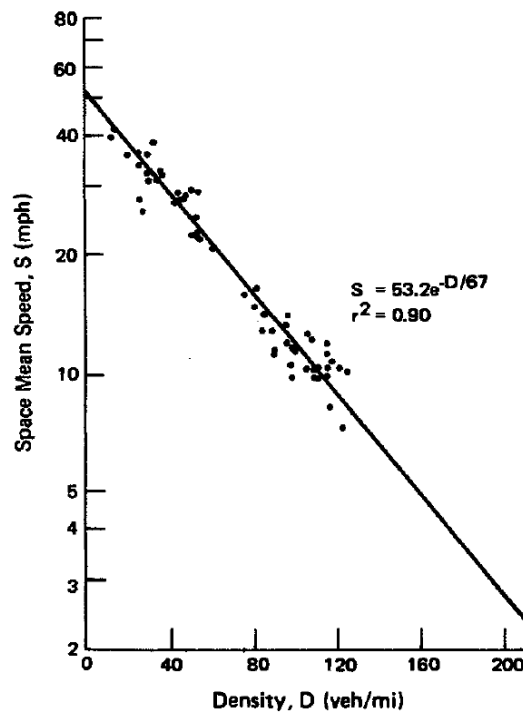
Obrázek 2: Srovnání Greenbergova logaritmického modelu s reálnými daty [2].

2.2.3 Underwoodův exponenciální model

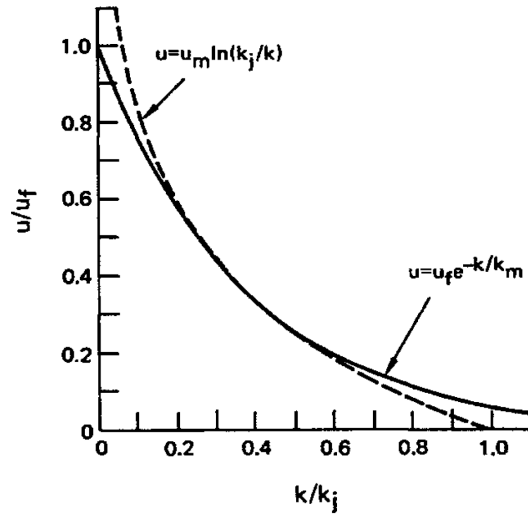
Problém logaritmického modelu s nestabilitou při nízkých hustotách se Underwood pokusil vyřešit exponenciálním modelem (obr. 3 a 4). Ten měl pro změnu zase problém při hustotách vysokých a nulová rychlost mohla nastat jenom při nekonečné hustotě. Pro běžné situace však fungoval dobře. Byl definován rovnicí

$$v = v_f \cdot e^{-\frac{\rho}{\rho_0}}, \quad (4)$$

kde ρ_0 je hustota provozu při maximální intenzitě.



Obrázek 3: Srovnání Underwoodova exponenciálního modelu s reálnými daty [2].

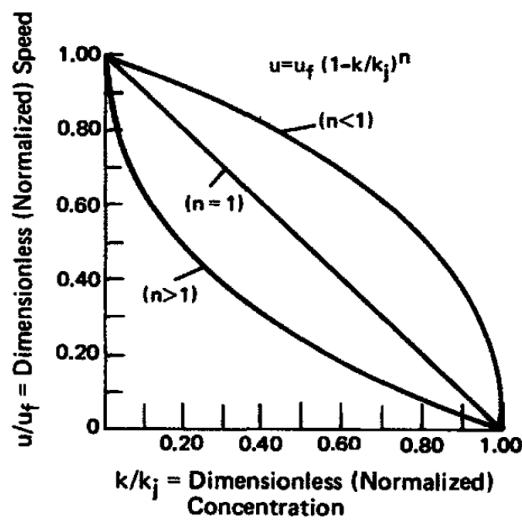


Obrázek 4: Teoretický průběh Greenbergova (čárkovaná čára) a Underwoodova (plná čára) modelu (v tehdejší době se pro označení hustoty ρ používal symbol k) [2] .

2.2.4 Pipesův a Munjalův zobecněný model

Další teoretické práce se ubíraly především zobecňováním a úpravou předešlých modelů. Například Pipesův a Munjalův zobecněný model, který pracuje s podílem aktuální rychlosti v a hustoty ρ k volné rychlosti v_f respektive hustotě při kongesci ρ_j a zavádí kalibrační konstantu $n = \{n \in R | n > 0\}$ ovlivňující tvar funkce, viz obr. 5 a rovnice

$$v = v_f \left(1 - \frac{\rho}{\rho_j}\right)^n . \quad (5)$$



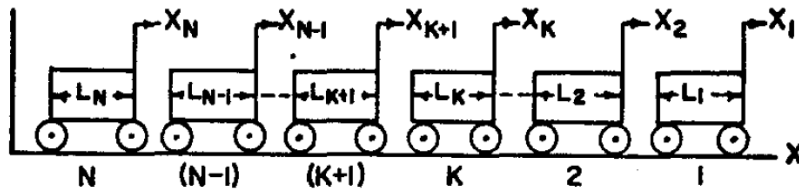
Obrázek 5: Pipesův a Munjalův zobecněný model (v tehdejší době se pro označení hustoty ρ používal symbol k) [2].

2.3 Mikroskopické „car-following“ modely dopravního proudu

Hlavním rozdílem mikroskopických modelů dopravního proudu oproti makroskopickým je zohlednění každého vozidla, jeho chování a interakce s ostatními vozidly. Jak už bylo zmíněno v úvodu této kapitoly, jedním z prvních byl model Luise A. Pipesa [22], který ovšem zároveň zjistil, že realizace těchto modelů analyticky není možná a musí být řešeny numericky pomocí počítačů. V roce 1952, kdy svoji práci publikoval, už sice byly používány tranzistorové počítače (tzv. druhá generace), ale byly dostupné jenom na těch nejbohatších univerzitách a pro špičkový výzkum. Rozvoj těchto modelů do praxe byl tedy do značné míry vázaný na dostupnost výpočetního výkonu, v teoretické rovině však vědci nezaháleli a v následujících letech po Pipesově publikaci vznikla celá řada dalších a přesnějších modelů. Jedno však tyto modely mají dodnes společné. Bez ohledu na to, jak komplexně je popsána interakce mezi vozidly, je tato interakce omezena pouze na vozidla sousední, v čase po sobě následující. Jsou tedy označovány jako „car-following“ modely. Praxe ukázala, že tento zjednodušující předpoklad je přijatelný, ale pokud se mají mikrosimulační modely dále vyvíjet, je potřeba tento předpoklad opustit a definovat širší interakce mezi více vozidly.

2.3.1 Model Luise A. Pipesa

Jeden z prvních mikrosimulačních modelů. Jak již bylo zmíněno, je postaven na předpokladu, že vzdálenost mezi vozidly je rovna jedné délce vozidla na každých 10 mil/h plus minimální vzdálenost mezi vozidly, pokud stojí. Pipes definoval řadu n po sobě jdoucích vozidel (obr. 6) a vztah mezi jejich rychlostí



Obrázek 6: Řada po sobě jdoucích vozidel jak ji definoval Luise A. Pipes [22].

$$a_{k+1} = \frac{1}{T}(v_k - v_{k+1}), k = 1, 2, 3 \dots (n-1), \quad (6)$$

kde a_k je akcelerace k -tého vozidla, v_k jeho rychlost a T časová konstanta (na základě předpokladu konstantního odstupu v závislosti na rychlosti stanovil Pipes $T = 1s$). Tím byla definovaná samotná podstata mikrosimulačních modelů. Pipesův model byl ale natolik zjednodušený, že nemohl podávat dobré výsledky. Velice záhy se ukázalo, že nelze předpokládat stejné chování u všech vozidel, toto chování je nutné variovat a například odstupy vozidel definovat statistickou funkcí, nikoliv konstantním parametrem. Rovněž se ukázalo jako nezbytné do rovnice zanést reakční dobu řidiče, která má na vzdálenost mezi vozidly zásadní vliv.

2.3.2 Model Gazis – Herman – Rothery (GHR)

Nedostatky Pipesova modelu se pokusil napravit model GHR, jehož základy byly položeny v roce 1958. Do výpočtu (7) zavedl reakční dobu řidiče a kalibrační parametry. I přesto, že byl tento model dlouze a podrobně studován, byl nakonec opuštěn. Kalibrační parametry bylo

potřeba získávat pro každý konkrétní model a neřídka tyto parametry nabývaly pro podobné situace protikladných hodnot, což nepůsobilo příliš důvěryhodně. Model je definován rovnicí

$$a_i(t) = c \cdot v_i^m(t) \frac{\delta v(t - T)}{\delta x^l(t - T)}, \quad (7)$$

kde $a_i(t)$ je zrychlení i -tého vozidla jako funkce času t , $v_i(t)$ je rychlost i -tého vozidla jako funkce času t , δv je relativní rychlost v porovnání s vedoucím vozidlem, δx je odstup od vedoucího vozidla, T je reakční doba řidiče a m, c, l jsou kalibrační konstanty.

2.3.3 Gippsův model

V následujících letech vznikala celá řada různě komplikovaných a úspěšných modelů. Jako další větve se rozvinuly i různé psycho-fyziologické modely (H. T. Fritsch [5], R. Wiedemann [26]) nebo modely založené na celuárních automatech (Nagel-Schreckenberg [21]). Některé z nich si našly cestu do praxe, jiné upadly v zapomnění. Cestou zachování deterministické podstaty matematického modelu se úspěšně vydal až P. G. Gipps v roce 1981 [6][7]. Z jeho modelu vychází většina poznatků v teorii dopravního proudu poslední doby a pravděpodobně se jedná o nejpřesnější model který je v současnosti užíván v inženýrské praxi. Je definován poměrně komplexní rovnicí (7) za předpokladu, že řidič je omezen maximální akcelerací a decelerací (zjednodušeně jsou tyto brány jako konstanty v celém rozsahu rychlostí), vzdáleností mezi vozidly a relativní rychlostí. Rychlost je nadále omezena tak, aby vozidlo vždy bezpečně zastavilo (takový model označujeme jako bezsrážkový), vozidla dále nepřekračují maximální rychlost a akcelerace je vždy maximální možná. Základní rovnice modelu je tedy

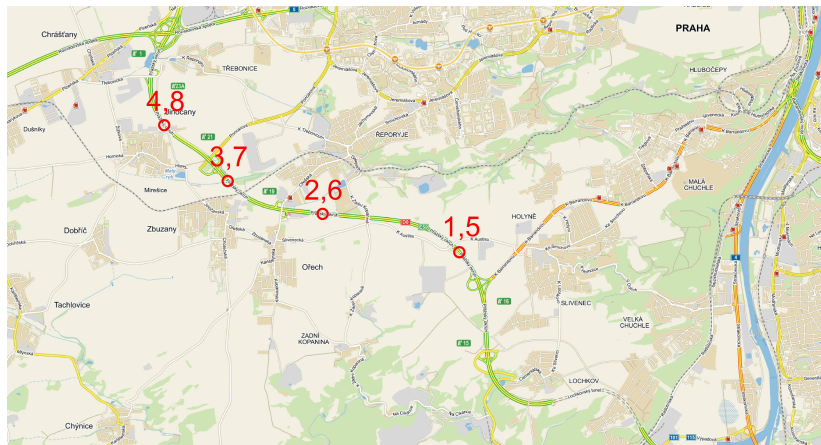
$$v_i(t + \tau) = \min\left\{v_i(t) + 2.5a_i\tau\left(1 - \frac{v_i(t)}{V_i}\right)\sqrt{0.025 + \frac{v_i(t)}{V_i}}, \right. \\ \left. b_i\tau + \sqrt{b_i^2\tau^2 - b_i\left[2\{x_{i-1}(t) - l_{i-1} - x_i(t)\} - v_i(t) - \frac{v_{i-1}(t)^2}{b}\right]}\right\}, \quad (8)$$

kde $v_i(t)$ je rychlost vozidla i v čase t , τ je reakční doba řidiče, respektive výpočetní krok, a_i je maximální akcelerace vozidla i , b_i je maximální decelerace vozidla i , l_i je efektivní délka vozidla i , která je součtem délky fyzické a minimální vzdálenosti mezi nárazníky vozidel, V_i je cílová (maximální) rychlost vozidla i odpovídající volné dopravě bez omezení jinými vozidly, $x_i(t)$ je poloha vozidla i v čase t . Číselné hodnoty jsou výsledkem kalibrace na základě plovoucího vozidla. První varianta je uplatňována převážně v oblasti volného dopravního proudu, druhá část potom pro oblast vysokých intenzit.

3 Data

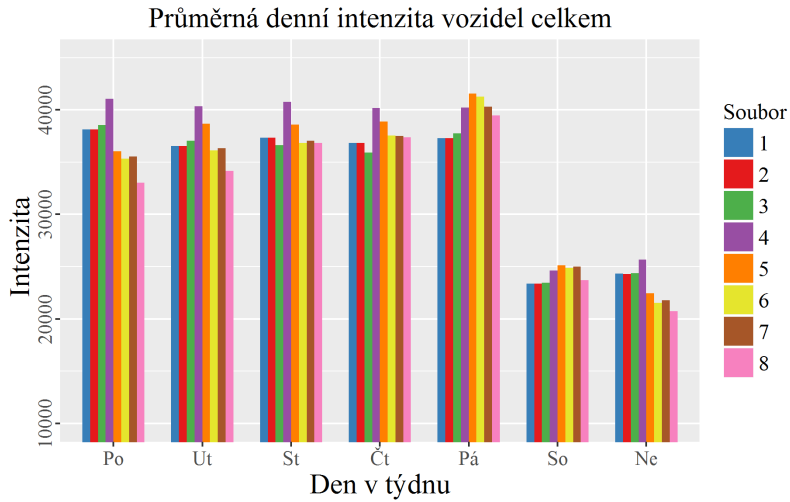
3.1 Měření

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. 7), jimiž naměřená data se podařilo získat.



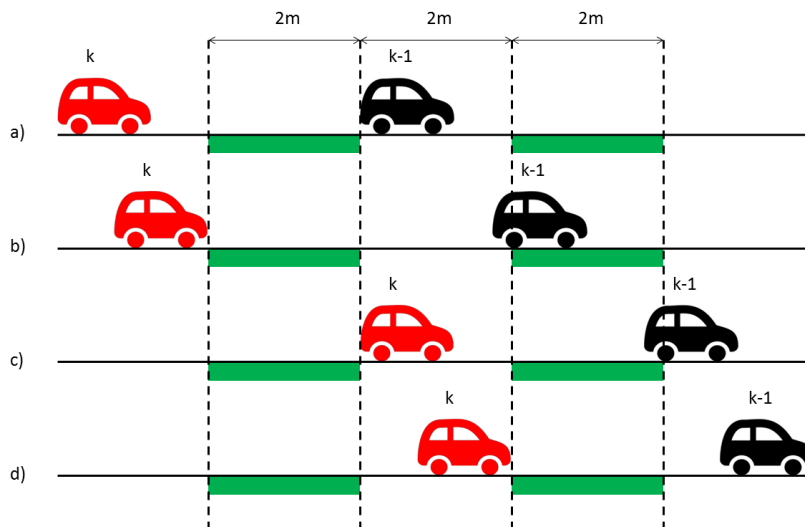
Obrázek 7: Přibližné umístění detekčních profilů na pražském okruhu mezi 16. a 22. kilometrem. Profily 1-4 jsou ve směru severo-západ (po směru staničení komunikace), profily 5-8 poté v protisměru. V každém směru jsou vedeny dva jízdní pruhy.

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. 7). Rozsah těchto souborů je značný. Pokrývají období od listopadu 2013 do května 2014, celkový počet záznamů v jednotlivých souborech se pohybuje kolem 6 milionů (vyjma souboru 4, který se od ostatních liší a obsahuje „pouze“ cca 0,5 milionu řádků). Takovéto množství dat je pro naše účely uspokojivé, ale jak se během práce ukázalo, pro některé okrajové případy hustot dopravního proudu je stále nedostatečné. Prostý průměr denních intenzit (pro oba jízdní pruhy dohromady) je uveden na obr. 8 a odpovídá údajům uvedeným v celostátním sčítání dopravy. Prakticky totožné hodnoty (rozdíl desítek vozidel za sledované období) u datového souboru 1 a 2, které jsou pořízeny v následujících profilech bez možnosti sjetí nebo najetí vozidel, potvrzují kvalitu pořízených dat.



Obrázek 8: Průměr denních intenzit pro jednotlivé datové soubory. Podrobnější informace o průběhu intenzit a rychlostí jsou uvedeny v příloze 8.2

Konstrukce detektoru je dvojitá indukční smyčka, která umožňuje měřit kromě četností i rychlosti vozidel. Smyčka měla délku 2 metry a rozstup náběžných hran smyček byl 4 m. Měřeny jsou celkem 4 události, ze kterých jsou následně dopočítávány charakteristiky dopravního proudu a vozidel (podrobně viz. obr. 9).



Obrázek 9: 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. 9 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

$$g_k = \tau_k^{(b)} - \tau_{k-1}^{(a)}. \quad (9)$$

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 (15). Následně je změřena obsazenost (occupancy), když k -té vozidlo opustí vstupní smyčku

$$o_k = \tau_k^{(c)} - \tau_k^{(b)} \quad (10)$$

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

$$tt_k = \tau_k^{(d)} - \tau_k^{(b)}. \quad (11)$$

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

$$v_k = l_{det}/tt_k \quad (12)$$

a délka

$$l_k = (v_k \cdot o_k) - l_{ic} \quad (13)$$

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 mezera mezi protnutím stejného místa stjenou částí (například zadní nárazník) $(k - 1)$ -ho a k -tého vozidla

$$th_k = g_k + o_k = \tau_k^{(c)} - \tau_{k-1}^{(a)}. \quad (14)$$

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

$$tc_k = th_k - \frac{l_k \cdot 3,6}{v_k}. \quad (15)$$

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 [3]). Pro každou takovou m -tici vozidel můžeme poté definovat intenzitu

$$I_j = \frac{m}{\sum_{k=1}^m th_k} 3600 \quad [\text{voz/h}], \quad (16)$$

hustotu

$$P_j = \frac{I_j}{\frac{1}{m} \sum_{k=1}^m v_k \cdot 3.6} \quad [\text{voz/km}], \quad (17)$$

a rychlost dopravního proudu

$$V_j = \frac{1}{m} \sum_{k=1}^m v_k \cdot 3.6 \quad [\text{km/h}], \quad (18)$$

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)

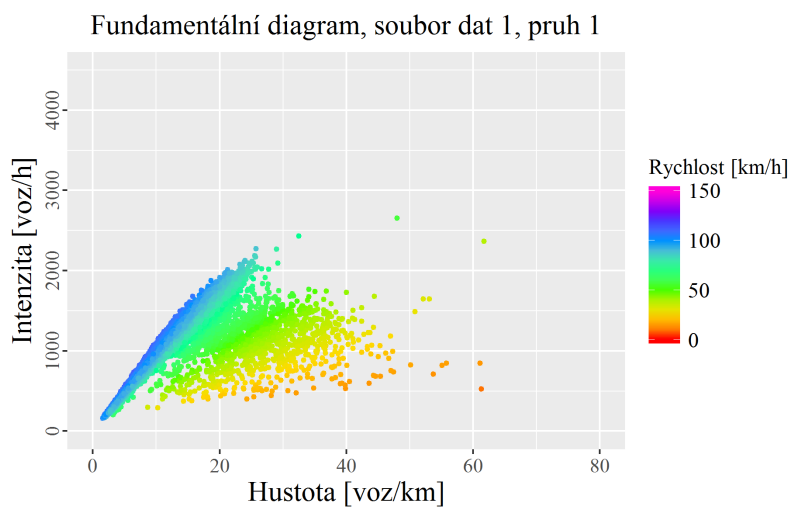
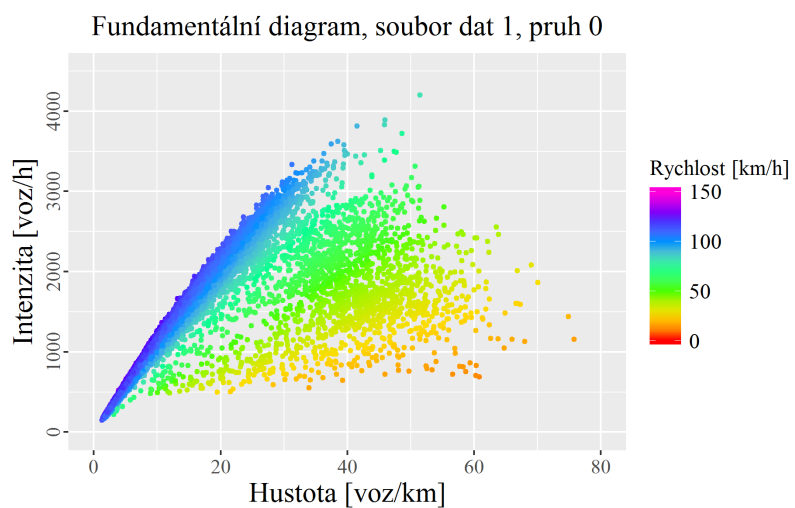
$$s_k = \frac{tc_k}{\frac{1}{m} \sum_{k=1}^m tc_k}, \quad (19)$$

Jednotlivé m -tice poté můžeme vynést do fundamentálního diagramu a přehledně tak zobrazit makroskopické charakteristiky datového souboru (obr. 10). Vzhledem k tomu, že zkoumáme po sobě jdoucí vozidla, je nutné datový soubor nadále rozdělovat na jednotlivé jízdní pruhy. Pruhy jsou číslovány od středu komunikace, 0 - levý jízdní pruh, 1 - pravý jízdní pruh.

Abychom měli představu o struktuře dat kompletní, uvedme 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
<i>tc</i> , průměr [s]	3,4	3,8
<i>tc</i> , s. odchylka [s]	4,4	3,9
<i>tc</i> , min. [s]	0,3	0,4
<i>tc</i> , median [s]	1,7	2,5
<i>tc</i> , 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



Obrázek 10: Fundamentální diagramy pro soubor dat 1. Fundamentální diagramy pro ostatní datové soubory jsou velice podobné a jsou uvedeny v příloze 8.3.

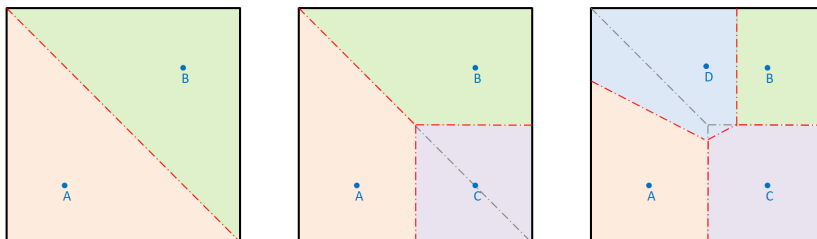
4 Použité statistické nástroje

4.1 Shluková analýza K-means

Pojem *k-means* poprvé použil ve své práci James MacQueen v roce 1967 [19]. Základy metody však sahají dále do historie. Otcem původní myšlenky je Hugo Steinhaus a formuloval ji v roce 1957 [24]. První algoritmus pro její řešení navrhl ve stejném roce Stuart Lloyd [18]. Matematická formulace je následující [27]: 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) (20)

$$\arg \min \sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|^2, \quad (20)$$

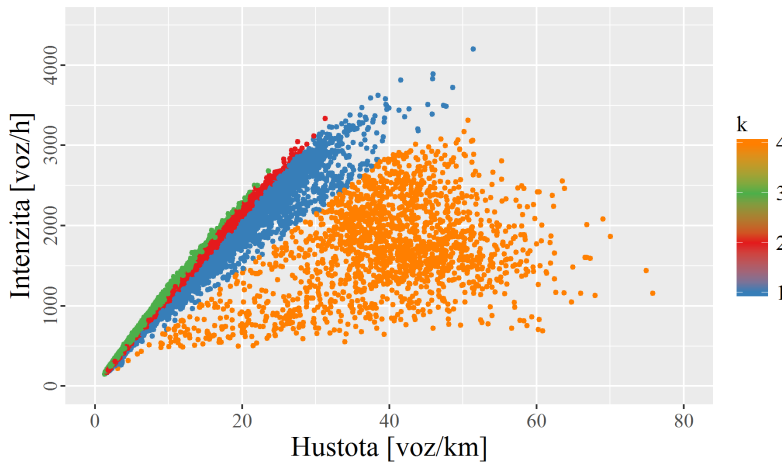
kde μ_i je střední hodnota pozorování v S_i . Výpočet probíhá v krocích. V prvním kroku jsou (zpravidla náhodně) vybrány středy shluků (k náhodně vybraných pozorování), tak zvané centroidy. Pomocí Voronoiovy teselace (obr. 11) je prostor rozdělen na k oblastí a do nich jsou přiřazena jednotlivá pozorování. 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 (20) nebo maximální počet iterací.



Obrázek 11: Voronoiova teselace je dekompozice spojitého prostoru na základě vzdáleností všech bodů tohoto prostoru k diskrétní podmnožině bodů M , kdy se každému bodu b z množiny M přidělí oblast $V(b)$ tak, aby všechny body v oblasti $V(b)$ byly blíže k bodu b než k jakémukoliv jinému bodu z množiny M . Jednou z možností aplikace je přírůstkový algoritmus, kdy najdeme řešení pro jednoduchý případ (např. dvou bodů), postupně přidáváme body další a dekomponujeme prostor na dílčí oblasti.

Proces konverguje velice rychle, ale vzhledem k náhodné počáteční volbě centroidů není pro některé případy zaručen stejný výsledek při každém opakování. Při aplikaci na dopravní data, konkrétně fundamentální diagram, se však proces chová velice stabilně a je dosahováno stejných výsledků. Na obr. 12 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 12: 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

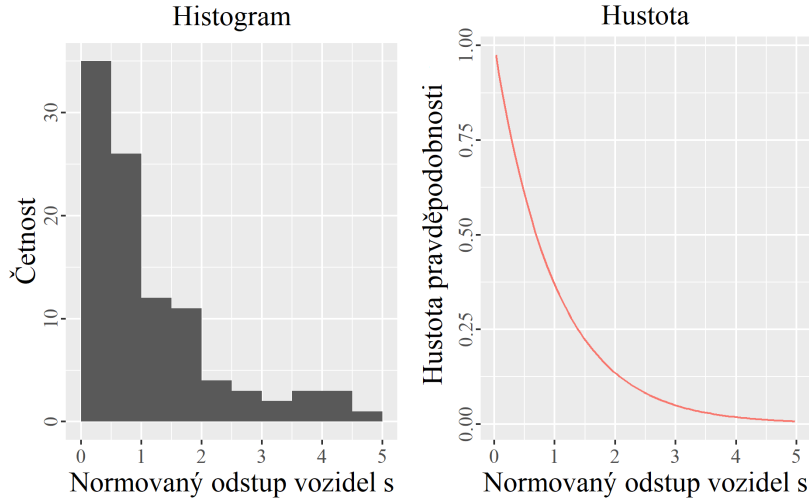
K popisu pravděpodobnosti výskytu náhodného diskrétního jevu používáme tak zvanou pravděpodobnostní funkci (tab. 2), která každému výsledku předepisuje pravděpodobnost výskytu. Součet pravděpodobností výskytu jednotlivých jevů je roven jedné.

Tabulka 2: Pravděpodobnostní funkce výsledku hodu hrací kostkou.

x_i	$P(X = x_i)$
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

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 13. 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 [20]



Obrázek 13: Podobnost histogramu a hustoty pravděpodobnosti (aproximace odstupů pomocí exponenciálního rozdělení (21)).

$$g(s) = \Theta(s)\lambda e^{-\lambda s}, \quad (21)$$

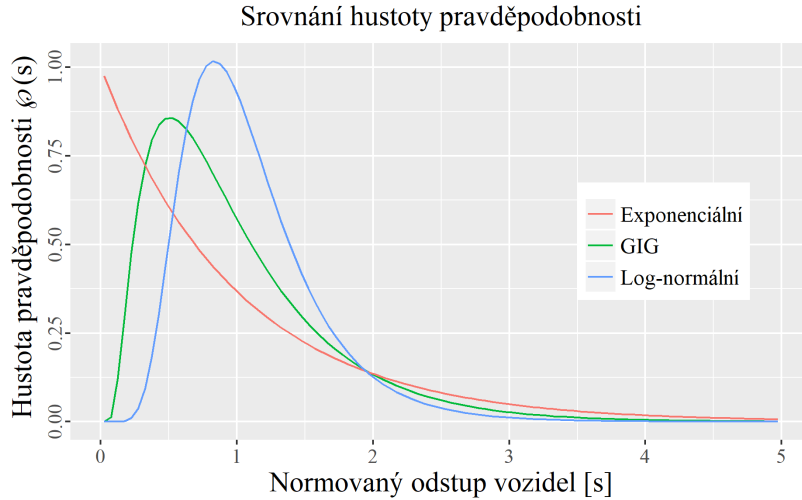
kde $\lambda > 0$ je parametr související s danou veličinou, v našem případě normovaným časovým odstupem vozidel s a Θ je Heavisidova funkce

$$\Theta(s) = \begin{cases} 1 & \text{pro } s > 0 \\ 0 & \text{pro } s \leq 0 \end{cases} \quad (22)$$

Tato aproximace je poměrně dobře použitelná při dopravních stavech s nižší hustotou dopravy. Při vysokých hustotách však začne selhávat (jak je vidět na obr. 14). Pravděpodobnost pro malé odstupy tak predikuje extrémně nepřesně. Lepší aproximaci empirických dat získáme s takzvaným dvouparametrickým log-normálním rozdělením (23)[23], které již respektuje minimální velikost časového odstupu mezi vozidly při vysokých hustotách

$$g(s) = \Theta(s) \frac{1}{s\sigma\sqrt{2\pi}} e^{-\frac{(\ln s - \mu)^2}{2\sigma^2}}, \quad (23)$$

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.



Obrázek 14: Srovnání exponenciální ($\lambda = 1$, (21)), log-normální ($\sigma = 0,43$; $\mu = 0$; (23)) a GIG ($\alpha = 0$; $\beta = 0,43$; (24)) hustoty pravděpodobnosti.

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 [3], které tyto interakce popisují nepřímo na základě sledování rozdělení hustoty pravděpodobnosti jistých mikroskopických veličin. Tuto metodu Krbálek úspěšně aplikoval na dopravní proud [14]. Bylo však potřeba splnit některé předpoklady. Rovnice použité pro termodynamický plyn jsou platné vždy pouze pro konkrétní systém o určité celkové energii. Pokud chceme zkoumat systém o jiné energii, je potřeba parametry rovnice upravit. Tento předpoklad je tedy nutné aplikovat i v případě, 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 [11] a potvrdil předpoklad, že chování kondenzovaného dopravního proudu může být lokálně (pro úzké hustotní pásmo) chápáno jako chování termodynamického systému (plynu). V následujících pracích [11] [17] [12] [16] 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 [15] poté doplňuje i druhý parametr. Právě toto dvouparametrické rozdělení hustoty pravděpodobnosti je použito v této práci (obr. 14) a

je definováno rovnicí

$$\wp(s) = A\Theta(s)s^\alpha e^{-\beta/s - Ds}, \quad (24)$$

kde

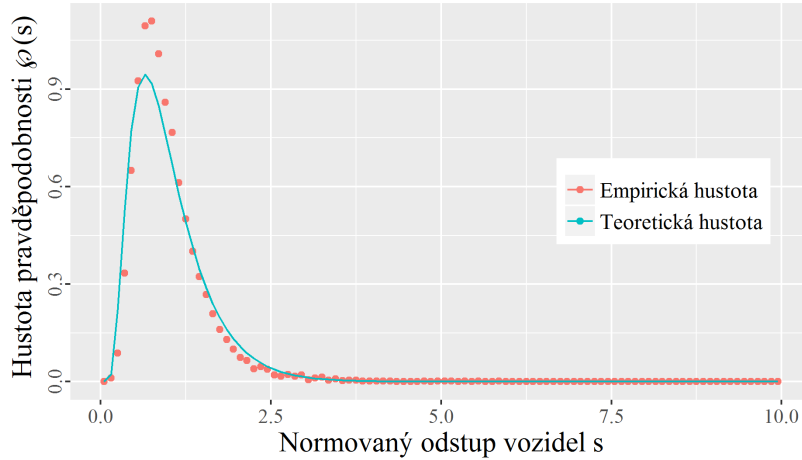
$$D = \alpha + \beta + \frac{3 - e^{-\sqrt{\beta}}}{2} \quad (25)$$

a

$$A^{-1} = 2\left(\frac{\beta}{D}\right)^{\frac{\alpha+1}{2}} \mathcal{K}_{\alpha+1}(2\sqrt{\beta D}). \quad (26)$$

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. 15, $s < 0,5$). 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. 15, $s \gg 2$), kdy se teoretická časová mezera mezi vozidly blíží nekonečnu (v praxi je maximální možná mezera dána nastavením detektoru).

Aproximace hustoty pravděpodobnosti



Obrázek 15: Výsledek fitovacího procesu pro datový soubor 1; pruh 0; $\rho = 55$ [voz/h]; $\alpha = 0$; $\beta = 0,96$.

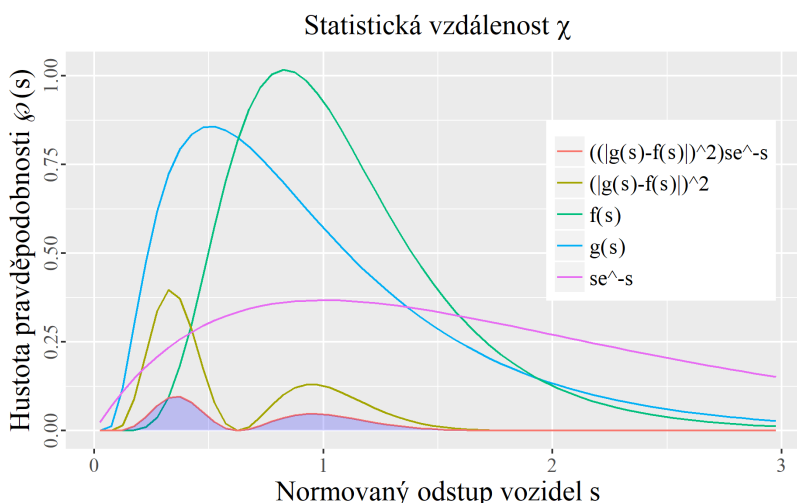
Jak je naznačeno v předchozím textu, odhad parametrů α a β je platný vždy jen pro úzké rozpětí hustoty (v této práci bylo zvoleno $\Delta\rho = 5$ voz/km a pro zjednodušení zápisu je dále v práci vždy uváděna pouze horní hranice rozpětí), která do jisté míry časové odstupy s determinuje a je zřejmé, že absolutní hodnoty s se budou pro různé hustoty a pozorování lišit. Abychom tedy aproximační funkci (24) 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 (24) jsou získány estimačním procesem, neboli fitováním. Při tomto procesu jsou do funkce (24) postupně dosazovány různé hodnoty α a β . Získané teoretické hustoty pravděpodobnosti jsou srovnávány s hustotami empirickými (obr. 15). Srovnávací kvantitou je statistická vzdálenost χ [16] a cílem je pro dané hustotní pásmo a definiční obor α a β najít minimum funkce

$$\chi_{(\alpha,\beta)} = \int_0^{\infty} |\varphi_{\alpha,\beta}(s) - h(s)|^2 se^{-s} ds, \quad (27)$$

kde $\varphi_{\alpha,\beta}(s)$ představuje teoretickou hustotu pravděpodobnosti podle (24), $h(s)$ 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. 16. Podrobněji je využití statistické vzdálenosti v procesu iterace popsáno v kapitole 5.4.



Obrázek 16: Statistické vzdálenosti χ (27) (modrá plocha). Aplikace váhy se^{-s} zvýší význam exponenciální části funkce.

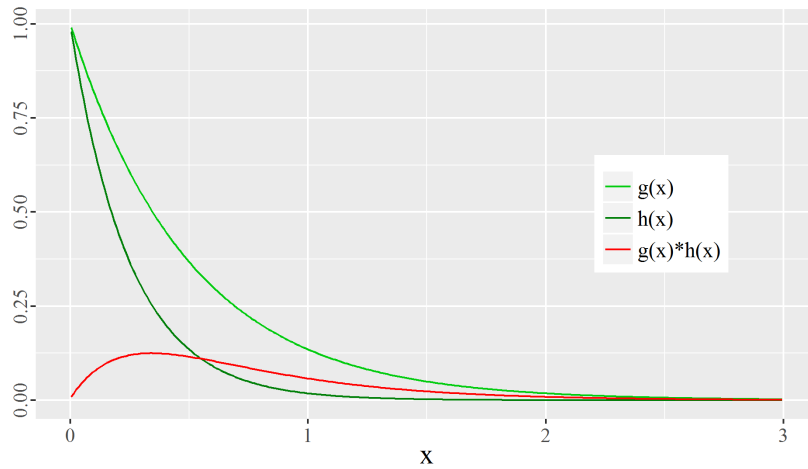
4.5 Konvoluce

Konvoluce je matematická operace, která zpracovává dvě funkce a je definována předpisem [28]

$$g(x) * h(x) = \int_0^x g(y) \cdot h(x - y) dy. \quad (28)$$

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. 17. Pro tuto práci je klíčové statistické využití konvoluce, kdy lze s její pomocí vypočítat kombinovanou hustotu pravděpodobnosti dvou nezávislých veličin (kap. 6.1).

Příklad operace Konvoluce



Obrázek 17: Výsledek konvoluce (28) funkcí $g(x) = \Theta(x)e^{-xa}$ a $h(x) = \Theta(x)e^{-xb}$ pro $a = 2$ a $b = 4$.

4.6 Kolmogorovův-Smirnovův test

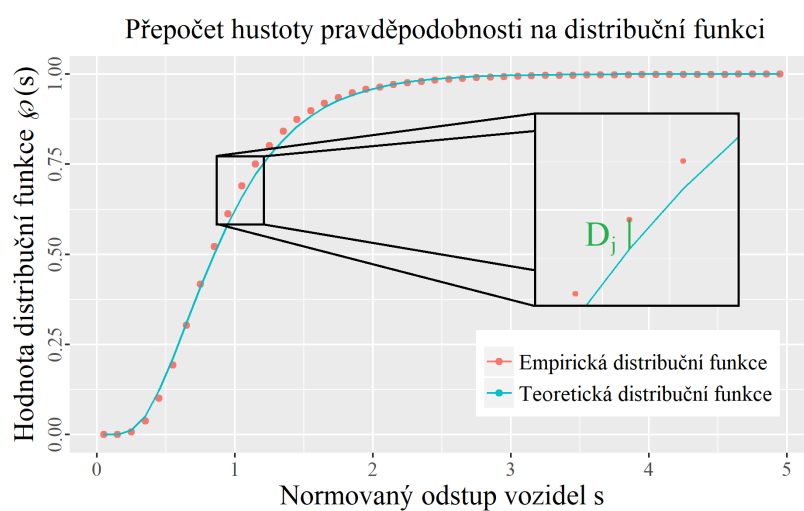
Základní jednovýběrový Kolmogorovův-Smirnovův test se používá k ověření hypotézy, zda náhodný výběr pochází z daného rozdělení pravděpodobnosti (tzv. nulová hypotéza H_0). Základem testu je určení maximální odchylky spojitě (teoretické) distribuční funkce $F_0(x)$ a výběrové (empirické) distribuční funkce $F_n(x)$ (29)

$$D_j = \sup_{-\infty < x < \infty} |F_j(x) - F_0(x)| = \max(D_1^*, D_2^*, \dots, D_j^*), \quad (29)$$

kde

$$D_i^* = |F_j(x_i) - F_0(x_i)|, \quad i = 1, 2, \dots, j. \quad (30)$$

Graficky je hodnota D_j znázorněna na obr. 18. 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 18: Grafická reprezentace D_j (29). Na rozdíl od statistické vzdálenosti χ (obr. 16) 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 3. 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 (10.)

Gap Časová mezera (9).

Ttime Cestovní čas (11).

Speed Rychlost vozidla (12).

Length Délka vozidla (13).

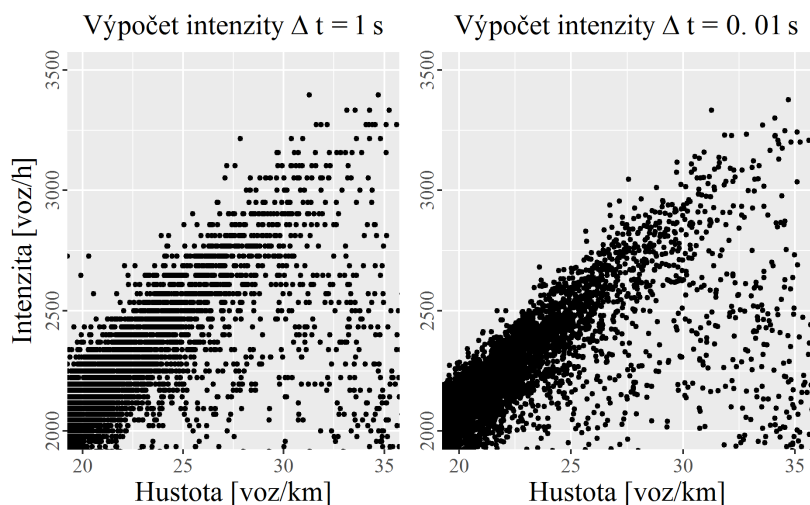
Category Kategorie vozidla (přiřazena podle délky).

Tabulka 3: 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. 19 vlevo). Naštěstí je možné kontinuální čas dopočítat s přesností 0,01 s ze sloupců Gap a Occ (14) a intenzity spočítat s potřebnou přesností (obr. 19 vpravo).

Jakmile je datový soubor doplněn o kontinuální čas, čistý časový odstup tc (15) a časový odstup th (14), 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).



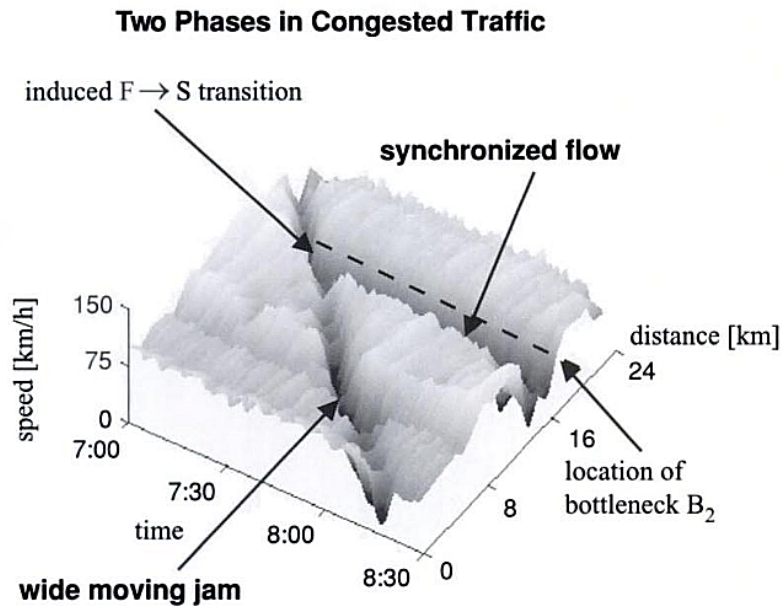
Obrázek 19: Ukázka diskredizace fundamentálního diagramu při nedostatečné přesnosti měření času průjezdu vozidel.

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 [10] předkládá rozdělení dopravního proudu na volný a kondenzovaný, ten dále na synchronizovaný dopravní proud a rozprostřenou pohybující se kolonu. Jak je vidět na obr. 20, pokud máme k dispozici časoprostorová data o vozidlech, můžeme tyto stavy jasně odlišit. Bohužel při použití profilového měření není jednoznačná separace možná.

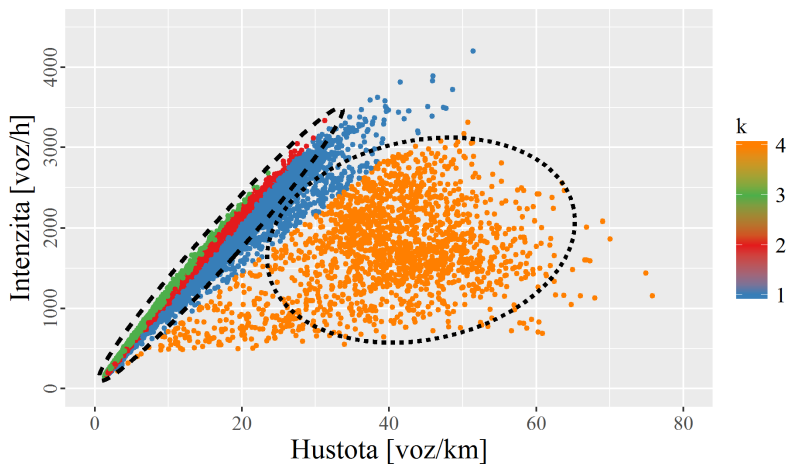
Možné řešení je využití fundamentální hypotézy třífázové teorie dopravy [9], která se pokouší jednotlivé fáze dopravního proudu označit v dvoudimenzionálním prostoru fundamentálního diagramu. Na obr. 21 je podle této teorie čárkovaně naznačena oblast volného dopravního proudu, tečkovaně potom oblast 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 [25]). 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. 21 a v tab. 4. Datový soubor byl vždy rozdělen na 4 shluky, přičemž první dva (sestupně dle rychlosti) byly z datového souboru vyřazeny (na obr. 21 červený(2) a zelený(3) shluk). Takto byly datové soubory upraveny pro další postup.



Obrázek 20: Kernerovo rozdělení dopravy na tři fáze zobrazené v časoprostorovém diagramu [10].

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



Obrázek 21: Naznačení aplikace fundamentální hypotézy třífázové teorie dopravy. Čárkovaně volný, tečkovaně synchronizovaný dopravní proud, k jsou jednotlivé shluky získané metodou K-means (kap. 4.1).

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 $k - 2$,

Tabulka 4: 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]			
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

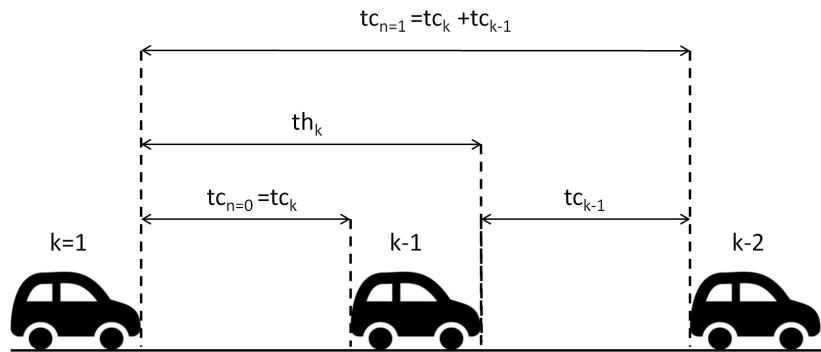
kteřou označíme jako $n = 1$ (viz obr. 22), musíme zvolit vhodnou veličinu. V duchu kapitoly 4.3 budeme sledovat čistý časový odstup vozidel, normovaný na střední hodnotu 1. Jak je však zřejmé z rovnice (15), nezahrnuje čistý časový odstup samotné vozidlo. Naše výpočty se tak posouvají do vyšší roviny 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

$$tc_n = \sum_{k=1}^{n+2} tc_k. \quad (31)$$

5.4 Dvouparametrická estimační metoda

Parametry α a β funkce (24) 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 (27) s přesností určení α a β na dvě desetinná místa. Jako prostředí pro implementaci iterační metody byl zvolen jazyk R [4] 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 (27) a ve výsledné matici je hledána



Obrázek 22: Princip výpočtu multi tc .

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. 23 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.

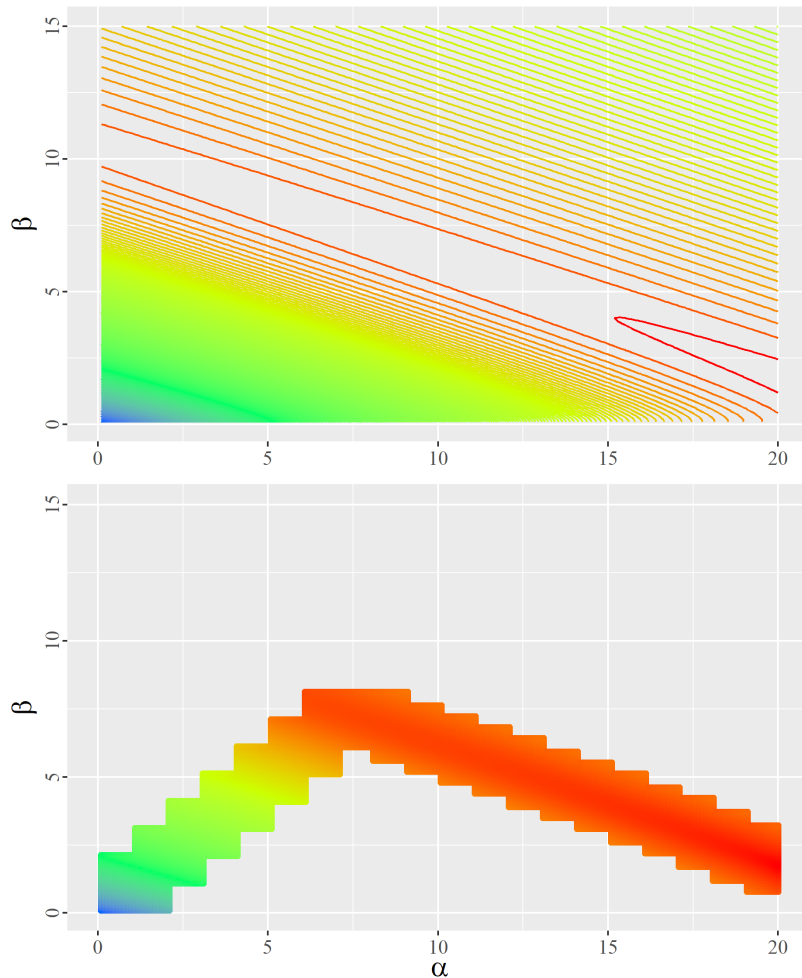
Výsledek tohoto procesu je vidět na obr. 23 dole. V tomto případě je ještě vidět, že minimum matice je omezené definičním oborem α a β . Algoritmus byl tedy doplněn o funkci, která definiční obor α a β rozšiřuje do doby, než je minimum matice nalezeno. Toto minimum často leží na kraji matice, kdy se α nebo $\beta = 0$ (v případě z obr. 23 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. 23 $\alpha \cong \beta \cong 7,5$), 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. 23 červená oblast dat).

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

Ukázka matice statistické vzdálenosti χ , $\rho = 40$ voz/km, $n = 9$



Obrázek 23: 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 [13]. 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

$$V_{h(s)} = V_{\varphi(s)}(\alpha, \beta) (\alpha \geq 0) (\beta \geq 0). \quad (32)$$

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

$$tc_0 = \frac{tc_n}{n+1}. \quad (33)$$

2. Následně musíme rovněž upravit rozptyl

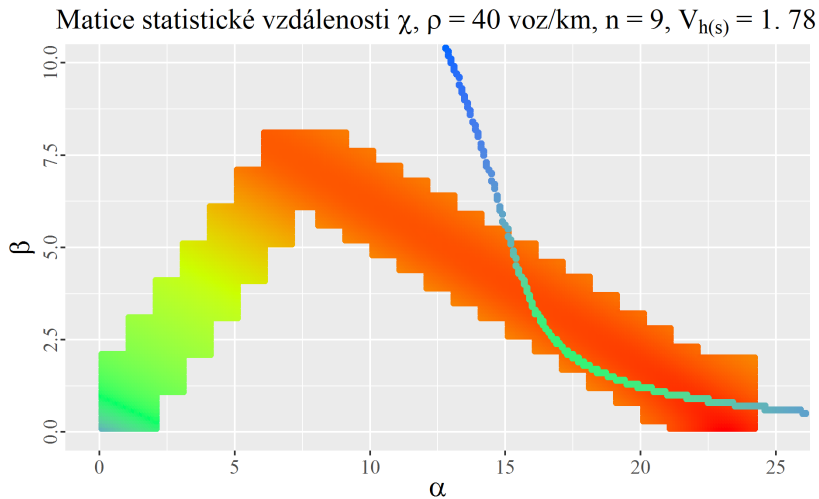
$$V_{h(s)}^{n=0} = \frac{V_{h(s)}^n}{(n+1)^2}. \quad (34)$$

3. Odhadneme hodnoty α a β pro $n = 0$ a splňující podmínku (32).

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

$$\begin{aligned} \alpha_n &= \alpha_0, \\ \beta_n &= \beta_0 \cdot (n+1). \end{aligned} \quad (35)$$

Podmínka (32) nám v matici χ definuje odpovídající dvojce α a β , pro které má smysl iteraci provádět. Na obr. 24 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 (32) 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 24: 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 (32).

5.6 Definiční obor estimace

Původním záměrem bylo provést fitovací proceduru zvláště 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). Jak je vidět z tabulky 6 a 7, 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 diference 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. 4).

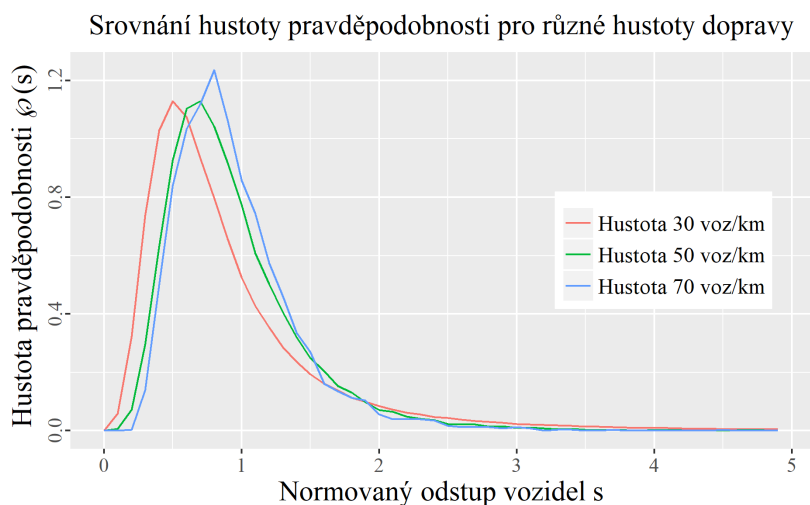
Na další úskalí v množství dat narazíme v situaci, kdy chceme na data uplatnit Kolmogorovův - Smirnovův test (kap. 4.6), jehož kritická hodnota vychází z počtu měření a je mu nepřímě úměrná. Pokud tedy bude j příliš velké, nulová hypotéza bude vždy zamítnuta. Byla tedy zvolena hodnota $j = 5000$, při 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. 8). 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. 5). Jako příklad jsou uvedeny hodnoty z datového souboru 1. Na obr. 27 jsou výsledné hodnoty α a β pro různé hustoty a n (pro sousední vozidla $n = 0$) znázorněny graficky (respektive číselně v tabulce 6 a 7). 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 (24) (viz obr. 26).

Stejným způsobem se chová i empirická hustota pravděpodobnosti časových odstupů tc (obr. 25), 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 (24) 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 25: 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 .

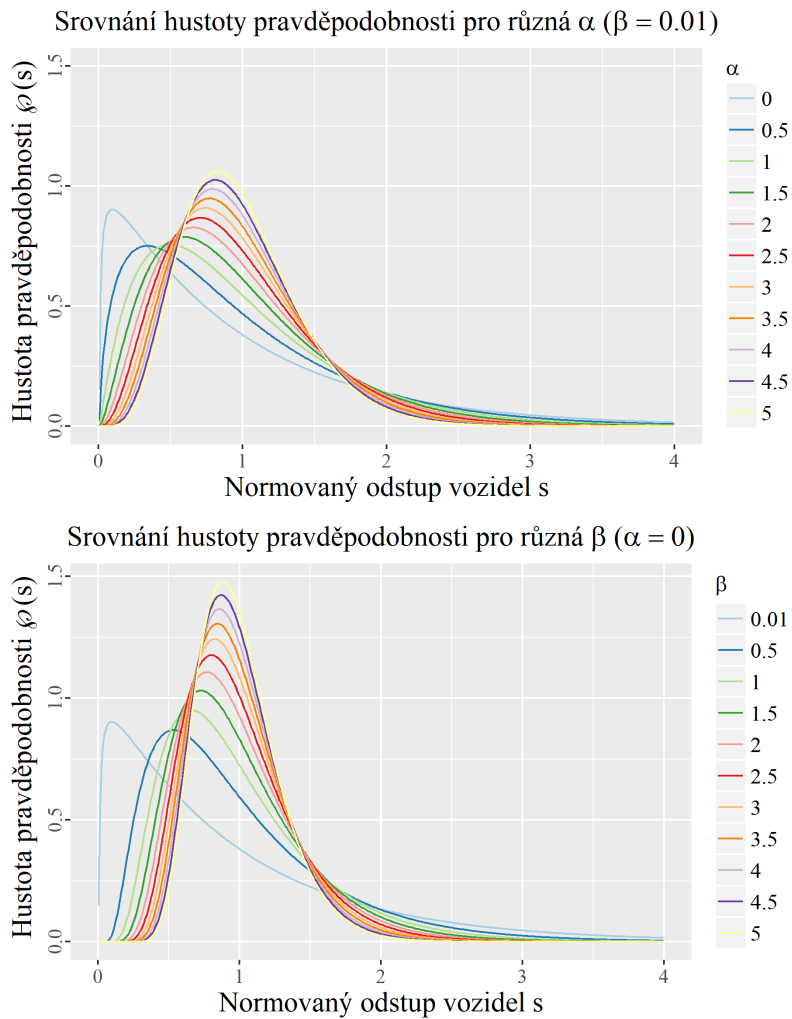
kteřá je způsobena malým počtem měření. Tato nestabilita se projeví rovněž skokovým nárůstem směrodatné odchylky odhadu α a β při vysokých hustotách v tabulce 5. Výsledky z prvního kroku nám poslouží jako kontrolní pro další postup, kdy jsme z důvodů uvedených v předchozí kapitole (5.6) sloučili všechny datové soubory do jednoho a pracovali v rozsahu hustot 25–65 voz/km.

Tabulka 5: Pruh 0, $n = 0$, datové soubory zpracované zvlášť.

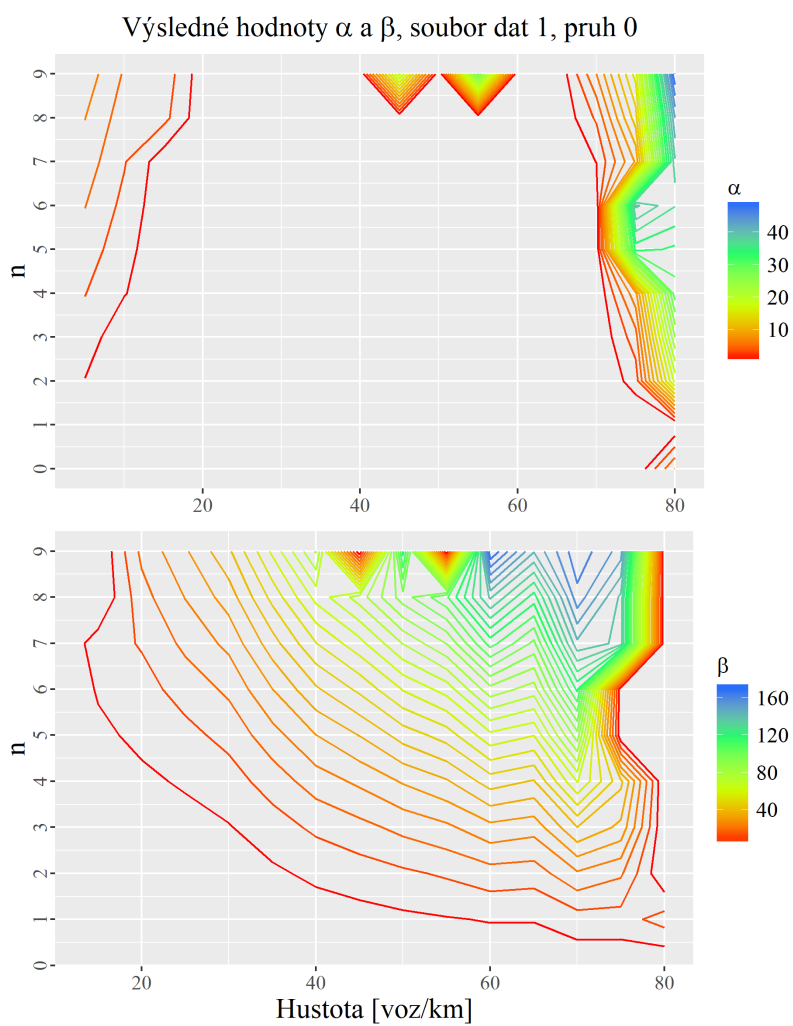
[voz/km]	α		β	
	Průměr	Směrodatná odchylka	Průměr	Směrodatná odchylka
5	0,00	0,00	0,00	0,00
10	0,00	0,00	0,00	0,00
15	0,00	0,00	0,00	0,00
20	0,00	0,00	0,00	0,00
25	0,00	0,00	0,00	0,00
30	0,16	0,15	0,05	0,07
35	0,00	0,00	0,33	0,12
40	0,00	0,00	0,57	0,18
45	0,00	0,00	0,80	0,19
50	0,04	0,12	0,97	0,22
55	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 8, ani po sloučení není dostupných dat příliš. Pro nejvyšší hustotu provozu 65 voz/km je k dispozici právě jedna sada 5000 měření. V tabulce 9 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 26: Ukázka vlivu parametrů α a β na tvar hustoty pravděpodobnosti (24).



Obrázek 27: 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 6: Výsledné hodnoty α a β , počet měření, soubor dat 1, pruh 0.

n	0	1	2	3	4	5	6	7	8	9
[voz/km]	α									
5	0,00	0,67	1,57	2,55	3,32	4,15	4,95	5,76	6,58	7,47
10	0,00	0,00	0,10	0,37	1,74	2,26	2,81	3,41	4,04	4,72
15	0,00	0,00	0,00	0,00	0,15	0,32	0,47	0,62	3,78	4,33
20	0,00	0,00	0,00	0,00	0,00	0,06	0,20	0,34	0,48	0,61
25	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,12	0,29
30	0,21	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
35	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
40	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
45	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	20,28
50	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
55	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	28,41
60	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
65	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
70	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,67	3,51	6,47
75	0,00	0,00	2,36	4,18	7,71	33,43	39,54	8,35	12,30	15,35
80	6,58	0,00	18,45	24,51	30,25	32,33	36,04	39,00	42,48	49,00
	β									
5	0,00	0,02	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00
10	0,00	0,08	0,63	1,12	0,03	0,04	0,04	0,05	0,04	0,04
15	0,00	0,08	0,77	2,03	3,23	4,58	6,32	8,36	0,04	0,05
20	0,00	0,18	1,08	2,59	4,66	7,06	9,38	12,08	15,16	18,67
25	0,00	0,47	1,77	3,77	6,45	9,80	13,83	18,66	23,23	27,99
30	0,00	0,92	2,75	5,42	8,96	13,32	18,49	24,51	31,52	39,27
35	0,20	1,82	4,78	8,88	14,14	20,80	28,73	37,84	47,89	58,81
40	0,43	2,83	6,97	12,72	20,07	28,74	39,40	51,62	65,36	81,23
45	0,66	3,66	8,63	15,54	24,26	34,77	47,45	61,36	77,65	0,10
50	0,88	4,54	10,51	18,92	29,41	42,81	57,99	75,21	95,15	117,28
55	1,07	5,33	12,31	21,88	33,89	48,74	65,81	86,72	108,50	0,14
60	1,17	6,13	14,93	27,11	43,14	61,90	84,39	111,98	139,29	172,90
65	1,22	6,07	14,19	25,40	39,83	57,69	77,85	96,45	118,38	144,59
70	1,85	8,90	22,18	40,29	63,84	89,71	116,49	142,74	159,51	165,72
75	1,49	9,01	18,16	33,14	41,84	0,39	0,33	136,55	136,81	152,97
80	0,00	14,03	0,10	0,18	0,07	0,09	0,12	0,15	0,18	0,24
	Počet měření									
5	10923	10950	10950	10950	10950	10950	10950	10950	10950	10950
10	34504	35041	35135	35148	35150	35150	35150	35150	35150	35150
15	112999	114521	114976	115079	115096	115099	115100	115100	115100	115100
20	143179	144801	145282	145428	145445	145445	145444	145443	145442	145441
25	91962	92837	93046	93088	93097	93096	93100	93100	93100	93100
30	37006	37287	37339	37347	37350	37350	37350	37350	37350	37350
35	20125	20233	20244	20249	20250	20250	20250	20250	20250	20250
40	22471	22530	22550	22550	22550	22550	22550	22550	22550	22550
45	19908	19938	19950	19950	19950	19950	19950	19950	19950	19950
50	13724	13744	13750	13750	13750	13750	13750	13750	13750	13750
55	7640	7649	7650	7650	7650	7650	7650	7650	7650	7650
60	3248	3250	3250	3250	3250	3250	3250	3250	3250	3250
65	1299	1300	1300	1300	1300	1300	1300	1300	1300	1300
70	250	250	250	250	250	250	250	250	250	250
75	100	100	100	100	100	100	100	100	100	100
80	50	50	50	50	50	50	50	50	50	50

Tabulka 7: Výsledné hodnoty α a β , počet měření, soubor dat 1, pruh 1.

n	0	1	2	3	4	5	6	7	8	9
[voz/km]	α									
5	0,17	1,45	2,90	4,46	6,02	7,54	9,13	10,73	12,51	14,15
10	0,00	0,18	2,71	3,88	5,15	6,51	7,88	9,34	10,95	12,65
15	0,00	0,14	0,26	0,34	0,43	7,09	8,39	9,72	11,15	12,66
20	0,00	0,12	0,24	0,25	0,28	0,23	0,20	10,94	12,39	14,00
25	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
30	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
35	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
40	0,00	0,00	0,00	0,00	0,00	0,00	17,87	20,74	23,39	26,03
45	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	18,61	20,74
50	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
55	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
60	0,00	0,00	0,00	0,00	0,00	0,00	1,13	2,39	22,72	24,51
65	0,00	1,16	0,00	0,00	0,00	0,69	1,42	3,05	40,23	49,26
	β									
5	0,02	0,02	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00
10	0,04	0,90	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
15	0,13	1,37	3,70	7,23	11,83	0,00	0,01	0,01	0,02	0,03
20	0,18	1,63	4,30	8,49	14,05	21,28	29,85	0,00	0,00	0,01
25	0,24	2,07	5,49	10,39	16,67	24,63	33,96	45,03	58,07	72,63
30	0,40	2,84	7,25	13,60	21,99	31,93	43,76	58,10	74,56	91,96
35	0,55	3,40	8,41	15,54	24,89	36,37	49,95	66,28	84,61	105,72
40	0,62	3,74	9,54	18,02	28,97	42,81	0,00	0,08	0,10	0,12
45	0,59	3,40	8,13	14,74	23,36	33,77	46,67	61,36	0,00	0,10
50	0,56	3,06	7,61	13,87	21,48	30,26	40,56	52,46	66,55	81,23
55	0,31	1,64	4,06	7,32	11,35	16,39	22,68	30,10	38,22	47,90
60	0,62	4,83	11,29	21,29	33,89	46,89	57,91	69,31	0,24	0,45
65	2,00	7,86	23,45	39,35	62,01	85,58	106,36	131,01	0,20	0,25
	Počet měření									
5	64340	64350	64350	62450	64350	64350	62450	64350	64350	64350
10	58604	58845	58850	56050	58850	58850	56050	58850	58850	58850
15	404365	405403	405449	391000	405450	405450	391000	405450	405450	405450
20	314081	314663	314694	310147	314696	314695	310144	314693	314692	314691
25	41000	41092	41099	41050	41100	41100	41050	41100	41100	41100
30	15022	15042	15050	15050	15050	15050	15050	15050	15050	15050
35	10777	10800	10800	10800	10800	10800	10800	10800	10800	10800
40	5541	5550	5550	5550	5550	5550	5550	5550	5550	5550
45	1698	1700	1700	1700	1700	1700	1700	1700	1700	1700
50	647	650	650	650	650	650	650	650	650	650
55	297	299	300	300	300	300	300	300	300	300
60	150	150	150	150	150	150	150	150	150	150
65	100	100	100	100	100	100	100	100	100	100

Tabulka 8: 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 9: 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

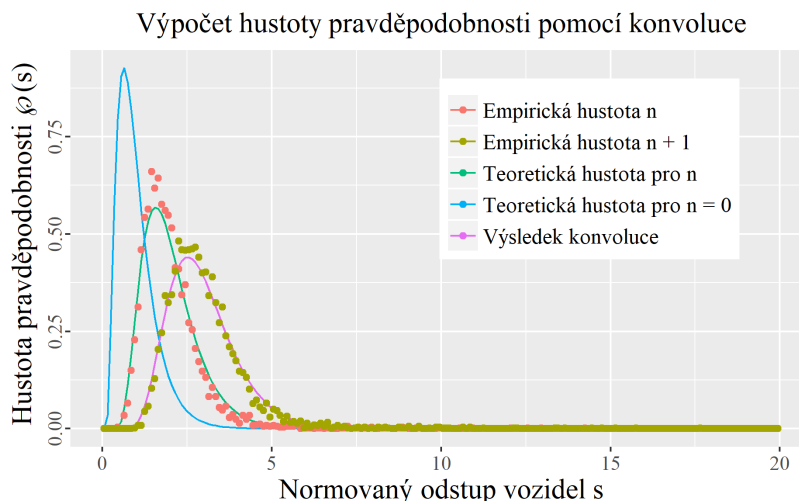
6.1 Aplikace principu konvoluce

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. 22). K aproximaci hustoty pravděpodobnosti jsme použili funkci (24) definovanou pomocí parametrů α a β , které jsme získali fitováním pro vybraná hustotní pásma a hodnoty n (viz tab. 8). 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

$$\varphi(x|n+1) = \varphi(x) * \varphi(x|n), \quad (36)$$

kde v našem případě je x teoretická hustota pravděpodobnosti dle funkce (24) pro $n = 0$ a $x|n$ je teoretická hustota pravděpodobnosti pro $n = \{0, 1, \dots, 9\}$. Výsledek tohoto procesu je vidět na obrázku 28, 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 $x|n$ [13].

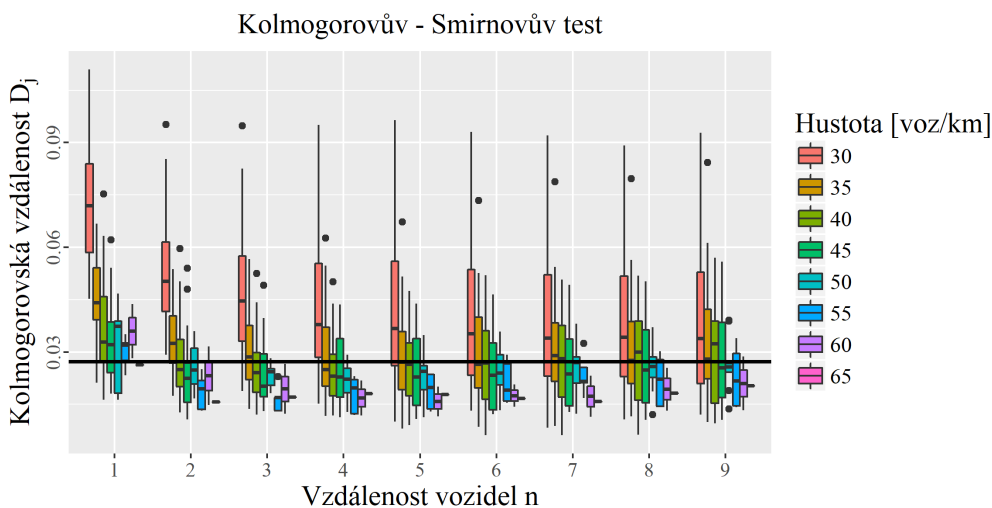


Obrázek 28: Ukázka výpočtu konvoluce, $n = 1$ a 40 [voz/km]

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.

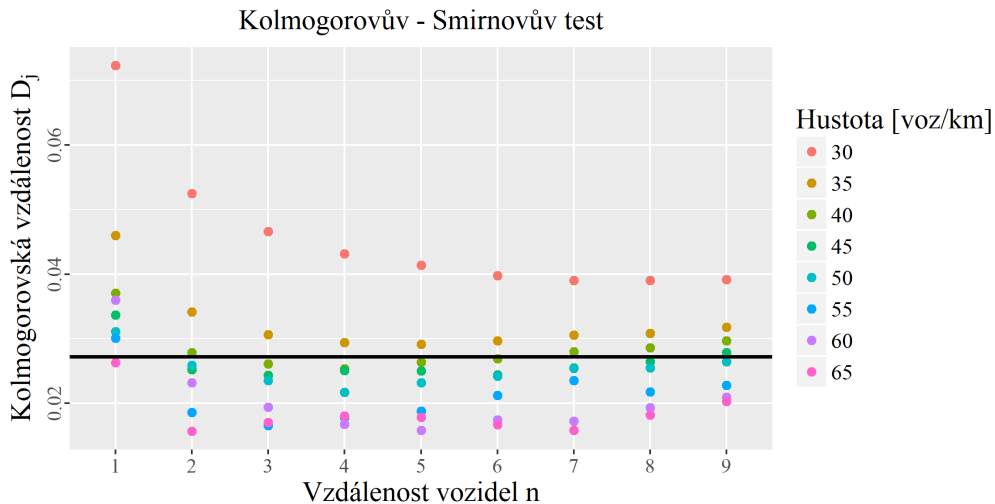
Samotné testování poté probíhá tak, že pro každý blok dat a každé n je vypočtena konvoluce, která je testována s empirickou hustotou pravděpodobnosti tohoto bloku dat, přičemž oboje je pro účel testu převedeno na distribuční funkci. Pro každý blok dat je tedy vyhotoven graf jako na obr. 28, empirická 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 29, 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 29: 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. 8 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 30: Výsledek Kolmogorov-Smirnovova testu - průměr hodnot D_j . Černá čára představuje hodnotu D_{krit} , při jejímž překročení zamítáme nulovou hypotézu.

Při pohledu na obrázek 30 můžeme vidět, pro které hodnoty n a hustoty dopravního proudu byla nulová hypotéza zamítnuta (nad černou čárou, není možné vyloučit interakce mezi vozidly) a pro které hodnoty nebyla zamítnuta (pod černou čárou, 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. 29 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. 22), 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

Podářilo 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 [13] 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. Uvedená 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] Daniel L. Gerlough, M. J. H.: *Traffic Flow Theory*. Transport Research Board, 1975.
- [3] Dereziński Jan, G. J.: *Scattering Theory of Classical and Quantum N-Particle Systems*. Springer-Verlag Berlin Heidelberg, 1997.
- [4] Foundation, R.: The R Project for Statistical Computing.
URL <https://www.r-project.org/>
- [5] Fritzsche, H.-T.: A model for traffic simulation. *Transportation Engineering And Control*, ročník 5, 1994: s. 317–321.
- [6] Gipps, P.: A behavioural car-following model for computer simulation. *Transportation Research Part B: Methodological*, ročník 15, č. 2, April 1981: s. 105–111.
URL <http://ideas.repec.org/a/eee/transb/v15y1981i2p105-111.html>
- [7] Gipps, P. G.: A model for the structure of lane-changing decisions. *Transportation Research Part B: Methodological*, ročník 20, č. 5, 1986: s. 403 – 414, ISSN 0191-2615, doi:DOI: 10.1016/0191-2615(86)90012-3.
- [8] Greenshields B. D., B. J. R.: A STUDY OF TRAFFIC CAPACITY. *Proceedings of the Fourteenth Annual Meeting of the Highway Research Board Held at Washington, D.C. December 6-7, 1934*, ročník 1, 1935: s. 448–477.
URL <https://trid.trb.org/view.aspx?id=120649>
- [9] 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>
- [10] 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.
- [11] Krbalek, M.: Inter-vehicle gap statistics on signal-controlled crossroads. *Journal of Physics A: Mathematical and Theoretical*, ročník 41, č. 20, 2008: str. 205004.
- [12] Krbalek, M.: Theoretical predictions for vehicular headways and their clusters. *Journal of Physics A: Mathematical and Theoretical*, ročník 46, č. 44, 2013: str. 445101.
- [13] 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>
- [14] 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.

- [15] 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.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] 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.bsm/1200512992>
- [20] Medelská, D. V.: *Dopravné Inženýrstvo*. Alfa, 1991.
- [21] Nagel Kai, S. M.: A cellular automaton model for freeway traffic. *J. Phys. I France*, ročník 2, 1992: s. 2221 – 2229.
- [22] Pipes, L. A.: An Operational Analysis of Traffic Dynamics. *Journal of Applied Physics*, ročník 24, 1953: str. 274.
- [23] Prof. Ing. Pavel Příbyl, C.: *Řídicí systémy silniční dopravy*. Fakulta dopravní (FD), 2003.
- [24] 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.
- [25] Szabová, Z.: *Correlation analysis in ensembles of empirical vehicular data*. České vysoké učení technické v Praze, 2016.
- [26] Wiedemann, R.: *Simulation des StraBenverkehrsflusses*, ročník 8. Instituts für Verkehrswesen der Universität Karlsruhe, 1974.
- [27] Wikipedia: k-means clustering.
URL https://en.wikipedia.org/wiki/K-means_clustering
- [28] Wikipedia: Konvoluce.
URL <https://cs.wikipedia.org/wiki/Konvoluce>

Seznam obrázků

1	Původní Greenshieldsovy diagramy závislostí základních veličin dopravního proudu. Vlevo lineární závislost rychlosti a hustoty, v pravo parabolická závislost mezi rychlostí a intenzitou (v diagramu ještě označena jako “Density-vehicles per Hour”) [8].	17
2	Srovnání Greenbergova logaritmického modelu s reálnými daty [2].	19
3	Srovnání Underwoodova exponenciálního modelu s reálnými daty [2].	19
4	Teoretický průběh Greenbergova (čárkovaná čára) a Underwoodova (plná čára) modelu (v tehdejší době se pro označení hustoty ρ používal symbol k) [2] . . .	20
5	Pipesův a Munjalův zobecněný model (v tehdejší době se pro označení hustoty ρ používal symbol k) [2].	20
6	Řada po sobě jdoucích vozidel jak ji definoval Luise A. Pipes [22].	21
7	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.	23
8	Průměr denních intenzit pro jednotlivé datové soubory. Podrobnější informace o průběhu intenzit a rychlostí jsou uvedeny v příloze 8.2	24
9	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. .	24
10	Fundamentální diagramy pro soubor dat 1. Fundamentální diagramy pro ostatní datové soubory jsou velice podobné a jsou uvedeny v příloze 8.3.	28
11	Voroniova teselace je dekompozice spojitého prostoru na základě vzdáleností všech bodů tohoto prostoru k diskrétní podmnožině bodů M , kdy se každému bodu b z množiny M přidělí oblast $V(b)$ tak, aby všechny body v oblasti $V(b)$ byly blíže k bodu b než k jakémukoliv jinému bodu z množiny M . Jednou z možností aplikace je přírůstkový algoritmus, kdy najdeme řešení pro jednoduchý případ (např. dvou bodů), postupně přidáváme body další a dekomponujeme prostor na dílčí oblasti.	29
12	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).	30
13	Podobnost histogramu a hustoty pravděpodobnosti (aproximace odstupu pomocí exponenciálního rozdělení (21)).	31
14	Srovnání exponenciální ($\lambda = 1$, (21)), log-normální ($\sigma = 0,43; \mu = 0$; (23)) a GIG ($\alpha = 0; \beta = 0,43$; (24)) hustoty pravděpodobnosti.	32
15	Výsledek fitovacího procesu pro datový soubor 1; pruh 0; $\rho = 55$ [voz/h]; $\alpha = 0$; $\beta = 0,96$	33
16	Statistické vzdálenosti χ (27) (modrá plocha). Aplikace váhy se^{-s} zvýší význam exponenciální části funkce.	34
17	Výsledek konvoluce (28) funkcí $g(x) = \Theta(x)e^{-xa}$ a $h(x) = \Theta(x)e^{-xb}$ pro $a = 2$ a $b = 4$	35

18	Grafická reprezentace D_j (29). Na rozdíl od statistické vzdálenosti χ (obr. 16) Kolmogorovská vzdálenost D_j zohledňuje pouze maximální odchylku mezi funkcemi.	36
19	Ukázka diskredizace fundamentálního diagramu při nedostatečné přesnosti měření času průjezdu vozidel.	38
20	Kernerovo rozdělení dopravy na tři fáze zobrazené v časoprostorovém diagramu [10].	39
21	Naznačení aplikace fundamentální hypotézy třífázové teorie dopravy. Čárkovane volný, tečkovaně synchronizovaný dopravní proud, k jsou jednotlivé shluky získané metodou K-means (kap. 4.1).	39
22	Princip výpočtu multi tc	41
23	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.	42
24	Znázornění matice χ pro datový soubor 1, rychlý pruh (barevné měřítko 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 (32).	43
25	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	45
26	Ukázka vlivu parametrů α a β na tvar hustoty pravděpodobnosti (24).	46
27	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	47
28	Ukázka výpočtu konvoluce, $n = 1$ a 40 [voz/km]	51
29	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.	52
30	Výsledek Kolmogorov-Smirnovova testu - průměr hodnot D_j . Černá čára představuje hodnotu D_{krit} , při jejímž překročení zamítáme nulovou hypotézu. . .	53

Seznam tabulek

1	Statistické charakteristiky základních veličin, datový soubor 1.	27
2	Pravděpodobnostní funkce výsledku hodu hrací kostkou.	30
3	Ukázka surových dat, soubor 1.	37
4	Výsledek aplikace metody K-means na datové soubory (vždy dva nejrychlejší shluky byly z datového souboru odstraněny).	40
5	Pruh 0, $n = 0$, datové soubory zpracované zvlášť.	45
6	Výsledné hodnoty α a β , počet měření, soubor dat 1, pruh 0.	48
7	Výsledné hodnoty α a β , počet měření, soubor dat 1, pruh 1.	49
8	Pruh 0, datové soubory sloučené.	50
9	Srovnání variant fitování, pruh 0, $n = 0$	50

8 Přílohy

8.1 R code

8.1.1 Úvod

Numerická aplikace analytických funkcí byla stěžejní činností této práce a byla provedena v prostředí jazyka R [4], který je pro podobné aplikace ideální. Hlavní výhodou je jeho univerzálnost, která vychází z open source licence a široké komunity uživatelů, kteří tvoří nepřehledné množství doplňujících balíčků a funkcí. Jazyk R je možné používat jednak jako konzoli pro přímou manipulaci a vyhodnocení dat, nebo vytvářet komplexní skripty a funkce, což byl i případ této práce. Využita byla i možnost paralelizace výpočtu (na 8 vláken) jelikož časová náročnost některých úloh byla při jednovláknové aplikaci neúnosná. Celkem bylo tedy v rámci této práce napsáno cca 1200 řádků kódu a vytvořena řada netriviálních funkcí, díky kterým bylo možné informace uložené v balíku 3,2 GB surových dat vytěžit a použít pro formulaci závěrů uvedených v této práci. V rámci přílohy jsou klíčové funkce stručně popsány a je uveden jejich zdrojový kód, což v první řadě umožní lépe pochopit postup výpočtu a v druhé řadě mohou případní zájemci tyto funkce využít pro svoje vědecké práce nebo ověření výsledků v této práci uvedených.

8.1.2 Načtení a příprava dat

Prvním krokem celého procesu je transformace surových dat do vhodného formátu. Bylo nutné odstranit prázdné a kontrolní sloupce (řádek 4) a sloupce přejmenovat (řádek 6). Poté převést proprietární textový formát data a času do formátu použitelného v jazyce R (řádek 9), vypočítat hodnoty tc a th (řádky 16 a 17) a nakonec datový soubor očistit od odlehlých nebo chybových měření (řádek 18 a 19).

```
1 nacteni_dat_1=function(nazev)
2 {
3   data=read.csv2(paste(nazev,"csv",sep="."),header=FALSE,sep=";",dec="
4     ",strip.white=TRUE,stringsAsFactors=FALSE)
5   data=data[-1,-c(10,9)]
6   #data=data[1:200000,] #testovací zkrácení
7   colnames(data)=c("time","line","occupancy","gap","travel_timme","
8     speed","length","category")
9   data$date=data$time
10  row.names(data)=NULL
11  data$time=as.numeric(as.POSIXct(strptime(data$time, "%Y-%m-%d_%H:%M
12    :%S"))) #represents the (signed) number of seconds since the
13    beginning of 1970 (in the UTC time zone)
14  #převod na číslo
15  for (k in 2:7)
16  {
17    data[,k]=as.numeric(gsub(",",".",data[,k]))
18  }
19  data$id=1
20  data$TH=data$gap+data$occupancy
21  data$TC=round((data$TH-((data$length*3.6)/data$speed),digits=2)
22  data=data[-(which(with(data,TC<0)))]
23  data=data[-(which(with(data,TC>quantile(data$TC, 0.99)))],]
```

```

20 data=data[,c(10,1,9,2,3,4,5,6,7,8,11,12)]
21 #write.table(data,file=paste(paste(nazev,"b",sep="_"),"csv",sep=".")
      ,sep=";",dec="," ,row.names = FALSE)
22 }

```

8.1.3 Generování *m*-tic

Pro účely výpočtu bylo nutné data rozdělit na stejně velké podsoubory *m* vozidel, v našem případě *m* = 50. Tento proces nutně vede ke ztrátě malého množství dat, pokud tato nejsou shodou okolností dělitelná *m*. V opačném případě dojde k oříznutí dat (řádek 7). Pro každou *m*-tici vozidel pak můžeme vypočítat makroskopické (řádky 12 až 15) a mikroskopické (řádky 16 až 18) charakteristiky dopravního proudu.

```

1 pade=function(data_jeden_pruh)
2
3 {
4   soubor_50=NULL
5
6   poradi=1
7   for (1 in seq(1, ((floor(nrow(data_jeden_pruh)/50))*50), by=50))
8   {
9     pade=data_jeden_pruh[1:(1+49),]
10    pade$poradi=poradi
11    doba=sum(pade$TH) #kontinuální čas
12    pade$flow=50/doba*3600
13    pade$dens=pade$flow/mean(pade$speed)
14    pade$factor_dens=factor_dens(pade$dens[1])
15    pade$mean_speed=mean(pade$speed)
16    pade$TH_1=pade$TH/mean(pade$TH)
17    pade$TC_1=pade$TC/mean(pade$TC)
18    pade$speed_1=pade$speed/pade$mean_speed
19    soubor_50[[poradi]]=as.list(pade)
20    poradi=poradi+1
21  }
22  soubor_50=rbindlist(soubor_50)
23  return(soubor_50)
24 }

```

8.1.4 Separace volného dopravního proudu metodou K-means

V tomto případě se ukázala výhoda jazyka R, jelikož metoda K-means (podobně jako celá řada dalších matematických operací) je do něj již integrována. Jediné, co bylo potřeba doplnit, je odstranění volného dopravního proudu (řádek 9).

```

1 separace=function(soubor_50)
2 {
3   cl=kmeans(soubor_50$mean_speed,4)
4   soubor_50$cl=cl$cluster
5   centres=as.data.frame(c(1,2,3,4))
6   centres$cl=cl$centers
7   odstranit=order(centres$cl)

```

```

8   soubor_50_b=soubor_50[-(which(with(soubor_50,c1==odstranit[3]|c1==
      odstranit[4]))),]
9   return(soubor_50_b)
10 }

```

8.1.5 Výpočet multi *tc*

Výpočet multi *tc* je poměrně triviální záležitost (31). Algoritmizace úlohy ovšem už zcela triviální nebyla z důvodu nutnosti zachování optimální struktury datové proměnné. Výpočet bylo tedy nutno rozdělit do několika kroků a s daty manipulovat postupně.

```

1  soucet_vzdalenosti=function(soubor_50_ovlivnene,n) #n je počet mezer!
2  {
3    TC=as.data.frame(soubor_50_ovlivnene$TC_1)
4    if (n==1)
5    {
6      colnames(TC)=paste("TC",101,sep="_")
7      TC_2_temp=TC
8      colnames(TC_2_temp)=paste("TC",1001,sep="_")
9      TC=cbind(TC,TC_2_temp)
10   soubor_50_ovlivnene=as.data.frame(soubor_50_ovlivnene)
11   soubor_50_ovlivnene=cbind(soubor_50_ovlivnene,TC)
12 } else
13 {
14   uprange=n-1
15   for (i in 1:uprange)
16   {
17     TC_temp=as.data.frame(TC[-1,i])
18     TC=as.data.frame(TC[-nrow(TC),])
19     TC[,i+1]=TC_temp+TC[,1]
20   }
21
22   colnames(TC)=paste("TC",seq(0:uprange)+100,sep="_")
23
24   #přidání dělených TC
25   TC_2_temp=TC
26   for (i in 1:n)
27   {
28     TC_2_temp[,i]=TC_2_temp[,i]/i
29   }
30   colnames(TC_2_temp)=paste("TC",seq(0:uprange)+1000,sep="_")
31   TC=cbind(TC,TC_2_temp)
32   soubor_50_ovlivnene=as.data.frame(soubor_50_ovlivnene)
33   soubor_50_ovlivnene=soubor_50_ovlivnene[-((nrow(soubor_50_ovlivnene)
      -n+2):nrow(soubor_50_ovlivnene)),]
34   soubor_50_ovlivnene=cbind(soubor_50_ovlivnene,TC)
35 }
36 return(soubor_50_ovlivnene)
37 }

```

8.1.6 Vytvoření podsouboru *tc* a rozdělení na hustotní pásma

Následující funkce vyjme z celého souboru dat hodnoty *tc* a rozdělí je do jednotlivých podsouborů podle hustotních pásem dopravního proudu. V tuto chvíli se tedy úloha posouvá do vyšší úrovně abstrakce, kde pracujeme s hustotou pravděpodobnosti *tc*, jakožto nepřímou reprezentací chování dopravního proudu. Z praktického hlediska je maximální hodnota *tc* omezena (řádek 9), aby bylo možné hustoty pravděpodobnosti pro jednotlivá hustotní pásma dopravního proudu srovnávat.

```
1 dataprohustotu_ekvi=function(soubor_50_vzdalenosti,max_dens,n)
2 {
3   factor=as.data.frame(unique(soubor_50_vzdalenosti$factor_dens))
4   factor=factor[order(factor),]
5   pdens_50=NULL
6   for (i in 1:length(factor))
7   {
8     temp=as.vector(soubor_50_vzdalenosti[(which(with(soubor_50_
9       vzdalenosti,factor_dens==factor[i]))),(36+n))])
10    pdens_50[[i]]=temp[temp<max_dens]
11  }
12  names(pdens_50)=t(factor)
13  return(pdens_50)
14 }
```

8.1.7 Výpočet hustoty pravděpodobnosti

Samotný výpočet hustoty pravděpodobnosti využívá zabudovanou funkci jazyka R (řádek 6). Hodnoty *max_dens* a *step* nám zaručují, že všechny hustoty pravděpodobnosti mají stejné rozlišení a je tedy možné je mezi sebou srovnávat.

```
1 pdens_50_graf=function(pdens_50,max_dens,step) #step=0.1, šířka
2   sloupce histogramu
3 {
4   pdens_50_graf=NULL
5   for (d in 1:length(pdens_50) )
6   {
7     graf_hustota=hist(pdens_50[[d]],freq=FALSE,breaks =seq(0,max_dens,
8       by=step)) #seq(0,max,by=max/200)
9     pdens_50_graf[[d]]=graf_hustota$density
10  }
11  names(pdens_50_graf)=names(pdens_50)
12  return(pdens_50_graf)
13 }
```

8.1.8 Fitování

Fitovací procedura je jednou z klíčových funkcí celého procesu a slouží k odhadu parametrů α a β . Prošla dlouhým vývojem a existuje ve třech klíčových verzích postupně použitých pro účely kapitol 5.4 a 5.5. Pro zachování přehlednosti je uvedena poslední verze využívající ekvivalentního rozptylu.

Prvním krokem je zjištění původního rozptylu daného hustotního pásma (řádek 11), přičemž tento rozptyl je upraven koeficientem 0,81. Tento koeficient je zjištěn empiricky a

upravuje nesoulad empirického rozptylu tc a teoretického rozptylu použitého pro výpočet možných hodnot α a β [13]. Vzhledem k tomu, že parametry α a β jsou stanoveny pro rozsah rozptylu 0,010 až 0,946, je nutné zavést podmínku pro případy, které leží mimo tento rozsah (řádek 13 a 14). Jakmile máme vybranou příslušnou množinu parametrů α a β (řádek 30), můžeme spočítat všechny hustoty pravděpodobnosti $\varphi(s)$ (24) (řádek 40), příslušné hodnoty χ (27) (řádek 41). Z množiny χ poté vybereme minimální hodnotu χ a k ní náležící parametry α a β (řádek 43 až 45).

```

1 chi_min_fast_ekvi=function(ekvidata,pdens_50,pdens_50_0,soubor_50_graf
  ,max_dens,step,n,soubor,line)
2 {
3   s=seq((step/2),max_dens,by=step)
4   vysledek=NULL
5   for (dens in 1:length(soubor_50_graf))
6     {
7       begTime=Sys.time()
8       s50=as.data.frame(soubor_50_graf[dens])
9       var50=round(var(as.data.frame(pdens_50_0[dens])),digits=3) #
      původní rozptyl z původního hustotního pásma
10      var50b=var50
11      var50=round(0.81*var50[1]/(n^2),digits=3)
12      poc_mer=nrow(as.data.frame(pdens_50[dens]))
13      if (var50<0.01) {var50=0.01}
14      if (var50>0.946) # pozor, pro velký rozptyl beta = 0
15        {
16
17          chi_alfa=0
18          chi_beta=0
19
20          alfa=chi_alfa
21          beta=chi_beta
22          D=beta+alfa+((3-exp(-((beta)^(1/2))))/2)
23          A=1
24          oh=as.data.frame(A*(s^(alfa))*exp(-(beta/s)-(D*s)))
25          chi_min=sum((((abs(oh-s50))^2)*(s*exp(-s))*step)) #bez váhy
      sum((((abs(oh-s50))^2)*(s*exp(-s))*step)) s váhou sum((((
      abs(oh-s50))^2)*(s*exp(-((1/n)*s))))*step))
26
27        } else
28        {
29
30          var=ekvidata[(which(with(ekvidata,var==var50)))]
31          chi=NULL
32
33          for (i in 1:nrow(var))
34            {
35
36              alfa=var$alfa[i]
37              beta=var$beta[i]
38              D=beta+alfa+((3-exp(-((beta)^(1/2))))/2)
39              A=1/(2*((beta/D)^((alfa+1)/2))*besselK(2*((beta*D)^(1/2)),
      alfa+1, expon.scaled = FALSE))
40              oh=as.data.frame(A*(s^(alfa))*exp(-(beta/s)-(D*s)))

```



```

41     chi[i]=sum((((abs(oh-s50))^2)*(s*exp(-s))*step)) #maticový
        nebo skalární součin?
42   }
43   chi_min=min(chi,na.rm = TRUE)
44   chi_alfa=var[which.min(chi),1]
45   chi_beta=var[which.min(chi),2]*n
46   }
47   Sys.time()-begTime
48   dens_name=names(soubor_50_graf[dens])
49   vysledek=rbind(vysledek,c(dens_name,chi_min,chi_alfa,chi_beta,
        var50b,var50,poc_mer))
50   colnames(vysledek)=c("hustota","chi","alfa","beta","var_orig","var
        _red","pocet_mer")
51   graf_vysledek(chi_alfa,chi_beta,s,s50,dens_name,n,soubor)
52   }
53   return(vysledek)
54 }

```

8.1.9 Vykreslení grafu

Funkce vykreslující graf, který je na obr. 15, zde uvedeme hlavně pro úplnost. Postup vykreslování ostatních grafů je obdobný. Řádky 3 až 22 slouží k přípravě dat do formátu požadovaného funkcí ggplot, jejíž parametry na řádcích 25 až 36 umožňují uživateli upravit výsledný vzhled do nejmenších detailů.

```

1 graf_vysledek=function(chi_alfa,chi_beta,s,s50,dens_name,n,soubor)
2 {
3   graf=NULL
4   alfa=chi_alfa
5   beta=chi_beta/n
6
7   if (beta==0)
8   {
9     D=beta+alfa+((3-exp(-((beta)^(1/2))))/2)
10    A=1
11  } else
12  {
13    D=beta+alfa+((3-exp(-((beta)^(1/2))))/2)
14    A=1/(2*((beta/D)^((alfa+1)/2))*besselK(2*((beta*D)^(1/2)), alfa+1,
        expon.scaled = FALSE))
15  }
16  graf$odhad=as.data.frame(A*(s^(alfa))*exp(-(beta/s)-(D*s)))
17  graf$graf_50=s50
18  graf$s=s
19  graf=as.data.frame(graf)
20  colnames(graf)=c("odhad","graf_50","s")
21  plotname=paste("ekvivar_soubor",soubor,"line",line,"n",n,"hustota",
        dens_name,"alfa",alfa,"beta",beta*n,sep="_")
22  plotname_popis=paste("Odhad_parametrů_pro_datový_soubor",paste(
        soubor,"",sep=""),"pruh",paste(line,"",sep=""),paste("hustota="
        ,dens_name,"[voz/h]",",",sep=""),paste("alfa=",round(alfa,digits =
        2),sep=""),paste("beta=",round(beta*n,digits = 2),sep=""),sep=" ")

```

```

23
24 ggplot(graf, aes(x=s)) +
25   geom_point(aes(y=graf_50, color="Empirická_hustota")) +
26   geom_line(aes(y=odhad, color="Teoretická_hustota")) +
27   theme(legend.title = element_blank(), legend.position=c(.8, .5)) +
28   theme(legend.text = element_text(size=12, family="Times_New_Roman")
29         ) +
29   ylab(TeX("Hustota_přavděpodobnosti_\\wp_{(s)}")) +
30   xlab(TeX("Normovaný_odstup_vozidel_{s}")) +
31   ggtitle(TeX("Aproximace_hustoty_přavděpodobnosti")) +
32   theme(axis.title.x = element_text(face="bold", colour="#000000",
33         size=15.7, family="Times_New_Roman"), #úprava vzhledu popisků
34         osy
35         axis.text.x = element_text(angle=0, vjust=0.5, size=11,
36         family="Times_New_Roman")) + #úprava vzhledu hodnot na
37         ose
38         theme(axis.title.y = element_text(face="bold", colour="#000000",
39         size=15.7, family="Times_New_Roman"),
40         axis.text.y = element_text(angle=90, hjust=0.5, size=11,
41         family="Times_New_Roman")) +
42         theme(plot.title = element_text(face="bold", size=15.7, family="
43         Times_New_Roman"))
44
45 ggsave(file=paste(plotname, "png", sep="."), width=160, height=100, units
46         ="mm", dpi=300)
47 write.csv2(graf_srovnani_distribuce, file=paste(paste(names(data_1_0_
48         pdens)[rychlost], file, beta, "d", sep="_"), "csv", sep="."))
49 }

```

8.1.10 Estimační procedura

Pokud výše uvedené funkce seřadíme do správné posloupnosti, dosáhneme kýženého výsledku (řádek 14). Poněkud složitější je zajistit správnou organizaci celého procesu tak, abychom výpočet provedli pro všechny relevantní kombinace ρ , n a výsledky jsme dostaly v přehledné a použitelné podobě (řádky 16 až 37).

```

1 data_vse=read.table("data_vse.csv", header=TRUE, sep=";", dec=".",
2   stringsAsFactors=FALSE)
3 soubor="all"
4 line="rychlý"
5 ekvidata=read.csv2("ekvidata.csv")
6
7 for (n in 1:1)
8 {
9   max_dens=20
10  step=0.1 #dělení histogramu
11  pdens_50_0=dataprohustotu_variance(data_vse, max_dens, n)
12  pdens_50=dataprohustotu_ekvi(data_vse, max_dens, n)
13  soubor_50_graf=pdens_50_graf(pdens_50, max_dens, step)
14  vysledek=as.data.frame(chi_min_fast_ekvi(ekvidata, pdens_50, pdens
15    _50_0, soubor_50_graf, max_dens, step, n, soubor, line),
16    stringsAsFactors=FALSE) #hustota pravděpodobnosti

```

```

15
16 if (n==1) {
17   vysledek_beta0=as.data.frame(chi_min_fast_ekvi_beta0(ekvidata,
18     pdens_50,pdens_50_0,soubor_50_graf,max_dens,step,n,soubor,line)
19     ,stringsAsFactors=FALSE) #hustota pravděpodobnosti
20   vysledek$beta0=vysledek_beta0$beta0
21 } else {
22   vysledek_beta0=read.table(file=paste(paste("ekvivar_vysledek",
23     soubor",soubor,"line",line,"n",1,sep="_"),"csv",sep="."),sep=";
24     ",dec="," ,header = TRUE,stringsAsFactors=FALSE)
25
26   if (sum(as.numeric(vysledek$hustota))==sum(as.numeric(vysledek_
27     beta0$hustota)))
28   {
29     vysledek$beta0=vysledek_beta0$beta0
30   } else {
31
32     for (i in 1:nrow(vysledek))
33     {
34       for (j in 1:nrow(vysledek_beta0))
35       {
36         if (vysledek_beta0$hustota[j]==vysledek$hustota[i]) {
37           vysledek$beta0[i]=vysledek_beta0$beta0[j]}
38       }
39     }
40   }
41
42   assign(paste("vysledek",n,sep="_"),vysledek)
43   write.table(vysledek,file=paste(paste("ekvivar_vysledek",soubor",
44     soubor,"line",line,"n",n,sep="_"),"csv",sep="."),sep=";","dec="," ,
45     row.names = FALSE)
46   gc()
47 }

```

8.1.11 Konvoluce

Předposledním krokem celého procesu je výpočet konvoluce (36), který je proveden na řádce 42. Předchozí řádky slouží k přípravě funkcí $\varphi(x)$ (řádky 25 až 38) a $\varphi(x|n)$ (řádky 5 až 21).

```

1 konvoluce=function(alfan,betan,alfa1,beta1,beta0,max_dens,step,
2   plotname,vd_zapis,vd_fit,empir_data0,empir_data1,n)
3 {
4   s=seq((step/2),max_dens,by=step)
5   konv_fce=NULL
6   #funkce gxn
7   if (betan==0)
8   {
9     alfa=as.numeric(alfan)
10    beta=as.numeric(betan)
11    D= (alfa/n)+(beta/(n^2))+((3-exp(-(beta/n)^(1/2)))/(2*n))
12    A=1
13    gxn=as.vector(A*(s^(alfa))*exp(-(beta/s)-(D*s)))

```

```

14
15 } else
16 {
17     alfa=as.numeric(alfan)
18     beta=as.numeric(betan)
19     D= (alfa/n)+(beta/(n^2))+((3-exp(-(beta/n)^(1/2)))/(2*n))
20     A=1/(2*((beta/D)^((alfa+1)/2))*besselK(2*((beta*D)^(1/2)), alfa+1,
        expon.scaled = FALSE))
21     gxn=as.vector(A*(s^(alfa))*exp(-(beta/s)-(D*s)))
22 }
23 #funkce gx0
24
25 if (beta0==0)
26 {
27     alfa=0
28     beta=as.numeric(beta0)
29     D=beta+((3-exp(-(beta)^(1/2)))/2)
30     A=1
31     gx0=as.vector(A*(s^(alfa))*exp(-(beta/s)-(D*s)))
32 } else
33 {
34     alfa=0
35     beta=as.numeric(beta0)
36     D=beta+((3-exp(-(beta)^(1/2)))/2)
37     A=1/(2*((beta/D)^((alfa+1)/2))*besselK(2*((beta*D)^(1/2)), alfa+1,
        expon.scaled = FALSE))
38     gx0=as.vector(A*(s^(alfa))*exp(-(beta/s)-(D*s)))
39 }
40 for (x in 1:length(s))#length(s)
41 {
42     konv_fce[x]=sum(gxn[1:x]*rev(gx0[1:x])*step)
43 }
44 gx1=gx1(alfa1,beta1,max_dens,step,n+1)
45 graf_konvoluce(gxn,gx1,gx0,konv_fce,empir_data0,empir_data1,s,
        plotname,vd_zapis,vd_fit,x)
46 return(konv_fce)
47 }

```

8.1.12 Kolmogorův-Smirnovův test

Posledním krokem celého procesu je provedení Kolmogorov-Smirnovova testu. Složitost této funkce spočívá především v optimalizaci pro paralelní výpočet a několika vnořeným cyklům výpočtu. Test je totiž nutné provést pro všechna n , všechna hustotní pásma a všechny bloky 5000 měření. Celkem je tedy nutné tuto funkci provést 14688x, přičemž je do ní vnořena i fitovací funkce, která sloužila pro alternativní srovnání konvoluce s teoretickou funkcí (nakonec nebyla tato alternativa využita). Parallelizace výpočtu na 8 vláken přinesla zkrácení celého výpočtu na cca 30 min., což umožnilo s funkcí relativně efektivně pracovat. V prvním kroku je sestaven seznam všech výsledků Estimační procedury (8.1.10), z něhož jsou čerpány především hodnoty parametrů α a β pro jednotlivé hustoty dopravního proudu a n (řádky 4 až 8). Na řádcích 10 až 15 je provedena paralelizace výpočtu, kdy jsme s výhodou využili toho, že máme 8 hustotních pásem a výpočet jsme tedy mohli paralelizovat na 8 vláken, díky čemuž jsme na

plno využili procesorový výkon. Na následujících řádcích je proveden Kolmogorův-Smirnovův test (29) pro příslušné veličiny a všechny možné kombinace dat, přičemž výsledky jsou zapsány do přehledné tabulky, který umožňuje jejich další vyhodnocení v libovolném nástroji.

```

1 ks_test=NULL
2   for (i in 1:(nrow(seznam)-1)) #cyklus přes řádky seznamu (přez
   soubory)
3     {
4       nazev=as.data.frame(strsplit(seznam[i,1],"\\."),stringsAsFactors =
         FALSE)[1,]
5       n=seznam$n[i]
6       line=seznam$line[i]
7       fit=read.table(seznam[i,1],header = TRUE ,stringsAsFactors = FALSE
         ,sep=";",dec=",")
8       fit_1=read.table(seznam[i+1,1],header = TRUE ,stringsAsFactors =
         FALSE,sep=";",dec=",")
9
10      detectCores()
11      cl<-makeCluster(8)
12      registerDoSNOW(cl)
13
14      ks_test_temp=NULL
15      ks_test_temp= foreach (d=1:8) %dopar% { #cyklus pře hustoty
16
17        library(parallel)
18        library(foreach)
19        library(doSNOW)
20        library(data.table)
21        library(ggplot2)
22        library(latex2exp)
23        library(extrafont)
24        fonts()
25        #font_import()
26        loadfonts(device="win")
27
28
29        data_d=data[(which(with(data, factor_dens==fit$hustota[d]))
          ,)]
30
31        ks_test_temp_2=NULL
32        #vytvoření vnitřní proměné pro záznam testu přes jednotlivé
          5000covky
33        if (fit$pocet_mer[d]>5000)
34          {
35
36            ks_test_temp_2$n[1:floor(nrow(data_d)/5000)]=n
37            ks_test_temp_2$d[1:floor(nrow(data_d)/5000)]=fit$hustota[d
          ]
38          } else{
39
40            ks_test_temp_2$n[1]=n
41            ks_test_temp_2$d[1]=fit$hustota[d]
42          }

```

```

43
44
45     if (nrow(data_d)>5000)
46     {
47         l2=1
48         for (l in (seq(1, ((floor(nrow(data_d)/5000))*5000), by
49                 =5000))) #cyklus přez 5000
50         {
51             empir_data0=e_data(data_d[1:(l+4999)],,n,max_dens,
52                               step)[[1]]
53             empir_data1=e_data(data_d[1:(l+4999)],,n+1,max_dens,
54                               step)[[1]]
55             empir_data1_1=e_data_1(data_d[1:(l+4999)],,n+1,max_
56                                   dens,step)[[1]]
57
58             #fit alfa beta pro 5000
59             var50=as.numeric(fit_1$var_red[d])
60             vysledek=chi_min_konvoluce(ekvidata,empir_data1_1,
61                                       var50,max_dens,step,n)
62
63             alfan=fit$alfa[d]
64             betan=fit$beta[d]
65             beta0=fit$beta0[d]
66             alfa1=vysledek[2]
67             beta1=vysledek[3]
68             plotname=paste(nazev,"dens",fit$hustota[d],"alfa",
69                             alfan,"beta",betan,"beta0",beta0,"l",l,sep="_")
70             konvoluce_temp=as.data.frame(konvoluce(alfan,betan,
71                                                   alfa1,beta1,beta0,max_dens,step,plotname,vd_zapis,
72                                                   vd_fit,empir_data0,empir_data1,n))
73             ks_test_temp_2$1[12]=1
74             ks_test_temp_2$g_f[12]=max(abs(dfce(konvoluce_temp,
75                                                 step)-dfce(as.data.frame(gx1(alfa1,beta1,max_dens,
76                                                 step,n+1)),step)))
77             ks_test_temp_2$g_e[12]=max(abs(dfce(konvoluce_temp,
78                                                 step)-dfce(as.data.frame(empir_data1),step)))
79             ks_test_temp_2$alfa_1[12]=alfa1
80             ks_test_temp_2$beta_1[12]=beta1
81             ks_test_temp_2$alfa_n[12]=alfan
82             ks_test_temp_2$beta_n[12]=betan
83
84             names(konvoluce_temp)=plotname
85             #konvoluce_vysledek=cbind(konvoluce_vysledek,
86                                       konvoluce_temp)
87             l2=l2+1
88         }
89     } else {
90         l2=1
91
92         empir_data0=e_data(data_d,n,max_dens,step)[[1]]
93         empir_data1=e_data(data_d,n+1,max_dens,step)[[1]]

```

```

84     empir_data1_1=e_data_1(data_d[1:(1+4999)],,n+1,max_dens,
      step)[[1]]
85
86     #fit alfa beta pro 5000
87     var50=as.numeric(fit_1$var_red[d])
88     vysledek=chi_min_konvoluce(ekvidata,empir_data1_1,var50,
      max_dens,step,n)
89
90     alfan=fit$alfa[d]
91     betan=fit$beta[d]
92     beta0=fit$beta0[d]
93     alfa1=vysledek[2]
94     beta1=vysledek[3]
95     plotname=paste(nazev,"dens",fit$hustota[d],"alfa",alfan,
      "beta",betan,"beta0",beta0,"1",1,sep="_")
96     konvoluce_temp=as.data.frame(konvoluce(alfan,betan,alfa1,
      beta1,beta0,max_dens,step,plotname,vd_zapis,vd_fit,
      empir_data0,empir_data1,n))
97     ks_test_temp_2$1[12]=1
98     ks_test_temp_2$g_f[12]=max(abs(dfce(konvoluce_temp,step)
      -dfce(as.data.frame(gx1(alfa1,beta1,max_dens,step,n
      +1)),step)))
99     ks_test_temp_2$g_e[12]=max(abs(dfce(konvoluce_temp,step)
      -dfce(as.data.frame(empir_data1),step)))
100    ks_test_temp_2$alfa_1[12]=alfa1
101    ks_test_temp_2$beta_1[12]=beta1
102    ks_test_temp_2$alfa_n[12]=alfan
103    ks_test_temp_2$beta_n[12]=betan
104
105    names(konvoluce_temp)=plotname
106    #konvoluce_vysledek=cbind(konvoluce_vysledek,konvoluce_
      temp)
107    l2=l2+1
108    }#cyklus přez 5000
109    ks_test_temp_2
110
111    }#cyklus pře hustoty
112    stopCluster(cl)
113    ks_test[[i]]=rbindlist(ks_test_temp)
114    }#cyklus přez řádky seznamu (přez soubory)
115
116    #konvoluce_vysledek=konvoluce_vysledek[,-1]
117    ks_test=rbindlist(ks_test)
118    ks_test=as.data.frame(ks_test)
119
120
121    setwd(vd_vysledky)
122    #write.table(konvoluce_vysledek,file="konvoluce_vysledek_20_01.csv",
      sep=";",dec=".",row.names = FALSE)
123    write.table(ks_test,file="ks_test_5000_fit_20_01.csv",sep=";",dec=".",
      row.names = FALSE)

```

8.1.13 Postprocessing dat

Vzhledem k tomu, že výsledky získané v rámci Estimační procedury (8.1.10) se ukládaly postupně do samostatných souborů, bylo potřeba vytvořit funkci, která jednotlivé výsledky spojí dohromady do jednoho souboru tak, aby bylo možné je přímo prohlížet. Následující funkce má tedy servisní charakter a jedná se pouze o manipulaci s daty. Nicméně stejně jako v předchozích případech bylo logickou výzvou uspořádat vnořené cykly tak, aby nebyly žádné hodnoty opomenuty a zároveň aby ve výsledném souboru nebyly prázdné řádky, protože počet hustotních pásem se u jednotlivých datových souborů lišil.

```
1 agregace_alfa_beta_dohr=function(vd_fit)
2 {
3   setwd(vd_fit)
4   agregace_dohromady=NULL
5
6   soubor="all"
7   line="rychlý"
8
9   seznam=as.data.frame(list.files(vd_fit,pattern=paste("^.*soubor_
10     ",soubor,"_line_",line,"*.csv$",sep="")),stringsAsFactors =
11     FALSE)
12   for (i in 1:nrow(seznam))
13   {
14     nazev=as.data.frame(strsplit(seznam[i,1],"\\. "),
15       stringsAsFactors = FALSE)[1,]
16     nazev2=as.data.frame(strsplit(seznam[i,1],"_"),
17       stringsAsFactors = FALSE)
18     n=as.data.frame(strsplit(nazev2[nrow(nazev2),1],"\\. "),
19       stringsAsFactors = FALSE)
20     seznam$n[i]=as.numeric(n[1,])
21     seznam$line[i]=(nazev2[6,])
22   }
23   seznam=seznam[order(seznam$n),]
24
25   tabulka=matrix(ncol=max(seznam$n)*4+3,nrow = 8)
26   tabulka=as.data.frame(tabulka)
27   tabulka[,3]=c(30,35,40,45,50,55,60,65)
28   tabulka[,2]=line
29   tabulka[,1]=soubor
30   for (k in 1:(nrow(seznam)))
31   {
32     vysledek=read.table(seznam[k,1],sep=";",dec=".",
33       stringsAsFactors = FALSE,header = TRUE)
34
35     for (i in 1:nrow(tabulka))
36     {
37       for (j in 1:nrow(vysledek))
38       {
39         if (tabulka[i,3]==vysledek$hustota[j]) {
40           tabulka[i,k+3]=vysledek$alfa[j]
41           tabulka[i,k+max(seznam$n)+3]=vysledek$beta[j]
42           tabulka[i,k+max(seznam$n)*2+3]=vysledek$var_orig[j]
43           tabulka[i,k+max(seznam$n)*3+3]=vysledek$pocet_mer[j]
```



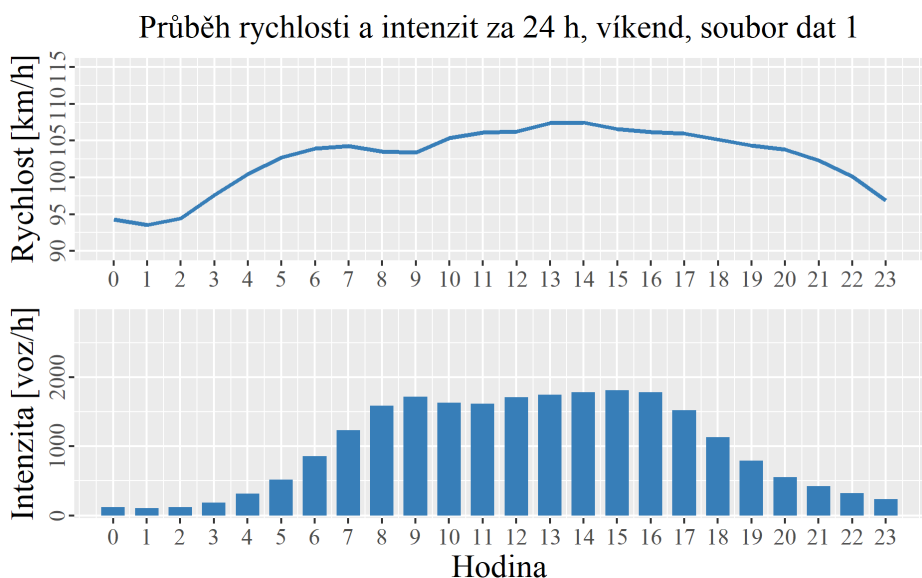
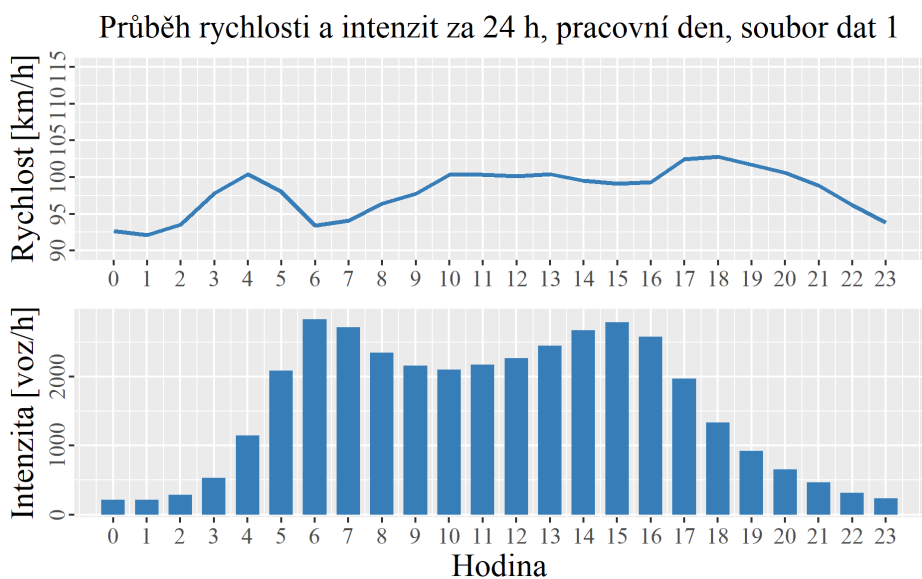
```

38     }
39   }
40
41   }
42 }
43 names(tabulka)=c("soubor","line","hustota",as.vector(paste("alfa
",seq(1,max(seznam$n),1),sep="_")),as.vector(paste("beta",seq
(1,max(seznam$n),1),sep="_")),as.vector(paste("var",seq(1,max
(seznam$n),1),sep="_")),as.vector(paste("poc_mer",seq(1,max(
seznam$n),1),sep="_")))
44 assign(paste("ab",soubor,line,"02",sep="_"),tabulka)
45 write.csv2(file="agregace_ab.csv",ab_all_rychlý_02,row.names =
FALSE)
46 }

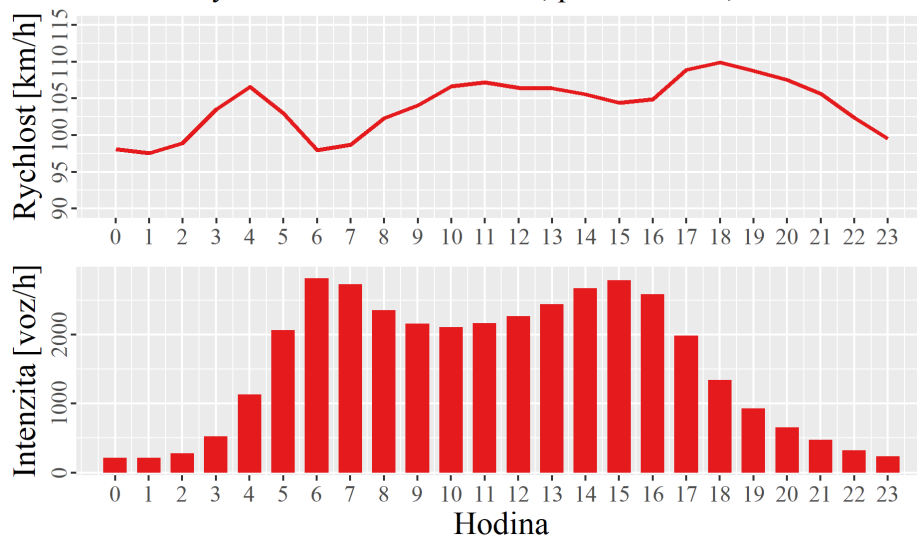
```

8.2 Intenzity a rychlosti

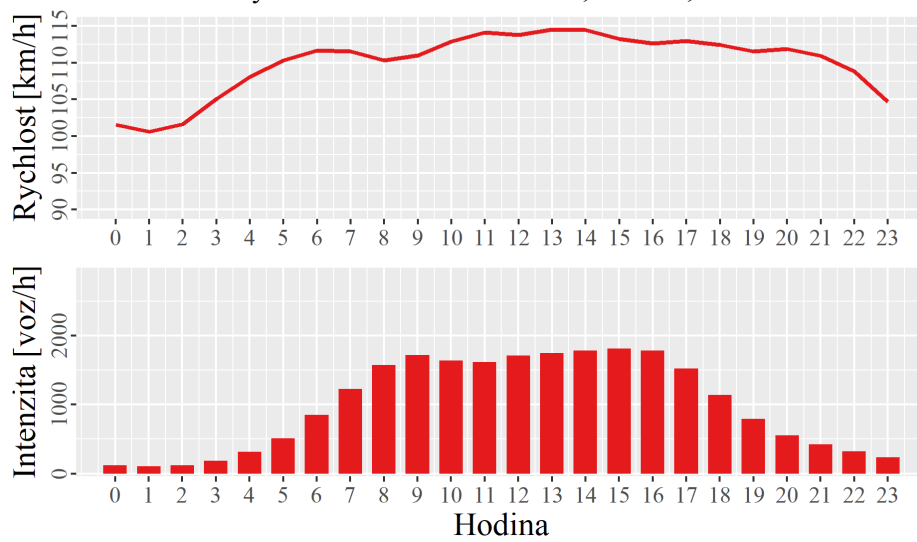
Vzhledem k tomu, že cílem práce je sledování ovlivněného dopravního proudu, bylo nutné před započítáním veškerých dalších prací ověřit, že se na měřených profilech taková doprava vyskytuje. Na následujících obrázcích jsou vykresleny průběhy rychlostí a intenzit pro všechny měřené profily, a to jak v pracovní den, tak o víkendu. Ve všech profilech se projevil výrazný pokles rychlostí v ranních hodinách ve všední den, indikující pravidelné kongesce. Mohli jsme tedy předpokládat, že měření v oblasti ovlivněného dopravního proudu bude v datech dostatek.



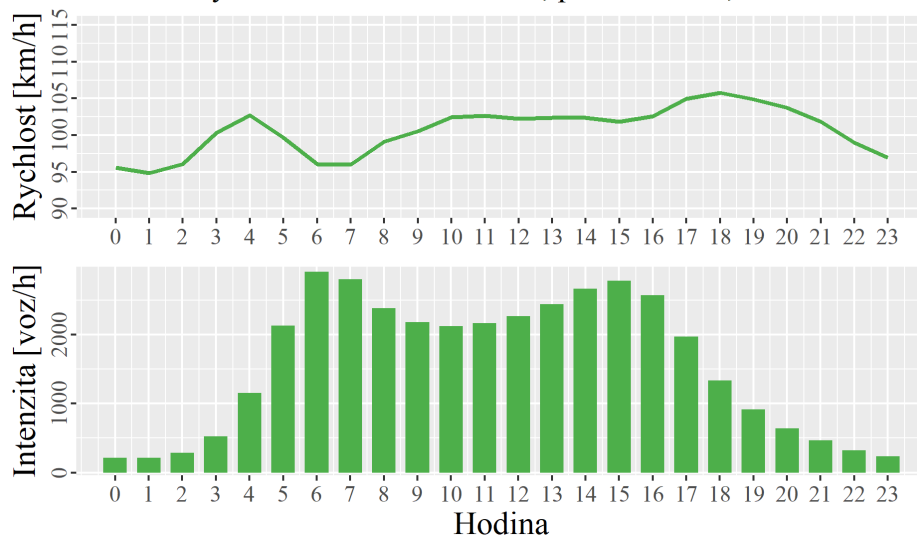
Průběh rychlosti a intenzit za 24 h, pracovní den, soubor dat 2



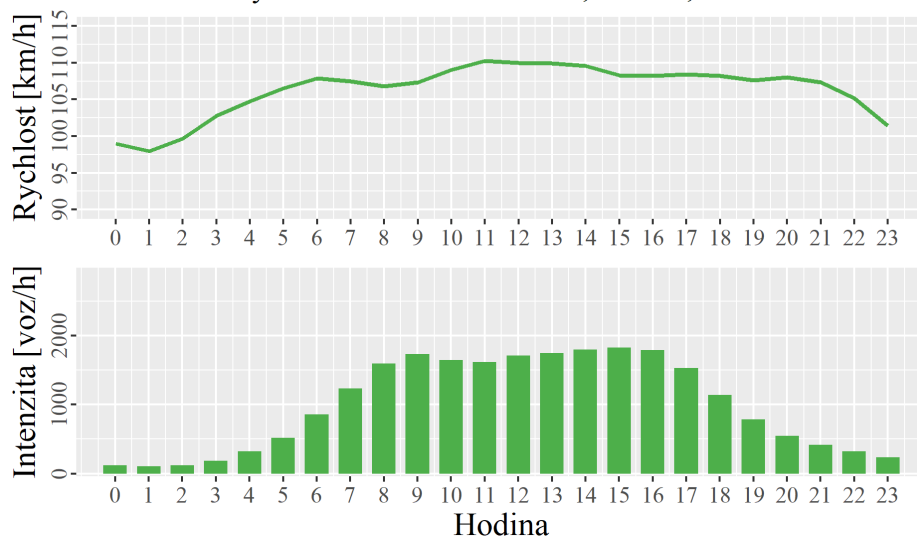
Průběh rychlosti a intenzit za 24 h, víkend, soubor dat 2



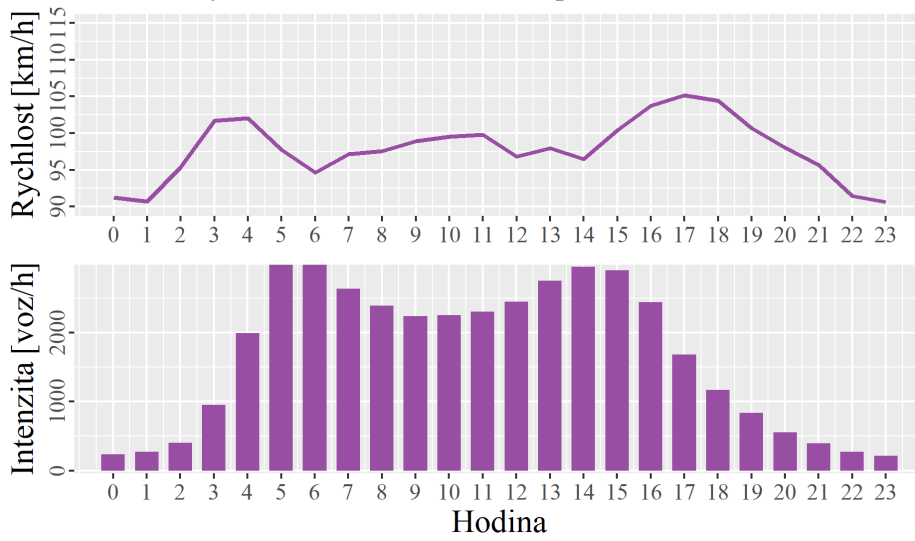
Průběh rychlosti a intenzit za 24 h, pracovní den, soubor dat 3



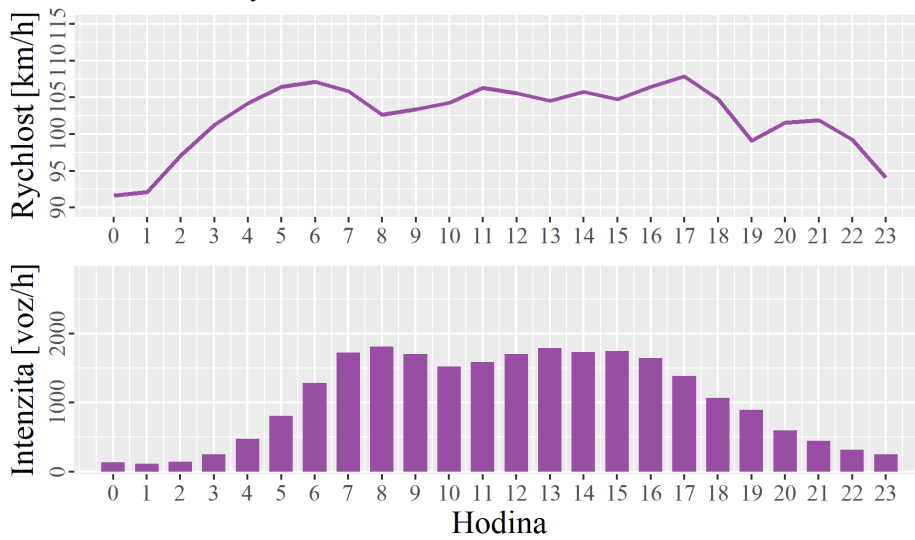
Průběh rychlosti a intenzit za 24 h, víkend, soubor dat 3



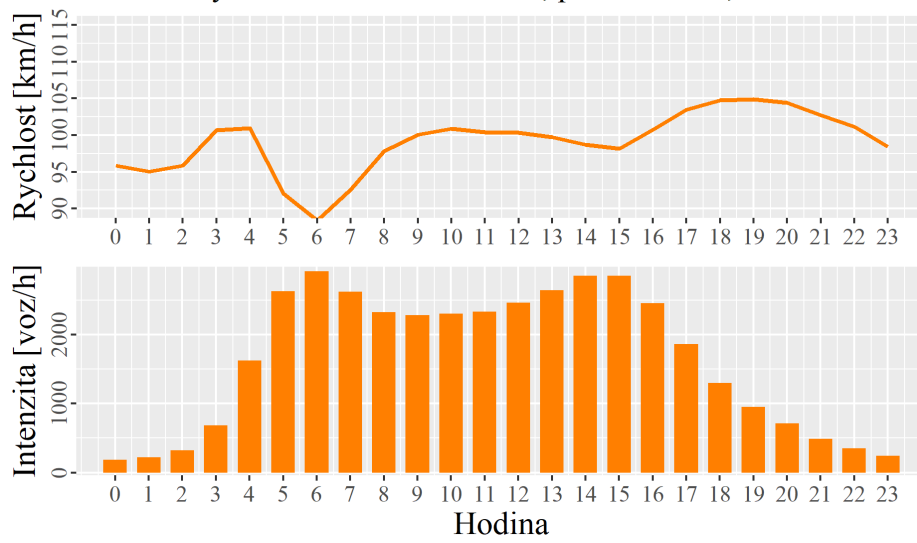
Průběh rychlosti a intenzit za 24 h, pracovní den, soubor dat 4



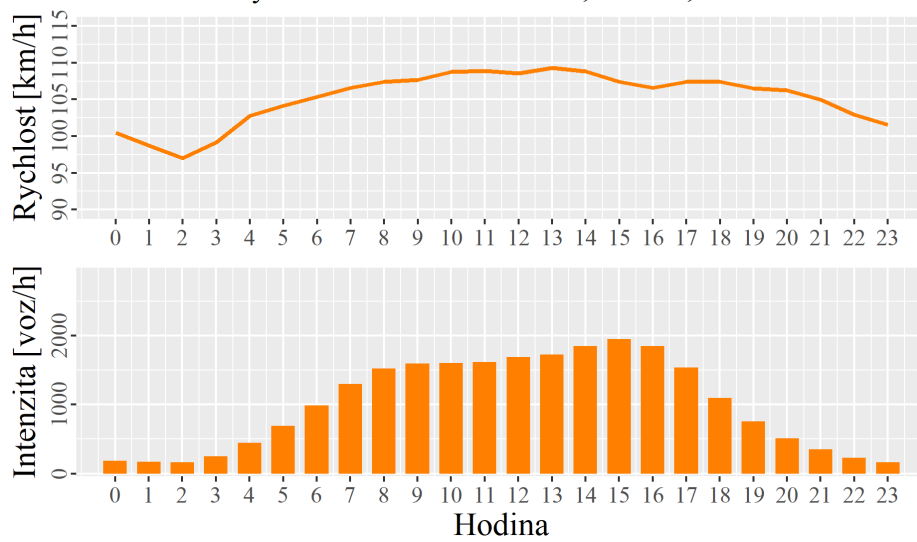
Průběh rychlosti a intenzit za 24 h, víkend, soubor dat 4



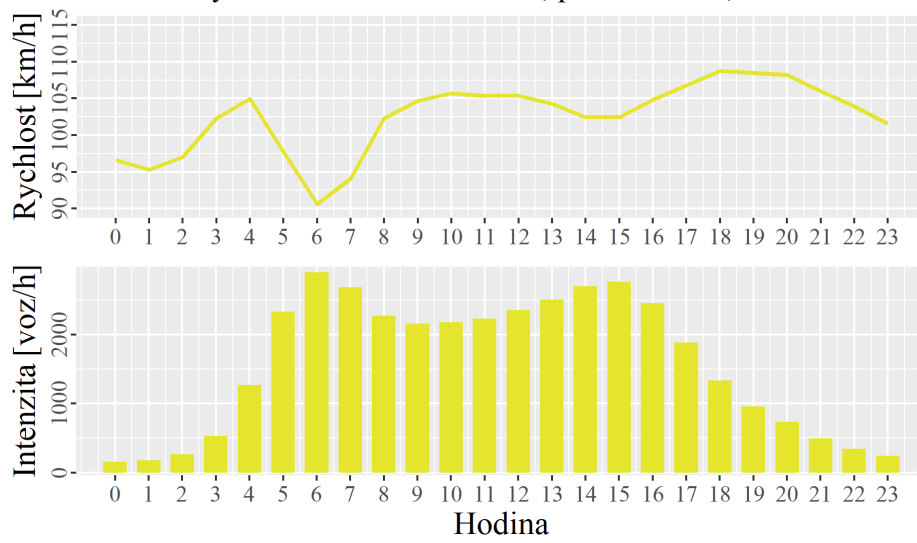
Průběh rychlosti a intenzit za 24 h, pracovní den, soubor dat 5



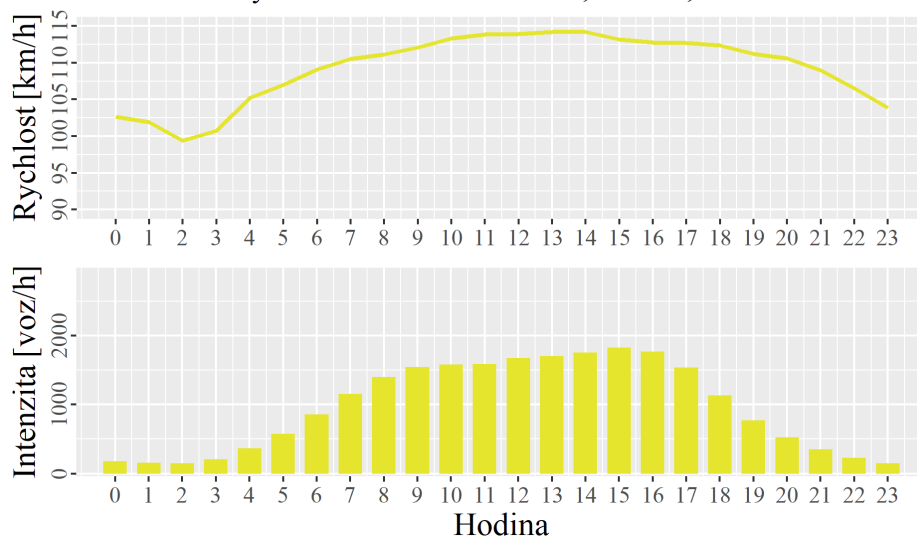
Průběh rychlosti a intenzit za 24 h, víkend, soubor dat 5



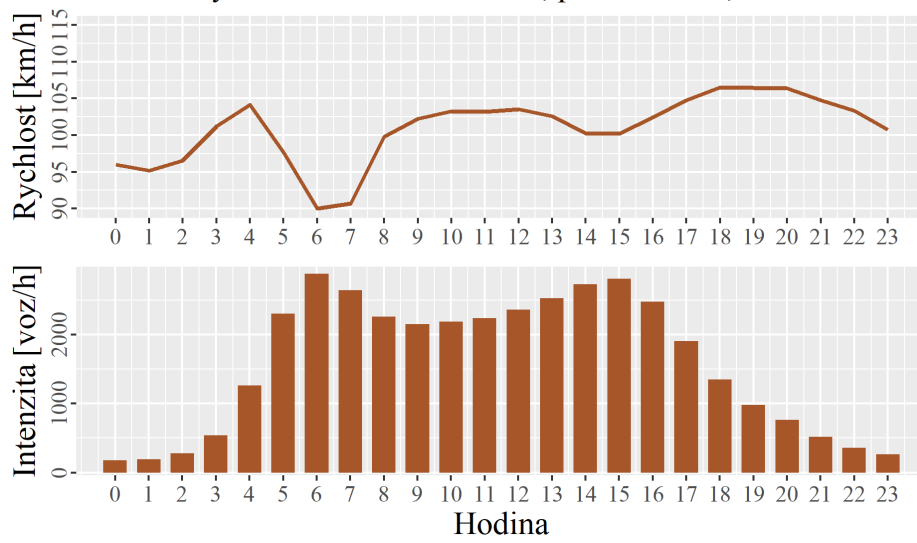
Průběh rychlosti a intenzit za 24 h, pracovní den, soubor dat 6



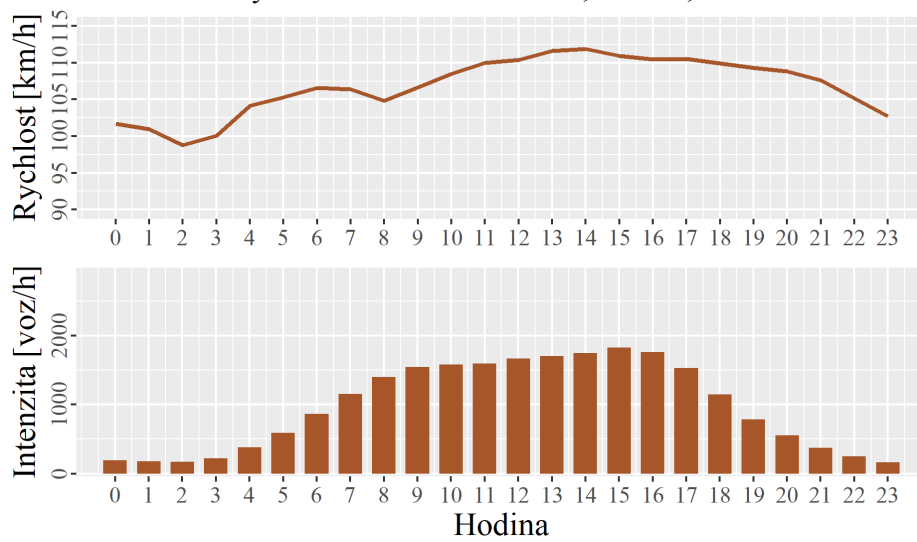
Průběh rychlosti a intenzit za 24 h, víkend, soubor dat 6



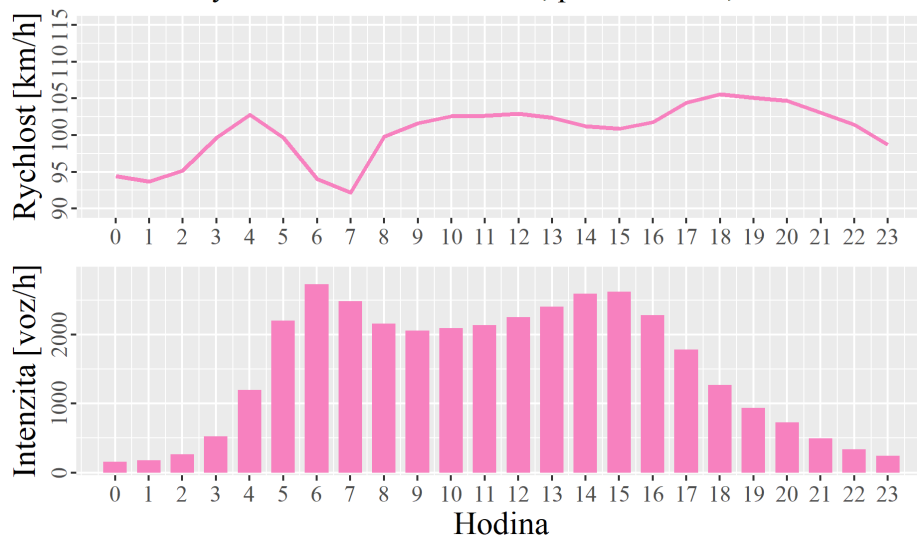
Průběh rychlosti a intenzit za 24 h, pracovní den, soubor dat 7



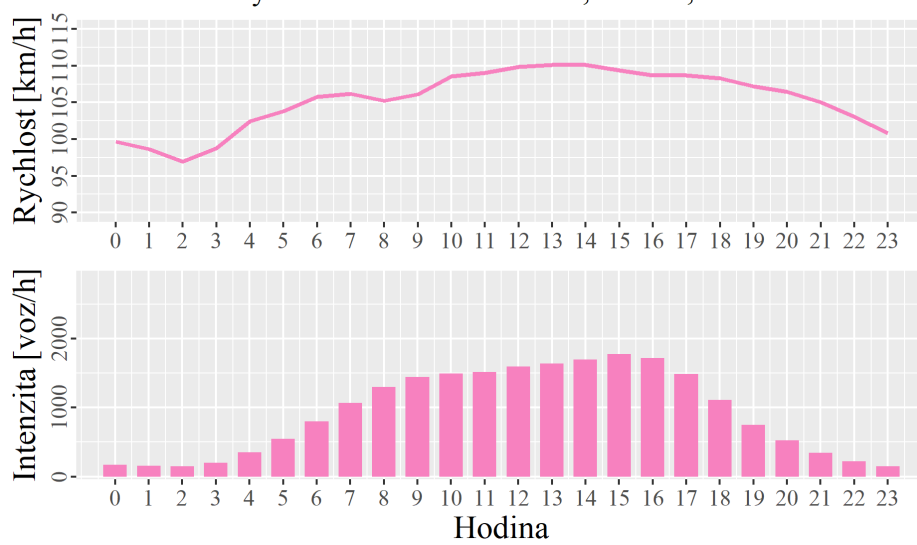
Průběh rychlosti a intenzit za 24 h, víkend, soubor dat 7



Průběh rychlosti a intenzit za 24 h, pracovní den, soubor dat 8



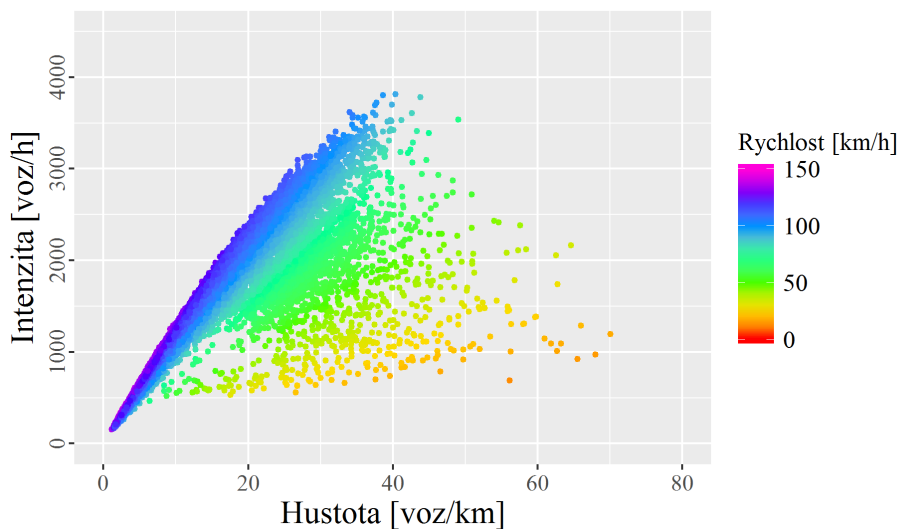
Průběh rychlosti a intenzit za 24 h, víkend, soubor dat 8



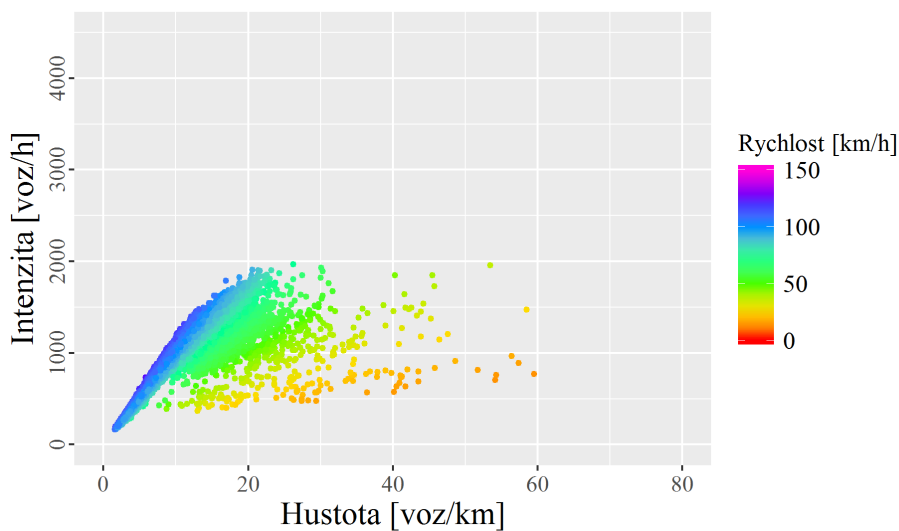
8.3 Fundamentální diagramy

Abychom si mohli udělat kompletní obrázek o charakteru dopravy na měřeném úseku, jsou v následující příloze uvedeny fundamentální diagramy pro jednotlivé jízdní pruhy a profily měření. Mimo jiné je z nich patrné, že vyjma souboru dat 4 jsou makroskopické charakteristiky dopravního proudu ve všech datových souborech velice podobné.

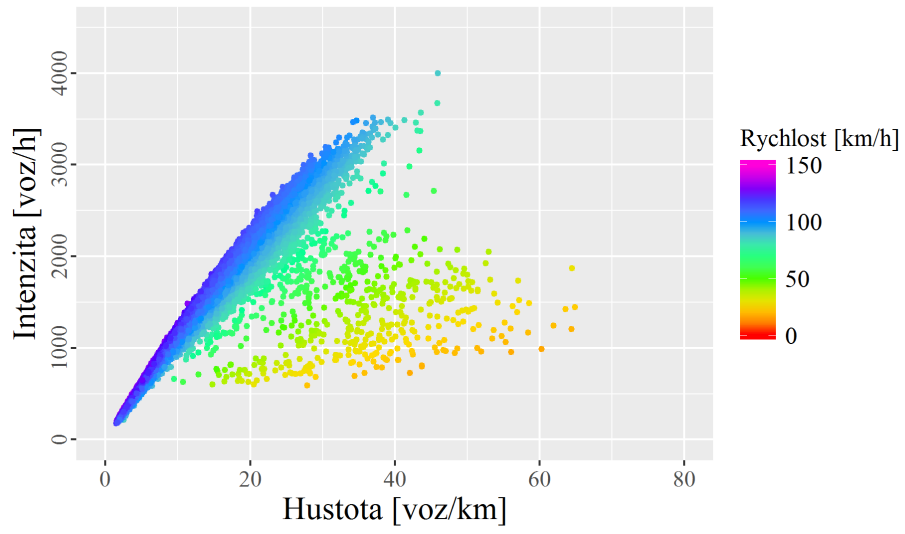
Fundamentální diagram, soubor dat 2, pruh 0



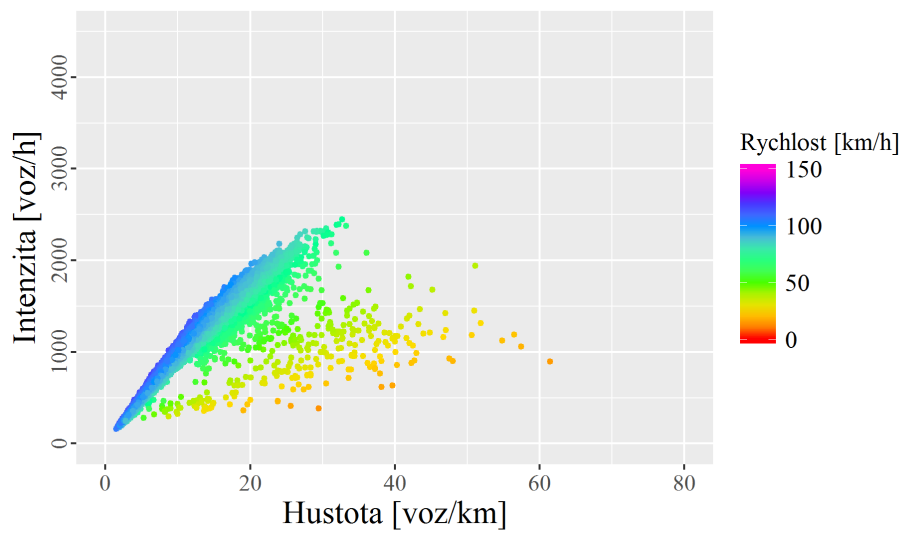
Fundamentální diagram, soubor dat 2, pruh 1



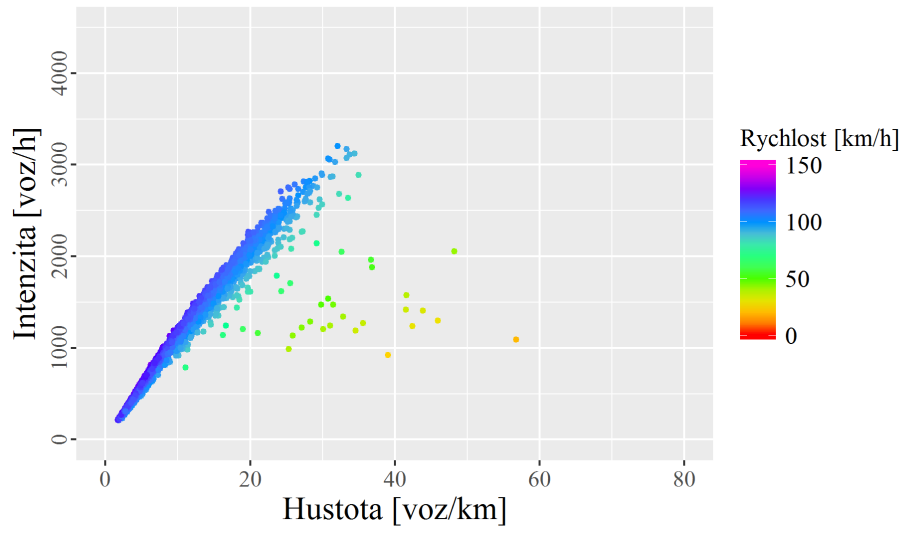
Fundamentální diagram, soubor dat 3, pruh 0



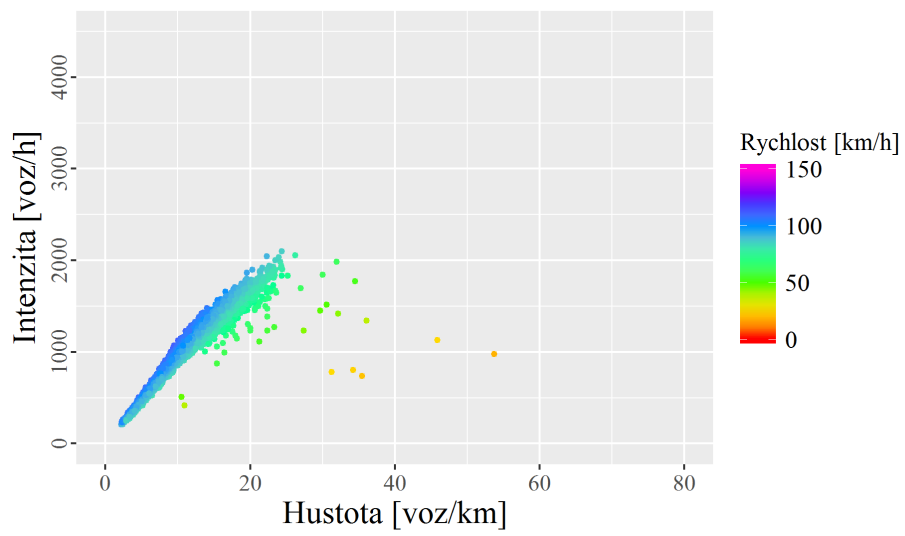
Fundamentální diagram, soubor dat 3, pruh 1



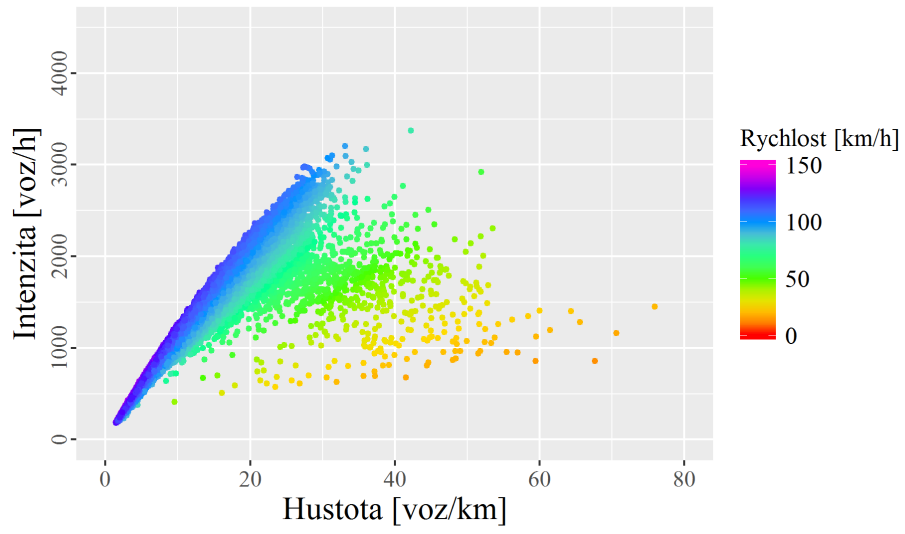
Fundamentální diagram, soubor dat 4, pruh 0



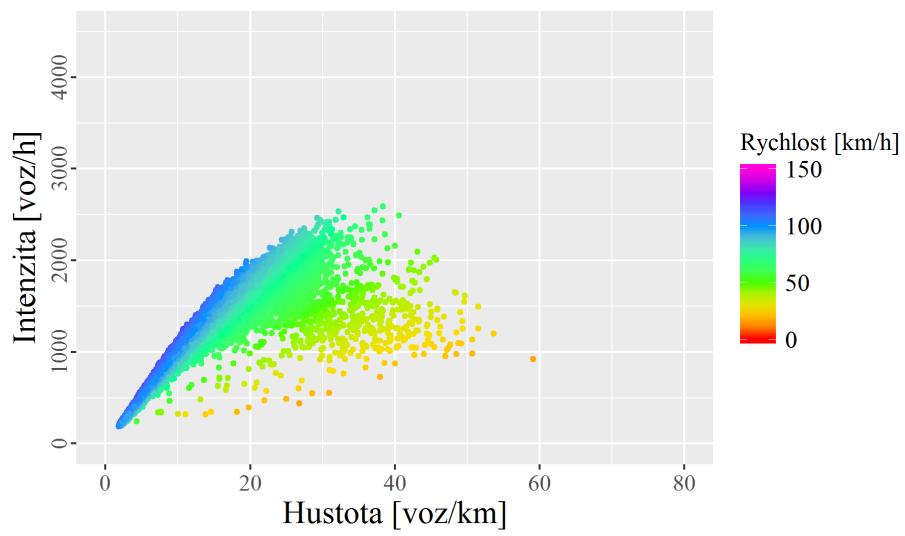
Fundamentální diagram, soubor dat 4, pruh 1



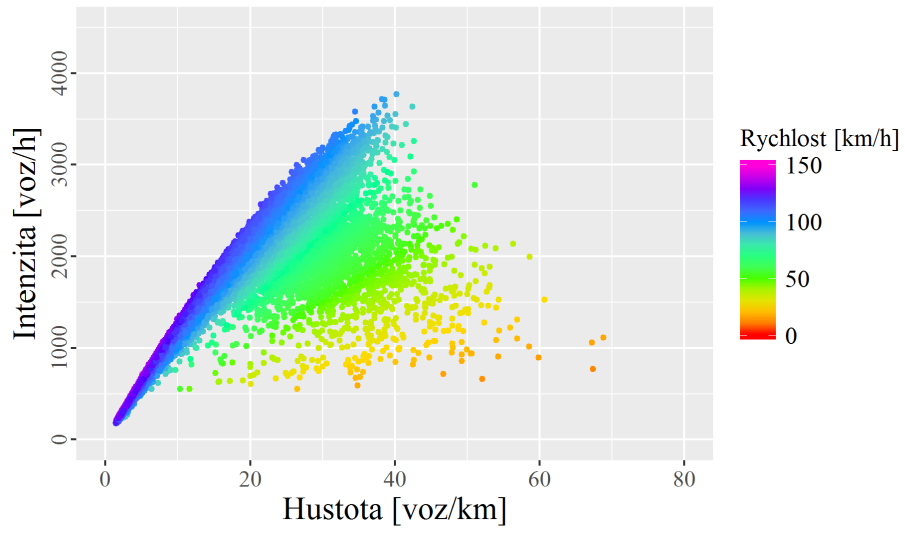
Fundamentální diagram, soubor dat 5, pruh 0



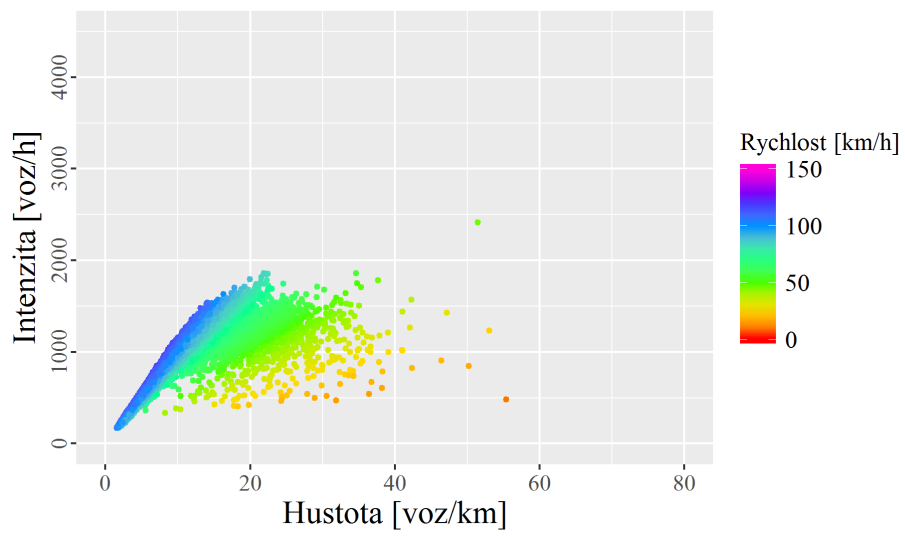
Fundamentální diagram, soubor dat 5, pruh 1



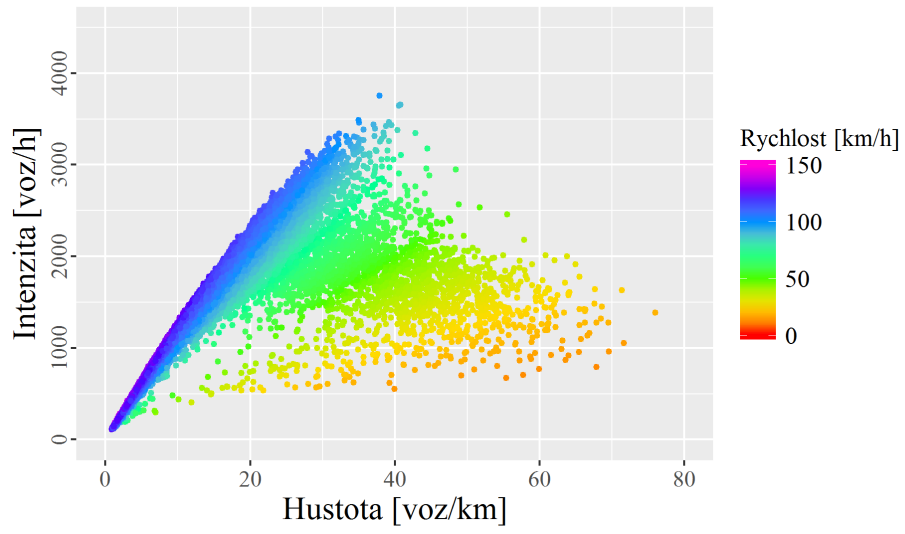
Fundamentální diagram, soubor dat 6, pruh 0



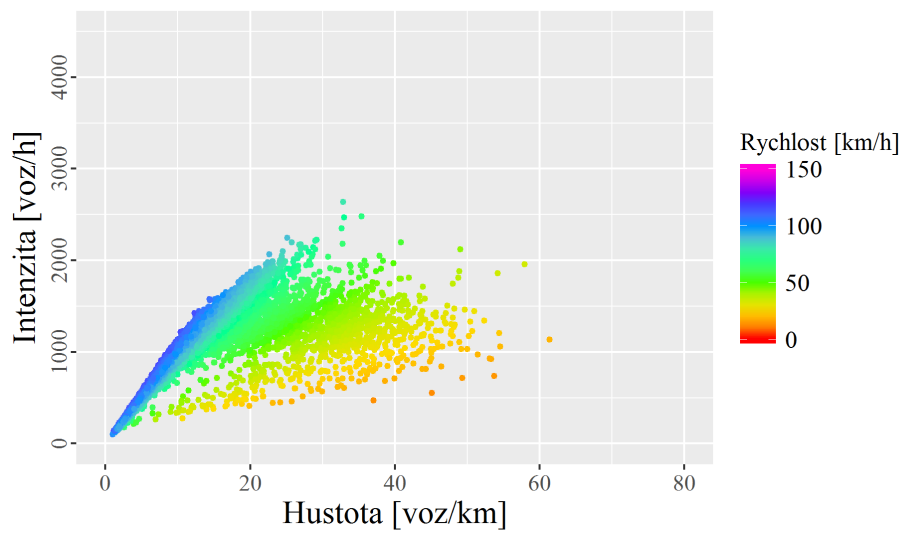
Fundamentální diagram, soubor dat 6, pruh 1



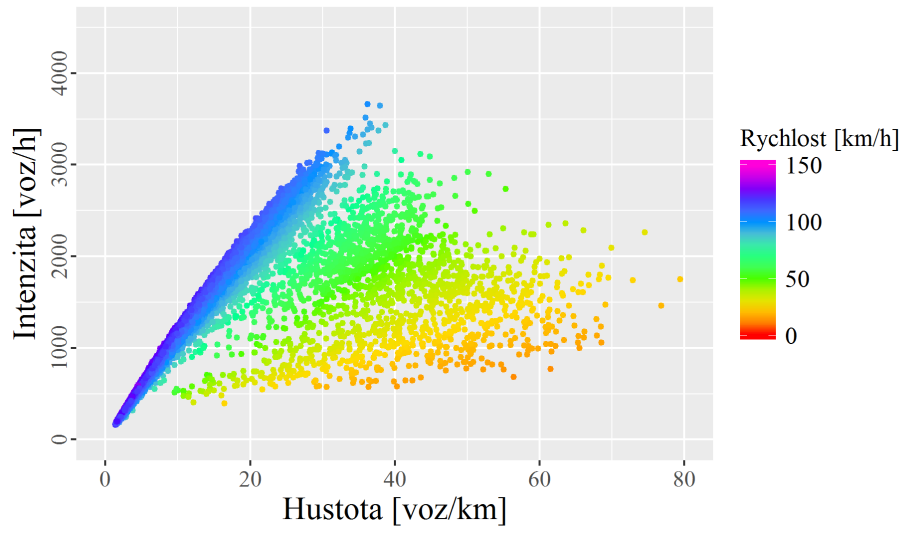
Fundamentální diagram, soubor dat 7, pruh 0



Fundamentální diagram, soubor dat 7, pruh 1



Fundamentální diagram, soubor dat 8, pruh 0



Fundamentální diagram, soubor dat 8, pruh 1

