

UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA

DIPLOMOVÁ PRÁCE

Analýza vlivu dopravních opatření na rychlost
vozidel v obcích



Katedra matematické analýzy a aplikací matematiky

Vedoucí práce: **doc. RNDr. Eva Fišerová, Ph.D.**

Vypracoval(a): **Bc. Kristýna Dadáková**

Studijní program: N0541A170026 Aplikovaná matematika

Studijní obor: Aplikovaná matematika

Forma studia: prezenční

Rok odevzdání: 2024

BIBLIOGRAFICKÁ IDENTIFIKACE

Autor: Bc. Kristýna Dadáková

Název práce: Analýza vlivu dopravních opatření na rychlost vozidel v obcích

Typ práce: Diplomová práce

Pracoviště: Katedra matematické analýzy a aplikací matematiky

Vedoucí práce: doc. RNDr. Eva Fišerová, Ph.D.

Rok obhajoby práce: 2024

Abstrakt: Cílem diplomové práce je analýza efektu dopravních opatření ve dvou španělských obcích s využitím vybraných metod analýzy funkcionálních dat. Prostorová data z plovoucích vozidel obsahující průjezdy obcemi jsou vnímána jako hladké funkce. V teoretické části práce jsou představeny základní pojmy a charakteristiky analýzy funkcionálních dat, metody vyhlazování a statistické testy vhodné pro analýzu tohoto reálného problému. Klíčovou částí práce je praktická část, ve které jsou teoretické znalosti aplikovány při řešení této úlohy, spočívající ve statistickém vyhodnocení významnosti efektu dopravních opatření v obci. Významný efekt dopravních opatření je dále detailněji analyzován s cílem odhalit, ve kterých konkrétních úsecích ke změně rychlosti dochází. V závěru diplomové práce je diskutován též problém sezónnosti dat z plovoucích vozidel a jeho možný vliv na posouzení efektu dopravních opatření.

Klíčová slova: analýza funkcionálních dat, plovoucí vozidla, efekt dopravních opatření, vyhlazování, intervalový test, sezónnost

Počet stran: 75

Počet příloh: 1

Jazyk: český

BIBLIOGRAPHICAL IDENTIFICATION

Author: Bc. Kristýna Dadáková

Title: Analysis of the impact of traffic measures on vehicle speeds in municipalities

Type of thesis: Master's

Department: Department of Mathematical Analysis and Application of Mathematics

Supervisor: doc. RNDr. Eva Fišerová, Ph.D.

The year of presentation: 2024

Abstract: The aim of this thesis is to analyze the effect of traffic measures in two Spanish municipalities using selected methods of functional data analysis. Spatial data from floating cars, containing passages through the municipalities, are viewed as smooth functions. The theoretical part of the thesis presents the basic concepts and characteristics of functional data analysis, smoothing methods, and statistical tests suitable for the analysis of this real problem. The key part of the thesis is the practical section, in which theoretical knowledge is applied to the solution of this problem. This problem involves the statistical evaluation of the significance of the effect of traffic measures in a municipality. The significant effect of the traffic measures is further analyzed in more detail to reveal in which specific sections the change in speed occurs. The issue of seasonality in floating car data and its possible influence on the assessment of the effect of traffic measures is discussed at the end of the thesis.

Key words: functional data analysis, floating cars, effect of traffic measures, smoothing, interval testing, seasonality

Number of pages: 75

Number of appendices: 1

Language: Czech

Prohlášení

Prohlašuji, že jsem diplomovou práci zpracovala samostatně pod vedením paní doc. RNDr. Evy Fišerové, Ph.D. a všechny použité zdroje jsem uvedla v seznamu literatury.

V Olomouci dne

.....

podpis

Obsah

Úvod	7
1 Úvod do analýzy funkcionálních dat	9
1.1 Motivační příklad	10
1.2 Funkcionální náhodná veličina a náhodný výběr	11
1.3 Výběrové charakteristiky a derivace	13
2 Metody vyhlazování	16
2.1 Bázová reprezentace funkcí	16
2.1.1 B -splajnová báze	17
2.1.2 Fourierova báze	19
2.2 Vyhlazování pomocí lineární regrese	19
2.3 Vyhlazování metodou regularizace	21
2.3.1 Volba vyhlazovacího parametru	24
3 Testy shody dvou funkcionálních výběrů	26
3.1 Globální test shody dvou funkcí středních hodnot	27
3.2 Intervalový test shody dvou funkcí průměru	29
4 Analýza efektu dopravních opatření	38
4.1 Představení dat a analyzovaných dopravních opatření	39
4.2 Pre-processing FCD	41
4.3 FCD jakožto funkcionální data	47
4.4 Analýza efektu instalovaných opatření	54
4.5 Problematika sezónnosti FCD	61
Závěr	70
Literatura	72
Seznam kódů v příloze	75

Poděkování

Ráda bych poděkovala své vedoucí práce, doc. RNDr. Evě Fišerové, Ph.D., za cenné rady, připomínky a za čas, který mi při konzultování práce věnovala. Dále patří mé poděkování Centru dopravního výzkumu, v. v. i. a společnosti Xouba Ingeniería, S.L. za poskytnutí dat a spolupráci. V neposlední řadě děkuji své rodině za podporu a trpělivost.

Úvod

Nepřiměřená rychlost vozidel je jedním z nejčastějších příčin dopravních nehod. Policejní prezidium České republiky eviduje jen za období leden až červen 2023 celkem 5334 nehod právě v důsledku nepřiměřené rychlosti. Jedná se o téměř 15 % všech dopravních nehod v tomto období. Neustále se proto rozšiřuje síť svislých a vodorovných dopravních značení, zpomalovacích prahů a dalších opatření s cílem snížit rychlost řidičů zejména v obcích. V rámci této problematiky hraje analýza dopravních opatření a účinnost jejich implementace klíčovou roli. Dochází ke snížení rychlosti vozidel v důsledku instalace opatření? Jsou instalovaná opatření na daných úsecích dostatečná? Právě těmito otázkami se zabývá tato diplomová práce. Cílem práce je tak představit vybrané metody analýzy funkcionálních dat, které mohou být pro analýzu vlivu dopravních opatření na rychlost vozidel využity. V teoretické části práce jsou představeny klíčové pojmy a metody analýzy funkcionálních dat, jakožto i metody vyhlazování a statistické testy vhodné pro tuto úlohu.

Hlavním přínosem diplomové práce je praktická část, ve které jsou teoretické poznatky aplikovány při řešení úlohy analýzy vlivu instalovaných opatření na rychlost vozidel v obcích Agolada a Cuntis ve Španělsku. Prostorová data z plovoucích vozidel obsahující průjezdy těmito obcemi jsou vnímána jako hladké funkce. Data jsou představena a vhodně vizualizována. Statistické vyhodnocení významnosti efektu dopravních opatření je zde klíčovým prvkem, přičemž se dále detailně analyzuje, ve kterých lokalitách úseku dochází k signifikantní změně rych-

losti. Toto nám může pomoci odpovědět na otázky, která opatření jsou účinná, která nikoliv, a v jakých oblastech úseku by bylo potřeba opatření rozšířit.

V závěru je diskutována také problematika sezónnosti dat z plovoucích vozidel, její vliv na hodnocení efektu dopravních opatření a navrženy možné strategie pro její zohlednění v praxi. Jsou zde ve stručnosti nastíněna další úskalí, jež mohou analýzu dat z plovoucích vozidel provázet.

Kapitola 1

Úvod do analýzy funkcionálních dat

Při datové analýze často pracujeme s mnohorozměrnými daty, kdy každému pozorovanému subjektu přísluší určitý konečný počet hodnot sledovaných veličin. Analýza funkcionálních dat (zkráceně FDA) představuje další odvětví analýzy dat pro data speciálního typu. Daty jako takovými jsou totiž funkce, nekonečně dimenzionální objekty, které získáme z replikovaných měření téhož procesu [11]. Procesem rozumíme v tomto kontextu sledování závislosti jedné veličiny na druhé (nejčastěji na čase, ale i například na věku, vzdálenosti apod.). Hodnoty provedeného měření tedy nechápeme jako posloupnost samostatných pozorování, nýbrž jako celek, funkci.

Typickými vlastnostmi funkcionálních dat je frekvence (velká hustota záznamu), kvantita (velký počet funkcionálních pozorování), podobný trend a v neposlední řadě hladkost [11]. Získaná data většinou zprvu hladká nejsou, a proto musíme využít metod vyhlazování funkcí, které si představíme v druhé kapitole.

V současné době se metody FDA stávají stále populárnějšími především díky neustále se rozvíjejícím technickým možnostem sběru dat. S funkcionálními daty se tak můžeme setkat opravdu v řadě odvětvích jako například v medicíně, biologii, meteorologii nebo dopravě.

V této kapitole si představíme motivační příklad, který ukáže, jak mohou

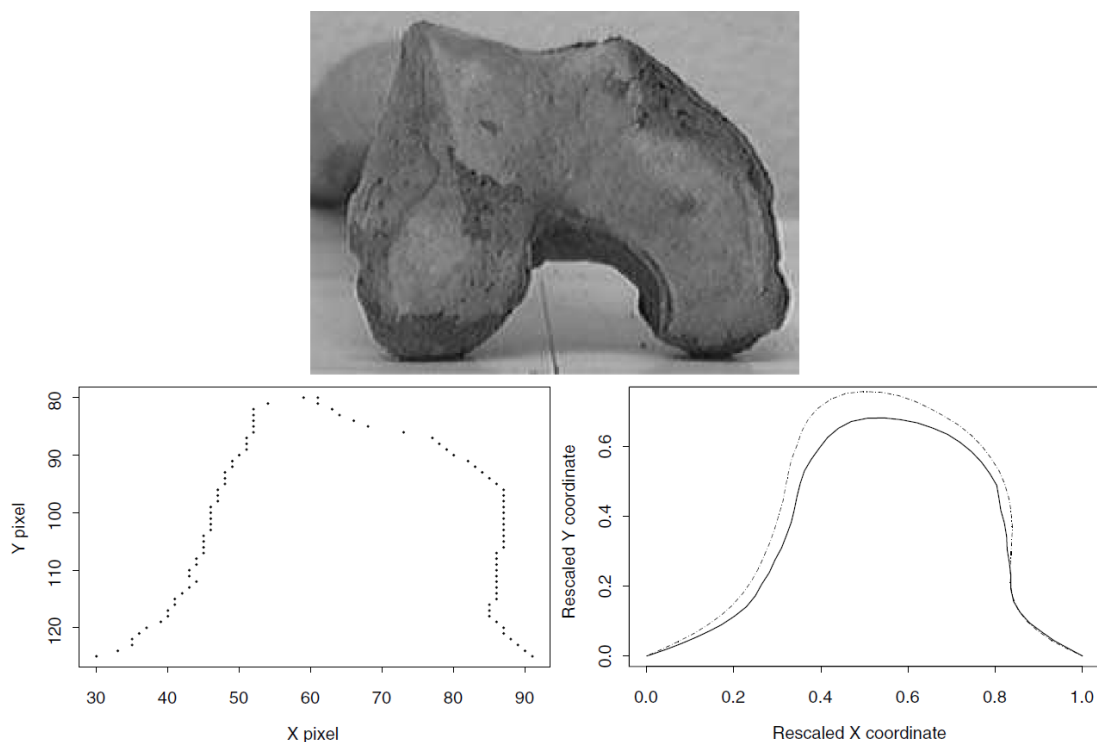
funkcionální data vypadat, a definujeme si základní pojmy a charakteristiky.

1.1. Motivační příklad

Jelikož funkce je nekonečně dimenzionální objekt, bylo potřeba zavést rozšíření pojmu náhodná veličina. Než však přejdeme k samotnému zavedení funkcionální náhodné veličiny, ukážeme si krátký motivační příklad, jenž vychází z Ramsaye a Silvermana (2002). S funkcionálními daty se lze setkat i v oblastech, kde bychom je na první pohled nečekali.

Mezi typické příklady funkcionálních dat patří například růstové křivky modelující výšku mladých chlapců v závislosti na věku, křivky průměrných teplot naměřených různými meteorologickými stanicemi v jednotlivých měsících či měření srdečního tepu sportovců při zátěžovém testu [20]. Funkcionální data však můžeme využít i pro modelování konkrétního útvaru, například rukopisu nebo i kosti.

Cílem jedné paleopatologické studie bylo zkoumat artrózu kolene z exhumovaných koster jedinců žijících na severu Anglie v letech 1000 – 1500 n. l. Bodový graf na obrázku 1.1 zachycuje útvar části kolenního kloubu vybraného jedince. Při vynášení bodů (pixelů) přitom využili pořízené digitální fotografie. Celkem archeologové do výzkumu zařadili 16 stehenních kostí se známkou kolenní artrózy a 52 kontrolních stehenních kostí. Vyhlazením bodových křivek vznikla funkcionální data, z nichž byly mimo jiné vypočteny průměrné útvary pro jedince s artrózou a bez artrózy kolene, viz obrázek 1.1.



Obrázek 1.1: Nahoře: digitální fotografie stehenní kosti vybraného jedince. Vlevo dole: originální data kloubního oblouku. Vpravo dole: průměrné křivky u jedinců s artrózou (plně), bez artrózy (čárkovaně). Zdroj: [19].

1.2. Funkcionální náhodná veličina a náhodný výběr

Zatím jsme funkcionální data vysvětlili pouze neformálně. Nyní již přejdeme k formálnímu zavedení základních pojmů FDA, tj. funkcionální náhodná veličina a funkcionální náhodný výběr.

Začneme se zavedením pojmu funkcionální náhodná veličina. Představme si náhodnou veličinu, kterou sledujeme v J_n různých po sobě jdoucích časových okamžicích z intervalu $[T_1, T_2]$. Takto danou náhodnou veličinu sledujeme celkem N -krát [10]. Již bylo zmíněno, že se na to můžeme dívat jako na replikace daného

procesu. Získáváme reálné hodnoty

$$x_n(t_{j,n}), \quad t_{j,n} \in [T_1, T_2], \quad n = 1, \dots, N, \quad j = 1, \dots, J_n. \quad (1.1)$$

Je tedy zřejmé, že obecně můžeme proces sledovat v různých časových okamžicích, kterých může být i různý počet pro každou jeho replikaci.

Všimněme si, že data (1.1) obsahují konečný počet bodů, ale v rámci FDA pracujeme s nekonečně dimenzionálními objekty. V praxi však nikdy nebudeme schopni získat měření ve všech bodech intervalu $[T_1, T_2]$. Z těchto bodů proto pomocí vhodných metod vytvoříme hladké křivky dostatečně reprezentující získaná data. Každá replikace procesu tak bude reprezentována konkrétní hladkou křivkou s nekonečně mnoha body, což již chápeme jako realizaci sledované funkcionální náhodné veličiny [10]. Z Ferratyho a Viea (2006) vychází následující definice 1.1.

Definice 1.1. *Náhodná veličina X se nazývá funkcionální náhodná veličina, jestliže nabývá hodnot v nekonečně rozměrném prostoru. Realizací x funkcionální náhodné veličiny X nazýváme funkcionální data.*

Je třeba si uvědomit, že pojem funkcionální náhodná veličina zaštiťuje všechny matematické objekty nabývající nekonečně mnoha hodnot. Hladké křivky jsou pouze speciálním případem pro $[T_1, T_2] \subset \mathbb{R}$, obecně můžeme uvažovat i složitější matematické objekty [5]. V tomto textu se však omezíme pouze na hladké křivky.

Nyní už můžeme snadno definovat funkcionální náhodný výběr. Ferraty a Vieu (2006) uvádějí následující definici 1.2.

Definice 1.2. *Funkcionálním náhodným výběrem (funkcionálními daty) x_1, \dots, x_N rozumíme N -tici realizací stejně rozdělených funkcionálních náhodných veličin X_1, \dots, X_n .*

Konečně formální zápis funkcionálních dat je potom

$$\{x_n(t), \quad t \in [T_1, T_2] \subset \mathbb{R}, \quad n = 1, \dots, N\},$$

kde hodnoty $x_n(t)$ již existují v každém bodě intervalu $[T_1, T_2]$ [10].

1.3. Výběrové charakteristiky a derivace

V metodách FDA budeme často potřebovat odhady teoretických charakteristik polohy a variability. Tyto funkcionální případy jsou přímým rozšířením charakteristik, se kterými pracujeme u mnohorozměrných dat.

Důležitou charakteristikou polohy je, jak je známo, střední hodnota. V mnohorozměrném případě se jedná o konstantu, kolem níž se data koncentrují. V případě funkcionálním je tomu tak rovněž s tím rozdílem, že se jedná o funkci. Empirickým odhadem tedy bude také funkce – funkce průměru. Matematicky jde pouze o bodově spočítaný aritmetický průměr přes všechny funkce [20], tedy

$$\bar{x}(t) = \frac{1}{N} \sum_{i=1}^N x_i(t), \quad t \in [T_1, T_2].$$

Podobně můžeme zavést také charakteristiky variability [20]. Míru variability dat reprezentuje funkce rozptylu, resp. funkce směrodatné odchylky, již odhadujeme jako

$$\widehat{\text{var}}_X(t) = \frac{1}{N-1} \sum_{i=1}^N [x_i(t) - \bar{x}(t)]^2, \quad t \in [T_1, T_2],$$

resp.

$$\widehat{\text{sd}}_X(t) = \sqrt{\widehat{\text{var}}_X(t)}.$$

Pro popisnou statistiku využíváme grafickou vizualizaci míry variability spolu s funkcí průměru na datech. Jde o pás kolem funkce průměru, který bodově vypočteme jako součet, resp. rozdíl, dvojnásobku směrodatné odchylky a průměru [10]. Ze základního kurzu statistiky víme, že řídí-li se data normálním rozdělením, zhruba 95 % dat se nachází uvnitř takto zkonstruovaného intervalu. Normalitou v FDA rozumíme mnohorozměrnou normalitu náhodných veličin v jednotlivých měřených bodech [7]. V případě funkcionálních dat se však s normálně rozděle-

nými daty v praxi často nesetkáme (přesto však v řadě metod FDA je normalita předpokladem). Zmíněnou charakteristiku budeme tedy využívat spíše jako grafický ukazatel variability dat.

Někdy se nám můžou hodit i kovarianční, resp. korelační, funkce [10]. Výběrová kovarianční funkce je dána předpisem

$$\hat{c}(t, s) = \frac{1}{N-1} \sum_{i=1}^N [x_i(t) - \bar{x}(t)][x_i(s) - \bar{x}(s)], \quad \forall t, s \in [T_1, T_2],$$

přičemž velké hodnoty $\hat{c}(t, s)$ znamenají významnou kovarianci hodnot funkcionální náhodné veličiny X v bodech t a s . Nutno zdůraznit, že kovarianční funkce je již funkcí dvou proměnných. Výběrovou korelační funkci potom vypočteme jako

$$\hat{r}(t, s) = \frac{\hat{c}(t, s)}{\widehat{\text{sd}}_X(t) \cdot \widehat{\text{sd}}_X(s)}, \quad \forall t, s \in [T_1, T_2].$$

Nejen pro popis funkcionálních dat budeme využívat také jejich derivace [18]. Obecně m -tá derivace funkcionální náhodné veličiny X je dána

$$D^m x_n(t) = \frac{d^m x_n(t)}{dt^m}, \quad n = 1, \dots, N.$$

Často se v FDA setkáváme s druhou derivací $D^2 x_n(t)$. Z hlediska popisné statistiky ji interpretujeme jako zrychlení (akceleraci). Lze tak analyzovat chování křivek co se týče rychlosti růstu nebo poklesu [18]. Obecně kladná akcelerace značí zrychlení a negativní zpomalení. Konkrétně je-li pozitivní akcelerace rostoucí, zrychlení průběhu funkce se zintenzivňuje. Jestliže je pozitivní akcelerace klesající, pak zrychlení průběhu funkce slábne. Analogicky je tomu tak u negativní akcelerace. Je-li negativní akcelerace rostoucí, zpomalení průběhu funkce se zintenzivňuje a je-li klesající, zpomalení naopak slábne. V bodě nulové akcelerace, na jehož okolí je její průběh současně klesající, zaznamenáváme lokálně nejrychlejší průběh funkce, tj. její rychlost dosahuje v tomto bodě lokálního maxima.

V bodě nulové akcelerace, na jehož okolí je rostoucí, pozorujeme naopak lokálně nejpomalejší průběh (rychlost dosahuje v tomto bodě lokálního minima).

Zkonstruovat můžeme také například 95% bodový pás spolehlivosti pro funkci střední hodnoty. Je již známo, že funkce průměru je pouhým odhadem této teoretické charakteristiky polohy a konstrukcí pásu spolehlivosti dostaneme informaci o jistotě odhadu. Skutečná hodnota funkce střední hodnoty v nějakém bodě t_j náleží s 95% pravděpodobností intervalu, který získáme řezem pásu spolehlivosti v bodě t_j . Abychom se vyhnuli jistým předpokladům na distribuci dat, provedeme konstrukci bootstrapovou procedurou [9]. Předpokládáme-li, že máme N křivek evaluovaných ve stejných bodech, provedeme konstrukci 95% bodového pásu spolehlivosti následovně

1. Z N křivek provedeme bootstrapový výběr, tj. náhodný výběr s opakováním rozsahu N .
2. Spočítáme funkci průměru na novém výběru a označíme $\bar{x}_1^*(t)$.
3. Toto provedeme například 1000krát, čímž získáme $\bar{x}_1^*(t), \dots, \bar{x}_{1000}^*(t)$.
4. Bodově vypočítáme dolní hranici pásu spolehlivosti jako 2,5% kvantil z křivek $\bar{x}_1^*(t), \dots, \bar{x}_{1000}^*(t)$ a označíme $l(t)$.
5. Analogicky vypočítáme horní hranici pásu spolehlivosti jako 97,5% kvantil z křivek $\bar{x}_1^*(t), \dots, \bar{x}_{1000}^*(t)$ a označíme $u(t)$.
6. Odhad 95% bodového pásu spolehlivosti pro funkci střední hodnoty je potom

$$[l(t), u(t)], \quad \forall t \in [T_1, T_2].$$

Skutečná hodnota funkce střední hodnoty v bodě t_j s 95% pravděpodobností náleží intervalu $[l(t_j), u(t_j)]$.

Kapitola 2

Metody vyhlazování

V první kapitole již bylo řečeno, že získaná data jsou nejprve diskrétního charakteru. Abychom dostali funkcionální data, musíme tato diskrétní měření pomocí vyhlazovacích metod transformovat do funkcionální podoby. Uplatnění jedné z metod vyhlazování, které si v této kapitole představíme, je tedy vždy nutným krokem před samotnou analýzou funkcionálních dat.

Přístupů je hned několik. Mezi nejznámější se řadí vyhlazování pomocí lineární regrese a vyhlazování metodou regularizace [20]. K dalším potom patří například jádrové vyhlazování, které podrobněji naleznete v [20]. Pro přehlednost vynecháme v této kapitole indexové značení křivek a budeme používat pouze $x(t)$. Parametrem J potom budeme značit počet naměřených bodů funkce $x(t)$.

2.1. Bázová reprezentace funkcí

Protože potřebujeme umět konstruovat i funkce složitějšího průběhu, obě výše zmíněné metody vycházejí z bázové reprezentace funkcí, tzv. bázové expanze [20]. Křivku $x(t)$ hledáme jako lineární kombinaci daného počtu známých bázových

funkcí, jež jsou lineárně nezávislé. Bázovou expanzi funkce $x(t)$ píšeme jako

$$x(t) = \sum_{k=1}^K c_k \phi_k(t) = \boldsymbol{\phi}^\top(t) \mathbf{c}, \quad t \in [T_1, T_2],$$

přičemž $c_k \in \mathbb{R}$ jsou koeficienty bázových funkcí $\phi_k(t)$ a K je počet bázových funkcí [20]. V maticovém značení $\boldsymbol{\phi}(t) = (\phi_1(t), \dots, \phi_K(t))^\top$ a $\mathbf{c} = (c_1, \dots, c_K)^\top$.

Pro diskrétní body $x(t_j)$, $j = 1, \dots, J$, budeme používat taktéž maticový zápis

$$\mathbf{x} = \boldsymbol{\Phi} \mathbf{c},$$

kde matice $\boldsymbol{\Phi}$ je rozměru $J \times K$ obsahuje prvky $\{\boldsymbol{\Phi}\}_{jk} = \phi_k(t_j)$ a vektor $\mathbf{x} = (x(t_1), \dots, x(t_J))^\top$. Předpokládáme plnou řádkovou hodnotu matice $\boldsymbol{\Phi}$. Počet bázových funkcí K volíme tak, aby $K < J$. Pokud by $K = J$, přešli bychom k úloze interpolace. Zvoli-li bychom K příliš velké, výsledek by reflektoval chyby měření a došlo by k podhlazení. Pokud bychom jej zadali naopak příliš malé, potlačili bychom lokální vlastnosti křivky [20]. Pro konkrétní úlohu je tak potřeba zvolit určitý kompromis.

Jelikož se v případě funkcionálních dat křivky chovají podobně, mají podobný trend, typicky je vyhlazujeme všechny najednou a využíváme k tomu stejné bázové funkce. Často používanými bázemi jsou B -splajnové a Fourierovy, které si přiblížíme podrobněji níže.

Méně využívanou bází je báze monomiální, kde bázovými funkcemi jsou polynomy, tj. $\phi_k(t) = t^{k-1}$. S ohledem na výpočetní problémy a použitelnost pouze pro velmi malé K se však téměř nepoužívá [18].

2.1.1. B -splajnová báze

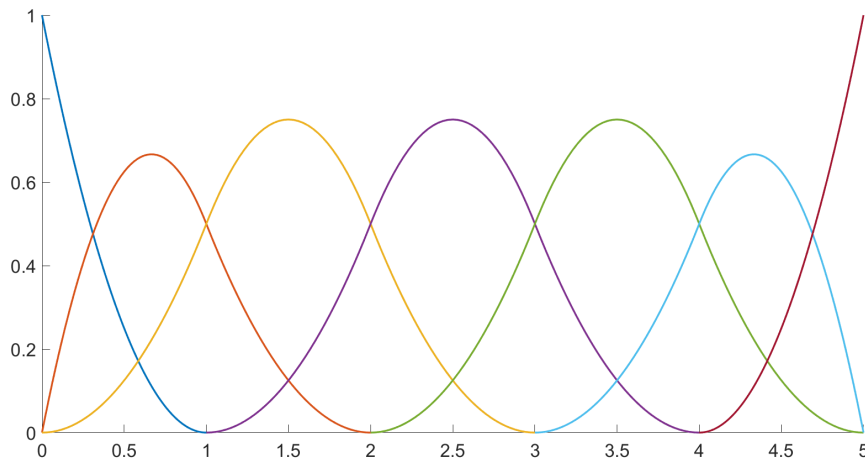
B -splajnová báze patří mezi vůbec ty nejpoužívanější pro vyhlazování neprerodických funkcí. Bázovými funkcemi jsou zde tzv. B -splajny, hladké, po částech polynomiálních funkce [4]. Uzly nazýváme body, v nichž se jednotlivé části splajnu

hladce pojí. Z původní definice dle de Boora (1978), byl pro B -splajn B_i^{k+1} stupně $k + 1$ s uzly $\lambda_i \leq \dots \leq \lambda_{i+k+1}$ odvozen následující rekurentní vztah

$$B_i^{k+1}(x) = \frac{x - \lambda_i}{\lambda_{i+k} - \lambda_i} B_i^k(x) + \frac{\lambda_{i+k+1} - x}{\lambda_{i+k+1} - \lambda_{i+1}} B_{i+1}^k(x), \quad (2.1)$$

$$B_i^1(x) = \begin{cases} 1, & \text{pro } x \in [\lambda_i, \lambda_{i+1}), \\ 0, & \text{pro } x \notin [\lambda_i, \lambda_{i+1}). \end{cases}$$

Výpočet dle vztahu (2.1) je preferován právě díky menší výpočetní náročnosti. Počet bázových funkcí K volíme jako $g + k + 1$, kde g je počet vnitřních uzlů (nepočítáme tedy uzly na krajích vyhlazovacího intervalu). Uzly volíme s ohledem na průběh funkce nebo jen ekvidistantně. Vyhlazování funkcí není vůči volbě uzlů invariantní [20].



Obrázek 2.1: Příklad báze z kubických B -splajnů s uzly $\{0, 1, 2, 3, 4, 5\}$.

Stručně si uvedeme některé vlastnosti B -splajnů. Na obrázku 2.1 je příklad báze s kubickými B -splajny (stupně 3) a uzly $\{0, 1, 2, 3, 4, 5\}$. Můžeme si všimnout, že bázové funkce jsou nezáporné a mají lokální nosič. Obecně má B -splajn B_i^{k+1} nosič na intervalu $[\lambda_i, \lambda_{i+k+1})$. Další podstatnou vlastností je spojitost derivací až do řádu $k - 1$ [4].

2.1.2. Fourierova báze

Speciálnějši bází je báze Fourierova, která se používá pro periodická data. Bázové funkce jsou ve tvaru

$$\{1, \sin(\omega t), \cos(\omega t), \sin(2\omega t), \cos(2\omega t), \dots, \sin(M\omega t), \cos(M\omega t)\},$$

kde $M \in \mathbb{N}$ a $\omega = \frac{2\pi}{T}$, přičemž T je perioda. Počet bázových funkcí K , který volíme, bude roven $2M + 1$, a bude tedy vždy lichý [11]. Bázovou expanzi funkce $x(t)$ můžeme potom zapsat jako

$$x(t) = c_1 + \sum_{j=1}^M c_{2j} \sin(\omega j t) + \sum_{j=1}^M c_{2j+1} \cos(\omega j t).$$

Fourierova báze poskytuje velmi dobré výpočetní vlastnosti. Její výpočet je nenáročný, stejně jako výpočet derivací bázových funkcí. V praxi se proto v některých aplikacích využívá i pro data neperiodická. Periodou T pak v této situaci uvažujeme celý interval pokrývající všechny hodnoty nezávisle proměnné, tj. $[T_1, T_2]$ [11].

2.2. Vyhlazování pomocí lineární regrese

Předpokládejme, že máme k dispozici data $y(t_j)$, pro $j = 1, \dots, J$. Poněvadž tyto diskrétní hodnoty jsou realizacemi sledované funkcionální náhodné veličiny, musíme brát v potaz nejistotu, se kterou se od hledané křivky $x(t)$ v daných bodech lišíme. Zavádíme proto model

$$y(t_j) = x(t_j) + \epsilon(t_j), \quad j = 1, \dots, J,$$

přičemž předpokládáme, že $\epsilon(t)$, $\forall t \in [T_1, T_2]$, je funkcí bílého šumu, tj. s nulovou střední hodnotou, rozptylem neměnným v čase a hodnotami na sobě nezávis-

lými [18].

Pomocí bázev expanze lze model přepsat následovně

$$y(t_j) = \sum_{k=1}^K c_k \phi_k(t_j) + \epsilon(t_j), \quad j = 1, \dots, J,$$

maticově

$$\mathbf{y} = \mathbf{\Phi} \mathbf{c} + \boldsymbol{\epsilon}. \quad (2.2)$$

Nyní využijeme model (2.2) pro vyhlazování metodou lineární regrese. Budeme hledat řešení $\mathbf{c} = (c_1, \dots, c_K)^\top$, jenž minimalizuje reziduální součet čtverců modelu. Právě aplikováním metody nejmenších čtverců na tento model dojde k jistému vyhlazení hodnot $x(t_j)$. V závislosti na \mathbf{c} tedy minimalizujeme

$$\text{SSE}(\mathbf{c}) = \sum_{j=1}^J \left[y(t_j) - \sum_{k=1}^K c_k \phi_k(t_j) \right]^2 = (\mathbf{y} - \mathbf{\Phi} \mathbf{c})^\top (\mathbf{y} - \mathbf{\Phi} \mathbf{c}), \quad (2.3)$$

kde řešením této úlohy, jak je známo z teorie lineárních modelů, je právě $\mathbf{c}^* = (\mathbf{\Phi}^\top \mathbf{\Phi})^{-1} \mathbf{\Phi}^\top \mathbf{y}$ [20]. Odhad vyhlazené funkce je potom

$$\hat{x}(t) = \boldsymbol{\phi}^\top(t) \mathbf{c}^* = \boldsymbol{\phi}^\top(t) (\mathbf{\Phi}^\top \mathbf{\Phi})^{-1} \mathbf{\Phi}^\top \mathbf{y}, \quad \forall t \in (T_1, T_2).$$

Předpokládáme-li, že funkci $\hat{x}(t)$ evaluujeme ve stejných bodech, ve kterých jsme daný proces sledovali, lze vyrovnané hodnoty psát pomocí tzv. vyhlazovací matice \mathbf{S} následovně

$$\hat{\mathbf{x}} = \mathbf{\Phi} \mathbf{c}^* = \mathbf{\Phi} (\mathbf{\Phi}^\top \mathbf{\Phi})^{-1} \mathbf{\Phi}^\top \mathbf{y} = \mathbf{S} \mathbf{y}.$$

Jelikož používáme klasickou metodu nejmenších čtverců, musíme dbát na splnění předpokladu, že rezidua modelu musí mít charakter bílého šumu. Vyskytují-li se ovšem v reziduiích problém heteroskedasticity, tj. jejich rozptyl není kon-

stantní, můžeme vyhlazování provést metodou vážených nejmenších čtverců

$$\text{WSSE}(\mathbf{c}) = \sum_{j=1}^J w_j^2 \left[y(t_j) - \sum_{k=1}^K c_k \phi_k(t_j) \right]^2 = (\mathbf{y} - \Phi \mathbf{c})^\top \mathbf{W}^\top \mathbf{W} (\mathbf{y} - \Phi \mathbf{c}),$$

s diagonální maticí \mathbf{W} obsahující příslušné váhy [11]. Řešením je

$$\mathbf{c}^* = (\Phi^\top \mathbf{W} \Phi)^{-1} \Phi^\top \mathbf{W} \mathbf{y}.$$

Vyhlazování pomocí lineární regrese je sice velmi jednoduché, avšak má své nevýhody. Míru vyhlazení nemůžeme nijak v tomto případě regulovat a mohli bychom tak získávat příliš flexibilní modely.

2.3. Vyhlazování metodou regularizace

Tato metoda nám umožňuje adaptovat vyhlazení dle aktuální potřeby. Stále požadujeme, aby hledaná funkce $x(t)$ naměřený proces dostatečně dobře reprezentovala, ale nyní chceme regulovat „hrbolatost“ funkce. Předpokládáme-li B -splajnovou bázi, mluvíme o tzv. vyhlazujícím splajnu [11]. Stejně jako v případě vyhlazování metodou nejmenších čtverců využijeme báze expanze funkce $x(t)$. Při vyhlazování metodou regularizace může být počet bázevých funkcí větší než bodů evaluace, tj. $K > J$ [11].

Hledáme bázevých koeficienty $\mathbf{c} = (c_1, \dots, c_K)^\top$, jenž minimalizují výraz obsahující, stejně jako v předchozí metodě, součet kvadrátů reziduí, ovšem doplněný o penalizační člen. Minimalizační úlohou je tedy tzv. penalizovaný součet čtverců reziduí

$$\text{PSSE}(\mathbf{c}, \lambda) = \text{SSE}(\mathbf{c}) + \lambda \int_{T_1}^{T_2} [D^2 x(s)]^2 ds, \quad (2.4)$$

s vyhlazovací konstantou λ , jež může nabývat libovolných nezáporných hodnot [10]. Vliv volby této konstanty na výslednou podobu vyhlazujícího splajnu vy-

světlné později.

Z výrazu (2.4) vidíme, že penalizační funkcí je zde právě druhá mocnina hledané funkce $x(t)$. Jedná se o nejpoužívanější tvar, obecně však můžeme použít mocninu libovolného řádu. Druhou mocninou penalizujeme křivost samotné funkce. Chceme-li analyzovat například také druhou derivaci funkce $x(t)$ tj. její akceleraci, použijeme $D^4x(t)$, čímž už zajistíme její větší hladkost [20].

Nyní hledejme řešení \mathbf{c}^{**} , jenž minimalizuje výraz (2.4) pro pevné λ . Nejprve vyjádříme penalizační výraz maticově takto

$$\begin{aligned} \int_{T_1}^{T_2} [D^2x(s)]^2 ds &= \int_{T_1}^{T_2} [D^2(\phi(s)\mathbf{c})]^2 ds = \\ &= \int_{T_1}^{T_2} [D^2(\phi(s)\mathbf{c})][D^2(\phi(s)\mathbf{c})]^\top ds = \\ &= \mathbf{c}^\top \left(\int_{T_1}^{T_2} [D^2\phi(s)][D^2\phi^\top(s)] ds \right) \mathbf{c} = \mathbf{c}^\top \mathbf{R} \mathbf{c}, \end{aligned} \quad (2.5)$$

kde nová matice \mathbf{R} je rozměrů $K \times K$ a obsahuje prvky

$$\{\mathbf{R}\}_{jk} = \int_{T_1}^{T_2} [D^2\phi_j(s)][D^2\phi_k^\top(s)] ds.$$

Pomocí maticového zápisu (2.3) a (2.5) přepíšeme penalizovaný reziduální součet čtverců

$$\text{PSSE}(\mathbf{c}, \lambda) = (\mathbf{y} - \Phi\mathbf{c})^\top (\mathbf{y} - \Phi\mathbf{c}) + \lambda \mathbf{c}^\top \mathbf{R} \mathbf{c}, \quad (2.6)$$

a v závislosti na \mathbf{c} hledáme minimum tohoto výrazu [20]. Derivujeme dle \mathbf{c}

$$\begin{aligned} \frac{\partial}{\partial \mathbf{c}} \text{PSSE}(\mathbf{c}, \lambda) &= \frac{\partial}{\partial \mathbf{c}} \left((\mathbf{y} - \Phi\mathbf{c})^\top (\mathbf{y} - \Phi\mathbf{c}) + \lambda \mathbf{c}^\top \mathbf{R} \mathbf{c} \right) = \\ &= \frac{\partial}{\partial \mathbf{c}} \left(\mathbf{y}^\top \mathbf{y} - 2\mathbf{c}^\top \Phi^\top \mathbf{y} + \mathbf{c}^\top \Phi^\top \Phi \mathbf{c} + \lambda \mathbf{c}^\top \mathbf{R} \mathbf{c} \right) = \\ &= -2\Phi^\top \mathbf{y} + 2\Phi^\top \Phi \mathbf{c} + 2\lambda \mathbf{R} \mathbf{c}, \end{aligned}$$

a položíme nule

$$\begin{aligned} -2\Phi^\top \mathbf{y} + 2\Phi^\top \Phi \mathbf{c} + 2\lambda \mathbf{R} \mathbf{c} &= 0 \\ -\Phi^\top \mathbf{y} + \Phi^\top \Phi \mathbf{c} + \lambda \mathbf{R} \mathbf{c} &= 0 \\ (\Phi^\top \Phi + \lambda \mathbf{R}) \mathbf{c} &= \Phi^\top \mathbf{y}. \end{aligned}$$

Formálně je potom výraz (2.6) minimální pro volbu bázových koeficientů

$$\mathbf{c}^{**} = (\Phi^\top \Phi + \lambda \mathbf{R})^{-1} \Phi^\top \mathbf{y}.$$

Získáme tak odhad vyhlazené křivky

$$\hat{x}(t) = \phi^\top(t) \mathbf{c}^{**} = \phi^\top(t) (\Phi^\top \Phi + \lambda \mathbf{R})^{-1} \Phi^\top \mathbf{y},$$

a podobně jako v případě vyhlazování pomocí lineární regrese, předpokládáme-li, že funkci $\hat{x}(t)$ evaluujeme ve stejných J bodech, ve kterých jsme daný proces sledovali, lze vyrovnané hodnoty psát pomocí vyhlazovací matice \mathbf{S}_λ o rozměrech $J \times J$ takto

$$\hat{\mathbf{x}} = \Phi (\Phi^\top \Phi + \lambda \mathbf{R})^{-1} \Phi^\top \mathbf{y} = \mathbf{S}_\lambda \mathbf{y}. \quad (2.7)$$

Z konstrukce minimalizační úlohy je již zřejmé, jaký vliv má volba vyhlazovacího parametru na výsledek. Čím je λ větší, tím více vyhlazení penalizujeme a funkce $x(t)$ data reflektuje méně. Zvolíme-li $\lambda = 0$, dostaneme úlohu metody nejmenších čtverců, kterou jsme rozebrali již dříve, a tím ignorujeme penalizaci. Pro $\lambda \rightarrow \infty$ bychom dostávali maximálně hladkou funkci, tj. konstantu [10].

2.3.1. Volba vyhlazovacího parametru

Zbývá otázka, jak vyhlazovací parametr λ vhodně zvolit. Můžeme jej zvolit na pevnou dle požadující míry vyhlazení. Nicméně existuje sofistikovanější přístup volby parametru, a to pomocí zobecněné křížové validace (z angl. zkráceně GCV).

V mnoha metodách analýzy mnohorozměrných dat volíme parametry metod pomocí klasické křížové validace (zkráceně CV), například tzv. k -fold CV či LOO CV (z angl. leave-one-out). Kritériem CV bývá obvykle střední čtvercová chyba, obecně

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2,$$

přičemž \hat{y}_i je i -tá vyrovnaná hodnota z modelu [9]. Leave-one-out CV se provádí tak, že jedno pozorování z dat ponecháme jako testovací a zbytek zahrneme do trénovací množiny. Na trénovací množině zkonstruujeme příslušný model s danou hodnotou parametru a tento model následně vyhodnotíme na testovacím pozorování pomocí čtvercové chyby v tomto bodě. Takto postupně vynecháváme každé pozorování a vypočteme hodnotu kritéria LOO CV jako průměr dílčích chyb. Ve výsledku potom vybereme takovou hodnotu parametru z námi zadané přípustné posloupnosti, pro niž jsme dostali nejnižší hodnotu parametru. k -fold CV funguje podobně s tím rozdílem, že data rozdělíme do k bloků a vždy jeden ponecháváme jako testovací [9].

Kritérium LOO CV má v našem kontextu tvar

$$\text{CV}_\lambda = \frac{1}{J} \sum_{j=1}^J \text{MSE}_j = \frac{1}{J} \sum_{j=1}^J [y(t_j) - \hat{x}^{(-j)}(t_j)]^2, \quad (2.8)$$

kde $\hat{x}^{(-j)}(t_j)$ je odhad funkce $x(t)$ v bodě t_j napočítaný na datech bez pozorování v bodě t_j , tj. při vynechání pozorování $y(t_j)$.

Lze ukázat, že kritérium (2.8) lze pro dané λ spočítat jednoduše následujícím

způsobem

$$\text{CV}_\lambda = \frac{1}{J} \sum_{j=1}^J \left[\frac{y(t_j) - \hat{x}(t_j)}{1 - \{\mathbf{S}_\lambda\}_{jj}} \right]^2. \quad (2.9)$$

V případě funkcionálních dat má ovšem tento přístup tendenci k podhlazování [20]. Tomuto problému se vyhneme právě použitím metody GCV. Ta spočívá v nahrazení členu $\{\mathbf{S}_\lambda\}_{jj}$ ve výrazu (2.9) průměrem diagonálních prvků vyhlazovací matice \mathbf{S}_λ , tedy

$$\text{GCV}_\lambda = \frac{1}{J} \sum_{j=1}^J \left[\frac{y(t_j) - \hat{x}(t_j)}{1 - \frac{\text{Tr}(\mathbf{S}_\lambda)}{J}} \right]^2, \quad (2.10)$$

kde $\text{Tr}(\mathbf{S}_\lambda)$ je stopa vyhlazovací matice \mathbf{S}_λ z (2.7) [12]. Výraz (2.10) lze přepsat do tvaru

$$\text{GCV}_\lambda = \left(\frac{J}{J - \text{Tr}(\mathbf{S}_\lambda)} \right) \left(\frac{\text{SSE}_\lambda}{J - \text{Tr}(\mathbf{S}_\lambda)} \right),$$

kde SSE_λ je reziduální součet čtverců pro odhad vyhlazené funkce s daným parametrem λ [20]. Optimální volbou je opět potom takové λ , pro něž kritérium GCV nabývá minimální hodnoty [12].

Je nutné zmínit, že i když máme k dispozici sofistikovaný přístup, prakticky nemusí být vždy tou nejlepší volbou. Většinou jde o subjektivní posouzení experta pro konkrétní problém. Někdy můžeme cíleně požadovat větší míru vyhlazení, než poskytuje parametr z GCV, a tudíž zadáme hodnotu parametru fixně takovou, jež nám poskytne požadovaný výsledek.

Kapitola 3

Testy shody dvou funkcionálních výběrů

V této kapitole uvedeme testy potřebné k řešení problému z praxe, který budeme rozebírat v praktické části práce. Chtěli bychom otestovat shodu dvou funkcionálních výběrů z hlediska funkcí průměru. Uvedeme si dva přístupy, které můžeme nazvat jako globální a lokální.

Předpokládáme, že máme již vyhlazené křivky $x_n(t)$, $n = 1, \dots, N$, evaluované ve stejných bodech t_j , $j = 1, \dots, J$, jenž jsou dostatečně husté. Dále předpokládáme, že lze data rozdělit na dva nezávislé výběry rozsahu N_1 a N_2 , kde $N_1 + N_2 = N$. Funkce potom značíme dle příslušnosti k danému funkcionálnímu výběru následovně

$$\begin{aligned} &x_1^1(t_j), x_2^1(t_j), \dots, x_{N_1}^1(t_j), \\ &x_1^2(t_j), x_2^2(t_j), \dots, x_{N_2}^2(t_j), \quad N = N_1 + N_2, \quad j = 1, \dots, J. \end{aligned}$$

Dále označme $\mu_1(t)$, $\mu_2(t)$ funkcemi středních hodnot těchto výběrů a $\bar{x}_1(t)$, $\bar{x}_2(t)$ jejich funkcemi průměru.

3.1. Globální test shody dvou funkcí středních hodnot

Jak již plyne z názvu, test, jenž si nyní popíšeme, posuzuje globální shodu výběrů z hlediska funkcí středních hodnot. Dostaneme z něj tedy odpověď na otázku, zda se tyto funkce signifikantně liší na intervalu $[T_1, T_2]$, nikoliv jak je tomu na různých menších podintervalech. Přesto nám globální test v analýze poskytne důležitou prvotní informaci.

Mějme nulovou hypotézu proti alternativní

$$H_0 : \mu_1(t) = \mu_2(t), \quad \forall t \in [T_1, T_2], \quad (3.1)$$

$$H_A : \exists t \in [T_1, T_2] : \mu_1(t) \neq \mu_2(t), \quad (3.2)$$

čili najdeme-li takové t z intervalu $[T_1, T_2]$, pro něž jsou funkce středních hodnot výběrů signifikantně odlišné, zamítáme nulovou hypotézu shody.

Konstrukce globálního testu vychází z jednoduché myšlenky J. O. Ramsaye, který použil kombinaci permutačního testu a statistiky Welchova dvouvýběrového t-testu. Ramsay algoritmus tohoto testu, který si vysvětlíme detailně, implementoval v softwaru R do balíčku `fda` pod funkcí `tperm.fd()` [17].

Jelikož je součástí algoritmu statistika Welchova dvouvýběrového t-testu, nejprve si jej krátce představíme. Jedná se o klasický dvouvýběrový t-test shody středních hodnot za předpokladu normality dat, ovšem bez předpokladu na shodu rozptylů výběrů [8]. Testová statistika vychází z rozdělení rozdílu průměrů výběrů. Za platnosti nulové hypotézy o shodě středních hodnot výběrů má testová statistika tvar

$$\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\widehat{\text{var}}(X_1)}{n_1} + \frac{\widehat{\text{var}}(X_2)}{n_2}}}, \quad (3.3)$$

kde n_1 je rozsah prvního a n_2 rozsah druhého výběru [8]. Abychom se vyhnuli silnému předpokladu normálního rozdělení dat, provedeme test s uplatněním tes-

tové statistiky (3.3) permutačně.

Algoritmus

Nechť je test proveden na hladině významnosti α a nechť je B počet permutací (doporučujeme alespoň 1000). Dále mějme oba funkcionální výběry, evaluované ve stejných bodech $t_j, j = 1, \dots, J$, uloženy v maticích takto

$$X_1 = \begin{bmatrix} x_1^1(t_1) & x_2^1(t_1) & \dots & x_{N_1}^1(t_1) \\ x_1^1(t_2) & x_2^1(t_2) & \dots & x_{N_1}^1(t_2) \\ \vdots & \vdots & & \vdots \\ x_1^1(t_J) & x_2^1(t_J) & \dots & x_{N_1}^1(t_J) \end{bmatrix}, \quad X_2 = \begin{bmatrix} x_1^2(t_1) & x_2^2(t_1) & \dots & x_{N_2}^2(t_1) \\ x_1^2(t_2) & x_2^2(t_2) & \dots & x_{N_2}^2(t_2) \\ \vdots & \vdots & & \vdots \\ x_1^2(t_J) & x_2^2(t_J) & \dots & x_{N_2}^2(t_J) \end{bmatrix},$$

čili ve sloupcích jsou jednotlivé křivky a v řádcích body evaluace. Spojením matic sloupci vznikne matice X o rozměrech $J \times N$, kde $N = N_1 + N_2$.

Následně proběhne iterační proces celkem B -krát, přičemž b -tá iterace postupuje takto

1. Z matice X vytvoříme novou matici X^* , která vznikla permutováním sloupců matice X , tj. ze sloupců původní matice X provedeme náhodný výběr s opakováním rozsahu N .
2. Z matice X^* vypočítáme nové funkce průměru $\bar{x}_1^*(t)$ z prvních N_1 sloupců a $\bar{x}_2^*(t)$ ze sloupců $N_1 + 1$ až N , přičemž je zřejmé, že máme stále k dispozici pouze evaluace těchto funkcí v bodech t_1 až t_J .
3. Analogicky jako v kroku 2 spočítáme odhady funkcí rozptylu $\widehat{\text{var}}_{X_1}^*(t)$ a $\widehat{\text{var}}_{X_2}^*(t)$.
4. Pro každý pevný bod t_j napočítáme Welchovu statistiku (3.3). Hodnoty

testové statistiky ukládáme do vektoru $\mathbf{s} = (s_1, \dots, s_J)^\top$, kde

$$s_j = \frac{|\bar{x}_1^*(t_j) - \bar{x}_2^*(t_j)|}{\sqrt{\frac{\widehat{\text{var}}_{X_1}^*(t_j)}{N_1} + \frac{\widehat{\text{var}}_{X_2}^*(t_j)}{N_2}}}. \quad (3.4)$$

Absolutní hodnota je přidána kvůli snazšímu výpočtu výsledné p-hodnoty testu.

5. Určíme maximální hodnotu testové statistiky $s_{\max,b} = \max_j s_j$.

Provedeme-li tyto kroky celkem B -krát, výsledkem bude vektor maximálních hodnot statistik pro jednotlivé permutace čili $\mathbf{s}_{\max} = (s_{\max,1}, \dots, s_{\max,B})$. Posledním krokem je určit výslednou p-hodnotu testu. Tu odhadneme jako

$$\text{p-value} = \frac{\sum_{b=1}^B \mathbb{1}\{s_{\max,b} \geq s_{\text{obs}}\}}{B},$$

kde s_{obs} je maximální hodnota testové statistiky (3.4) pro původní data a $\mathbb{1}$ je indikátorová funkce. Finální p-hodnotou globálního testu je tedy relativní podíl maximálních statistik, jež jsou větší či rovny maximální hodnotě testové statistiky pro původní data.

Je-li p-hodnota menší než zadaná hladina významnosti α , zamítáme nulovou hypotézu o shodě funkcí středních hodnot funkcionálních výběrů. Z globálního hlediska jsou funkce středních hodnot signifikantně odlišné.

3.2. Intervalový test shody dvou funkcí průměru

Globální test, jenž jsme popsali výše, slouží ke globálnímu posouzení shody dvou funkcionálních výběrů z hlediska funkcí středních hodnot. Chceme-li znát přesné úseky z intervalu $[T_1, T_2]$, v nichž k signifikantnímu rozdílu funkcí středních hodnot dochází, aplikujeme intervalový test dle Pini a Vantiniho. Zmínění autoři v roce 2016 popsali tento test v článku [14] a v softwaru R jej implementovali do

balíčku `fdatest` pod funkcí `ITP2bspline()` [16].

Intervalový test se skládá ze třech fází [15]. První fází je básová expanze, které jsme se věnovali v druhé kapitole. Tento test je postaven na reprezentaci funkcí právě básovými koeficienty, a proto jsou důležitým výstupem první fáze dvě matice koeficientů pro každý výběr

$$\begin{bmatrix} c_{11}^{(1)} & c_{12}^{(1)} & \dots & c_{1p}^{(1)} \\ \vdots & \vdots & & \vdots \\ c_{N_11}^{(1)} & c_{N_12}^{(1)} & \dots & c_{N_1p}^{(1)} \end{bmatrix}, \quad \begin{bmatrix} c_{11}^{(2)} & c_{12}^{(2)} & \dots & c_{1p}^{(2)} \\ \vdots & \vdots & & \vdots \\ c_{N_21}^{(2)} & c_{N_22}^{(2)} & \dots & c_{N_2p}^{(2)} \end{bmatrix},$$

rozměrů $N_1 \times p$ a $N_2 \times p$. Každá funkce daného výběru je jednoznačně reprezentována p koeficienty [14]. Koeficient $c_{jk}^{(1)}$ je k -tý básový koeficient j -té funkce prvního výběru.

Zavedme dále matici všech koeficientů \mathbf{C}

$$\mathbf{C} = \begin{bmatrix} c_{11}^{(1)} & \dots & c_{1p}^{(1)} \\ \vdots & & \vdots \\ c_{N_11}^{(1)} & \dots & c_{N_1p}^{(1)} \\ c_{11}^{(2)} & \dots & c_{1p}^{(2)} \\ \vdots & & \vdots \\ c_{N_21}^{(2)} & \dots & c_{N_2p}^{(2)} \end{bmatrix},$$

o rozměrech $N \times p$.

Jelikož intervalový test posuzuje opět shodu funkcí středních hodnot, potřebujeme jejich reprezentaci prostřednictvím koeficientů. Funkci střední hodnoty prvního výběru $\mu_1(t)$ budeme reprezentovat vektorem básových koeficientů $\mathbf{c}^{(1)} = (c_1^{(1)}, \dots, c_p^{(1)})^\top$. Analogicky pak funkci střední hodnoty druhého výběru $\mu_2(t)$ reprezentuje vektor koeficientů $\mathbf{c}^{(2)} = (c_1^{(2)}, \dots, c_p^{(2)})^\top$. Při testování pracujeme s odhady funkcí středních hodnot, tj. s funkcemi průměru, které lze reprezentovat vektorem průměrných koeficientů daného výběru, a sice $\bar{\mathbf{c}}^{(1)} = (\bar{c}_1^{(1)}, \dots, \bar{c}_p^{(1)})^\top$

pro první výběr, kde $\bar{c}_k^{(1)} = \frac{1}{N_1} \sum_{i=1}^{N_1} c_{ik}^{(1)}$. Podobně potom pro druhý výběr je tento vektor ve tvaru $\bar{\mathbf{c}}^{(2)} = (\bar{c}_1^{(2)}, \dots, \bar{c}_p^{(2)})^\top$, přičemž $\bar{c}_k^{(2)} = \frac{1}{N_2} \sum_{i=1}^{N_2} c_{ik}^{(2)}$.

Ve druhé fázi provedeme celkem p jednorozměrných permutačních testů, přičemž k -tý test ověřuje shodu k -tých koeficientů prvního a druhého výběru pro $k = 1, \dots, p$ [14]. Nulová hypotéza tohoto testu je ve tvaru

$$H_0^{(k)} : c_k^{(1)} = c_k^{(2)}.$$

Jelikož testujeme shodu koeficientů výběrů, je napozorovaná testová statistika T_{obs} založena na absolutním rozdílu jejich průměrných koeficientů [15]. Tato statistika T_{obs} je tedy p -rozměrným vektorem ve tvaru

$$T_{obs} = (|\bar{c}_1^{(1)} - \bar{c}_1^{(2)}|, \dots, |\bar{c}_p^{(1)} - \bar{c}_p^{(2)}|)^\top,$$

přičemž pro k -tý test použijeme právě k -tý prvek této statistiky, který označíme jako $T_{obs}^{(k)}$.

Následujícím algoritmem otestujeme všechny hypotézy $H_0^{(k)}$, $k = 1, \dots, p$

1. Provedeme celkem B permutací řádků matice \mathbf{C} (bez opakování). Nějakou b -tou permutací této matice koeficientů zapisujeme takto

$$\begin{bmatrix} c_{11}^{(1)*} & \dots & c_{1p}^{(1)*} \\ \vdots & & \vdots \\ c_{N_{11}}^{(1)*} & \dots & c_{N_{1p}}^{(1)*} \\ c_{11}^{(2)*} & \dots & c_{1p}^{(2)*} \\ \vdots & & \vdots \\ c_{N_{21}}^{(2)*} & \dots & c_{N_{2p}}^{(2)*} \end{bmatrix}.$$

2. Pro b -tou permutaci, $b = 1, \dots, B$, spočteme testovou statistiku

$$T_b = (|\bar{c}_1^{(1)*} - \bar{c}_1^{(2)*}|, \dots, |\bar{c}_p^{(1)*} - \bar{c}_p^{(2)*}|)^\top,$$

kde $\bar{c}_k^{(1)*} = \frac{1}{N_1} \sum_{i=1}^{N_1} c_{ik}^{(1)*}$ a $\bar{c}_k^{(2)*} = \frac{1}{N_2} \sum_{i=1}^{N_2} c_{ik}^{(2)*}$ [15]. Sestavíme matici \mathbf{T} obsahující v řádcích vektory testových statistik jednotlivých permutací

$$\mathbf{T} = \begin{bmatrix} T_1^\top \\ T_2^\top \\ \vdots \\ T_B^\top \end{bmatrix}.$$

3. Nakonec pro k -tou hypotézu $H_0^{(k)}$, $k = 1, \dots, p$, odhadneme p-hodnotu jako proporci případů, kdy testová statistika vypočtena z permutovaných k -tých koeficientů byla větší či rovna napozorované hodnotě $T_{obs}^{(k)}$

$$\lambda_k = \frac{\sum_{i=1}^B \mathbb{1}\{\{\mathbf{T}\}_{ik} \geq T_{obs}^{(k)}\}}{B}.$$

V druhé fázi jsme tak získali celkem p marginálních p-hodnot $\lambda_1, \dots, \lambda_p$ příslušejících hypotézám o shodě daných koeficientů prvního a druhého nezávislého výběru.

Ve třetí fázi bude naším cílem vyhodnotit průniky sousedních dílčích hypotéz $H_0^{(k)}$ [14]. Budeme postupovat od průniku dvou sousedních hypotéz, poté třech, až po průnik všech p dílčích hypotéz

$$H_0^{(k,k+1)} = H_0^{(k)} \cap H_0^{(k+1)}, \quad k = 1, \dots, p-1,$$

$$H_0^{(k,k+1,k+2)} = H_0^{(k)} \cap H_0^{(k+1)} \cap H_0^{(k+2)}, \quad k = 1, \dots, p-2,$$

⋮

$$H_0^{(1,\dots,p)} = \bigcap_{i=1}^p H_0^{(k)}.$$

Párové hypotézy vyhodnotíme vhodným zkombinováním marginálních p-hodnot. Pro tento účel je v intervalovém testu použit NPC algoritmus (z angl. nonpa-

rametric combination), jenž kombinuje marginální p-hodnoty pomocí tzv. kombinující funkce [2]. Pini a Vantini použili Fisherovu kombinující funkci ψ ve tvaru

$$\psi(x_1, \dots, x_n) = -2 \sum_{j=1}^n \log x_j.$$

Nejprve vypočítáme napozorovanou testovou statistiku pro průnik hypotéz $H_0^{(k)}$ a $H_0^{(k+1)}$ takto

$$T_{obs}^{(k,k+1)} = \psi(\lambda_k, \lambda_{k+1}) = -2(\log \lambda_k + \log \lambda_{k+1}), \quad k = 1, \dots, p-1,$$

a dále zkonstruujeme matici \mathbf{L} rozměru $B \times p$, jež obsahuje prvky $L_b^{(k)}$, $b = 1, \dots, B$, $k = 1, \dots, p$ [6]. Hodnotu $L_b^{(k)}$ získáme jako

$$L_b^{(k)} = \frac{\sum_{j=1}^B \mathbb{1}\{\{\mathbf{T}\}_{jk} \geq \{\mathbf{T}\}_{bk}\} + \frac{1}{2}}{B+1},$$

a tedy jde o (upravenou) proporcii případů, kdy prvek matice \mathbf{T} v b -tém řádku a k -tém sloupci je větší či roven než prvky v jejím k -tém sloupci. Úprava hodnoty dle NPC algoritmu v podobě přičtení $1/2$ v čitateli a 1 ve jmenovateli je z toho důvodu, aby výsledná p-hodnota náležela otevřenému intervalu $(0, 1)$, nikoliv intervalu uzavřenému [6].

Výslednou párovou testovou statistiku pro b -tou permutaci získáme opět pomocí kombinující funkce jako

$$T_b^{(k,k+1)} = \psi(L_b^{(k)}, L_b^{(k+1)}) = -2(\log L_b^{(k)} + \log L_b^{(k+1)}),$$

přičemž tuto statistiku vypočteme pro každé $k = 1, \dots, p-1$ a $b = 1, \dots, B$ [6]. Párovou p-hodnotu testu hypotézy $H_0^{(k,k+1)}$ vyhodnotíme potom následovně

$$\lambda_{k,k+1} = \frac{\sum_{b=1}^B \mathbb{1}\{T_b^{(k,k+1)} \geq T_{obs}^{(k,k+1)}\}}{B}.$$

Podobným postupem bychom vypočítali kombinované p-hodnoty $\lambda_{k,k+1,k+2}$ pro $k = 1, \dots, p-2$. Napozorovanou testovou statistiku pro průnik hypotéz $H_0^{(k)}$, $H_0^{(k+1)}$ a $H_0^{(k+2)}$ spočteme jako

$$T_{obs}^{(k,k+1,k+2)} = \psi(\lambda_k, \lambda_{k+1}, \lambda_{k+2}) = -2(\log \lambda_k + \log \lambda_{k+1} + \log \lambda_{k+2}).$$

Dále pro každou permutaci $b = 1, \dots, B$ vyhodnotíme testovou statistiku

$$T_b^{(k,k+1,k+2)} = \psi(L_b^{(k)}, L_b^{(k+1)}, L_b^{(k+2)}) = -2(\log L_b^{(k)} + \log L_b^{(k+1)} + \log L_b^{(k+2)}),$$

a p-hodnoty potom jako

$$\lambda_{k,k+1,k+2} = \frac{\sum_{b=1}^B \mathbb{1}\{T_b^{(k,k+1,k+2)} \geq T_{obs}^{(k,k+1,k+2)}\}}{B}.$$

Analogickým postupem vypočteme zbývající p-hodnoty $\lambda_{k,k+1,k+2,k+3}$ pro $k = 1, \dots, p-3$ až finální p-hodnotu $\lambda_{1,\dots,p}$ příslušnou průniku všech hypotéz.

Zavedme značení dle Pini a Vantiniho [14]. $\lambda_{\mathbf{k}}$ reprezentuje p-hodnotu testu hypotézy $H_0^{(\mathbf{k})} = \bigcap_{k \in \mathbf{k}} H_0^{(k)}$, kde \mathbf{k} je podmnožinou množiny indexů $\{1, \dots, p\}$. Hypotéza $H_0^{(\mathbf{k})}$ je tedy průnikem všech hypotéz, v nichž se vyskytuje index k . Stejným způsobem zavedme $\lambda_{\mathbf{k}^C}$ jakožto p-hodnotu testu hypotézy $H_0^{(\mathbf{k}^C)} = \bigcap_{k \notin \mathbf{k}} H_0^{(k)}$, kde \mathbf{k}^C je komplementem (doplňkem) množiny \mathbf{k} vzhledem k množině $\{1, \dots, p\}$. Výslednou, už korigovanou, p-hodnotu testu shody k -tých koeficientů výběrů určíme následovně

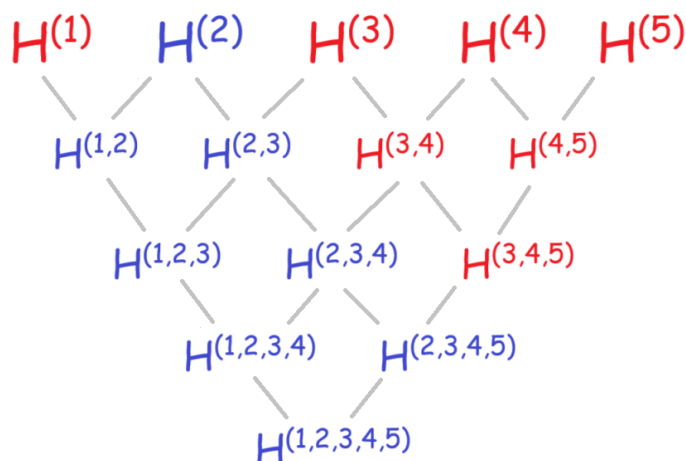
$$\lambda_k^{\text{ITP}} = \max \left(\max_{k \in \mathbf{k}} \lambda_{\mathbf{k}}, \max_{k \in \mathbf{k}^C} \lambda_{\mathbf{k}^C} \right). \quad (3.5)$$

Do výpočtu bereme v potaz právě i výše uvedené komplementy kvůli korekci počtu výskytů krajních hypotéz (odpovídajících krajním koeficientům), kterých by bylo jinak ve výpočtech p-hodnot podstatně méně oproti hypotézám „uprostřed“ [15].

Je-li hodnota λ_k^{ITP} menší než zadaná hladina významnosti α , zamítáme nulo-

ovou hypotézu $H_0^{(k)}$ o shodě k -tých koeficientů prvního a druhého výběru [14].

Pro názornost způsobu výpočtu konečné p -hodnoty si nastiňme příklad výpočtu λ_2^{ITP} pro $p = 5$. Na obrázku 3.1 je sestavena hierarchie hypotéz, jak jsme ji popsali v algoritmu intervalového testu. Modře jsou vyznačeny hypotézy zahrnující index $k = 2$ a červeně hypotézy komplementu, tedy hypotézy neobsahující tento index. Ke všem hypotézám mějme vypočtené příslušné p -hodnoty. Dále z obou množin vybereme dle vzorce (3.5) maximální p -hodnoty a ta větší z nich je pak výsledná p -hodnota λ_2^{ITP} .



Obrázek 3.1: Příklad hierarchie hypotéz pro 5 koeficientů.

Zbývá vyřešit otázku, jak z reprezentace bázevými koeficienty přejít zpět do původních funkcí na intervalu $[T_1, T_2]$. Na základě p -hodnot λ_k^{ITP} , $k = 1, \dots, p$, bychom chtěli identifikovat úseky intervalu $[T_1, T_2]$, v nichž dochází k významnému rozdílu funkcí středních hodnot výběrů.

Mějme kubickou B -splajnovou bázi a interval $[T_1, T_2]$ pro jednoduchost přeškálovaný na $[0, 1]$. Počet bázevých funkcí, jak již víme, je $g + k' + 1$, kde k' je stupeň B -splajnu a g je počet vnitřních uzlů. V našem příkladu je $k' = 3$ a báze obsahuje celkem $g + 4$ známých koeficientů, přičemž krajní koeficienty přísluší krajním B -splajnům doplňujícím bázi. Vnitřních koeficientů tudíž máme $p = g + 2$ a jedná se o koeficienty c_1, \dots, c_p s nimiž jsme pracovali v intervalovém testu. Dále před-

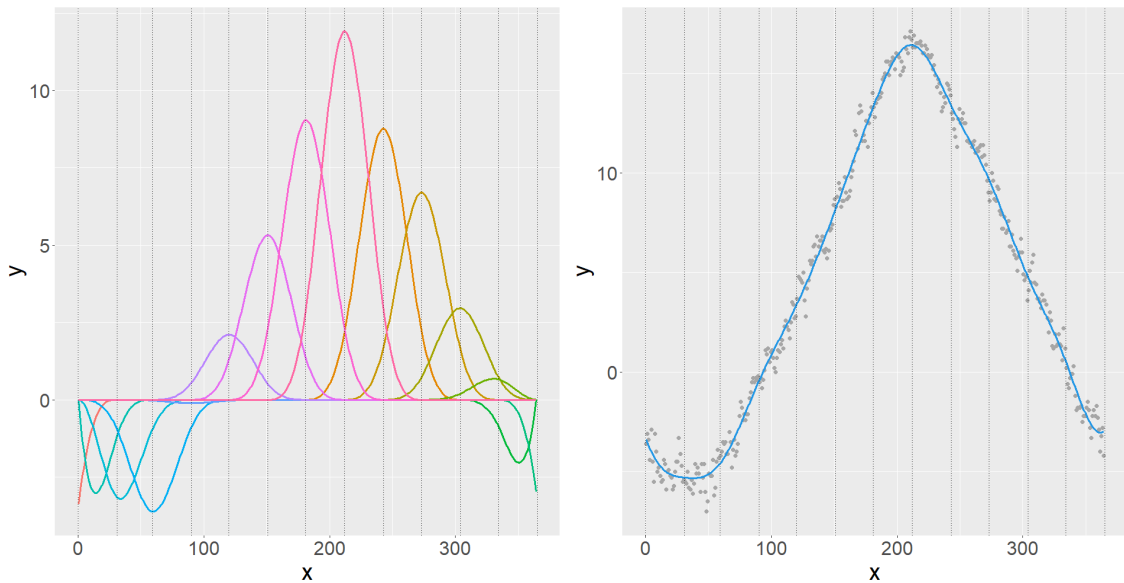
pokládejme, že uzly báze jsou dány ekvidistantně. Interval $[0, 1]$ poté stačí pouze rozdělit na $p-1$ podintervalů délky $\frac{1}{p}$. Koeficient c_1 tedy odpovídá intervalu $[0, \frac{1}{p})$, koeficient c_2 odpovídá intervalu $[\frac{1}{p}, \frac{2}{p})$ až c_p intervalu $[\frac{p-1}{p}, 1)$. Obecně interval odpovídající nějakému koeficientu c_k je tedy dán jako

$$\left[\frac{k-1}{p}, \frac{k}{p} \right). \quad (3.6)$$

Přeškálujeme-li interval $[0, 1]$ zpět do $[T_1, T_2]$, dostaneme z intervalu (3.6) interval

$$\left[\frac{k-1}{p}(T_2 - T_1) + T_1, \frac{k}{p}(T_2 - T_1) + T_1 \right). \quad (3.7)$$

Zamítne-li intervalový test shodu koeficientů c_k u prvního a druhého výběru, na intervalu (3.7) došlo k signifikantnímu rozdílu funkcí středních hodnot výběrů.



Obrázek 3.2: Příklad kubické B -splajnové báze s 13 uzly vybraného funkcionálního pozorování. Zdroj: [11].

Uveďme si následující příklad. Na obrázku 3.2 lze názorně vidět, jak koeficienty jednotlivých bázeových funkcí souvisí s podintervaly $[T_1, T_2]$. Obrázek znázorňuje kubickou B -splajnovou bázi vybraného funkcionálního pozorování na inter-

valu $[0; 364,5]$, kde bylo zvoleno 13 ekvidistantních uzlů. Vnitřních uzlů je tudíž 11 a počet bázových funkcí potom 15. Parametr p je v této situaci tedy roven 13. Například koeficientu c_1 tedy dle vzorce (3.7) odpovídá interval

$$\left[\frac{1-1}{13} 364,5; \frac{1}{13} 364,5 \right) = [0; 28,04).$$

Tudíž pokud by výsledek intervalového testu ukázal statisticky významný rozdíl koeficientů c_1 prvního a druhého výběru, znamenalo by to, že k signifikantnímu rozdílu ve funkcích středních hodnot došlo právě na intervalu $[0; 28,04)$.

Kapitola 4

Analýza efektu dopravních opatření

Jednou z úloh ve světě dopravy je analýza efektu již instalovaných dopravních opatření. V potřebných lokalitách pozemních komunikací jsou instalována dopravní opatření různého typu, jako jsou svislá a vodorovná dopravní značení, zpomalovací prahy a podobně. Abychom zjistili skutečný efekt těchto opatření, lze se v některých případech opřít o regionální data nehodovosti. Podrobnou analýzu efektu opatření v daných lokalitách lze provést pomocí tzv. plovoucích vozidel. Jedná se o vozidla, která jsou registrována u určitého poskytovatele dat, vybavená systémem GPS, jenž s určitými časovými odstupy neustále zaznamenává údaje o vozidle v reálném čase během běžného provozu. Zaznamenává se zejména aktuální čas, rychlost vozidla a GPS poloha [1]. Plovoucí vozidla zaznamenávají údaje po určitý čas, typicky několik měsíců. Abychom mohli posoudit vliv instalovaných opatření na rychlost těchto vozidel, je nezbytné tyto údaje měřit již před jejich instalací. Jak se však v této kapitole přesvědčíme, podrobná a správná analýza vlivu opatření na rychlost vozidel není problematikou jednoduchou zejména z důvodu potřebného množství finančně náročných dat a nemalého množství intervenujících vnějších faktorů.

Data z plovoucích vozidel, zkráceně FCD (z angl. floating car data) jsou data velmi objemná a vyžadující náročný pre-processing. FCD jsou původně prosto-

rová data a konkrétní zkoumané úseky silnic je nutné převést do klasické podoby dvourozměrných dat. Později si tento postup lehce nastíníme. Problematice pre-processingu FCD se v České republice mimo jiné věnuje Centrum dopravního výzkumu, v. v. i. (CDV), jenž poskytlo data pro tuto práci.

4.1. Představení dat a analyzovaných dopravních opatření

V této kapitole budeme podrobně analyzovat FCD ze dvou španělských obcí Agolada a Cuntis. Data z Cuntisu obsahují silniční úsek, v němž jsou dvě oblasti zájmu, a tudíž bude tato obec dále rozdělena na Cuntis dolní a Cuntis horní. Máme k dispozici FCD od ledna roku 2022 do března roku 2023 (včetně). Instalace opatření probíhala konkrétně během září 2022, tudíž toto období z analýzy vyloučíme. Data z obou obcí zahrnují oba směry průjezdů. Tyto směry dále budeme rozlišovat čísly 1 a 2. Jednotlivé jízdy plovoucích vozidel jsou popsány identifikačními čísly. Plovoucími vozidly byla v tomto případě nejen osobní vozidla, ale rovněž nákladní vozy, dodávky, kamiony aj. Každá jízda obsahuje několik časových záznamů se všemi měřenými údaji. Obecně je pro různé jízdy různý počet záznamů.



Obrázek 4.1: Instalovaná dopravní opatření v obcích Agolada a Cuntis. Vlevo: dračí zuby. Vpravo: klikaté čáry. Zdroj: CDV.

Dopravními opatřeními, jejichž efekt budeme v této práci zkoumat jsou tzv. dračí zuby a bílé klikaté čáry. Dračí zuby, viz fotografie vlevo na obrázku 4.1, slouží k optickému zúžení vozovky, což by mělo zapříčinit zpomalení řidičů při vjezdu do obce. Bílé klikaté čáry na krajnicích vozovky se objevují bezprostředně před přechody pro chodce a jejich účelem je větší opatrnost řidičů vůči přecházejícím chodcům. V některých úsecích byly také přidány značky rychlostního limitu 70 km/h a 50 km/h.

	Agolada		Cuntis dolní		Cuntis horní	
	Směr 1	Směr 2	Směr 1	Směr 2	Směr 1	Směr 2
Opatření	Zuby	Zuby	Limit 70	Čáry	Limit 70	Limit 70
	Čáry	Čáry	Limit 50			Limit 50
	Čáry	Čáry	Zuby			Zuby
			Čáry			

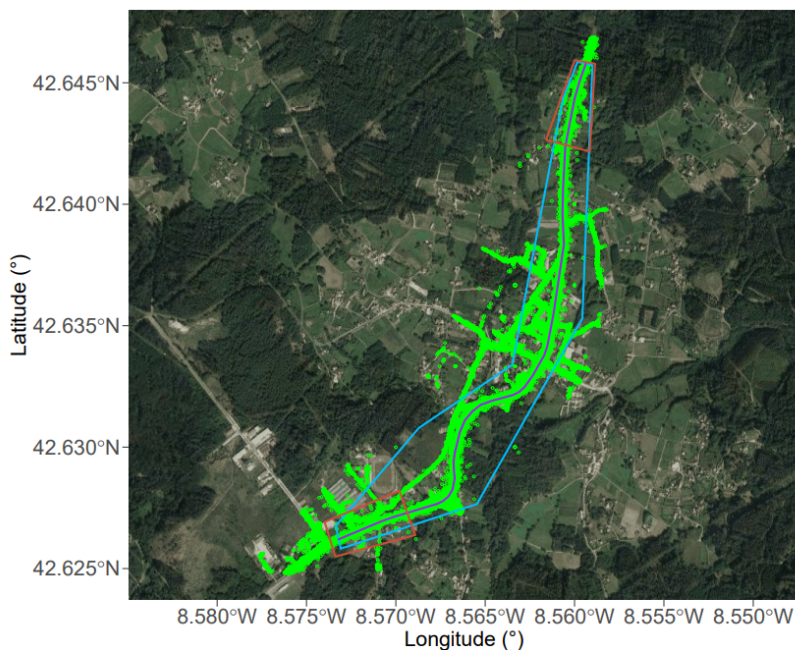
Tabulka 4.1: Seznam chronologicky seřazených instalovaných opatření pro daný úsek a směr.

Pro lepší orientaci v instalovaných opatřeních v jednotlivých zájmových úsecích byla vytvořena tabulka 4.1, kde je chronologický seznam nainstalovaných opatření v daném úseku a daném směru. V obci Agolada byly při vjezdu do obce instalovány dračí zuby a v obci před dvěma přechody pro chodce klikaté čáry. Všechna tato dopravní značení byla instalována v obou směrech. V dolním Cuntisu byly před vjezdem do obce (tj. ve směru 1) přidány značky rychlostního limitu, přičemž ve vzdálenosti asi 650 metrů limit 70 km/h a asi 300 metrů před obcí limit 50 km/h. Dále byly při vjezdu do obce přidány dračí zuby a v obci před přechodem pro chodce klikaté čáry. Ve směru 2 dolního Cuntisu se nachází pouze klikaté čáry rovněž před daným přechodem pro chodce. V úseku horního Cuntisu se přechod pro chodce nenachází, a tudíž zde byly nainstalovány jen dračí zuby při vjezdu do obce (tj. směr 2) a v tomto směru byla také přidána značka rychlostního limitu 50 km/h 50 metrů před vjezdem do obce a značka limitu 70

km/h 800 metrů před vjezdem do obce. Ve směru 1 tohoto úseku byla přidána značka rychlostního limitu 70 km/h zhruba 50 metrů po výjezdu z obce. GPS souřadnice začátku a konce všech opatření jsou známa.

Některé značky limitu 70 km/h se nachází na kraji měřeného úseku a tyto části v úsecích budou během pre-processingu dat z určitých, později vysvětlených, důvodů odseknuty. Nicméně je potřeba stále brát v potaz vliv dané značky na rychlost vozidel v další analyzované části úseku.

Je zřejmé, že na většině zkoumaných úseků budeme zjišťovat efekt opatření jakožto celku, nikoliv nějakého konkrétního.

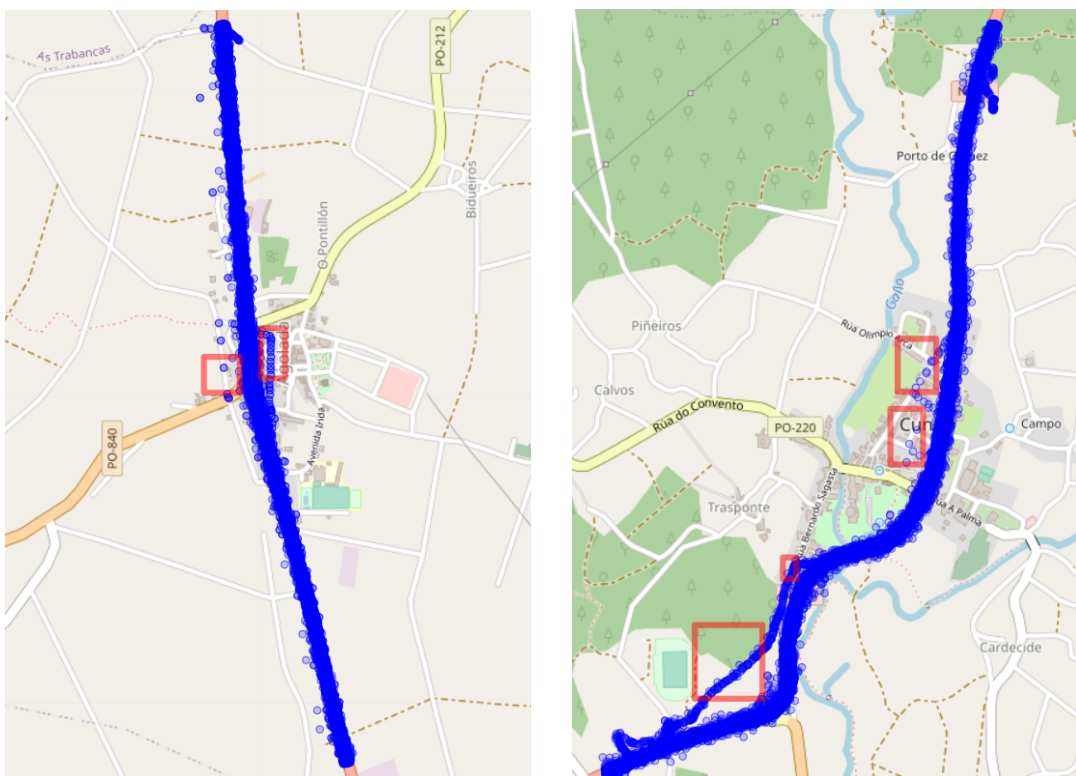


Obrázek 4.2: Ukázka filtrace zkoumaného úseku FCD vnějším polygonem.

4.2. Pre-processing FCD

Jelikož zdrojová FCD obsahují veškerá data z registrovaných plovoucích vozidel, která se pohybují v měřené lokalitě dle svých potřeb, je nezbytné úseky zájmu

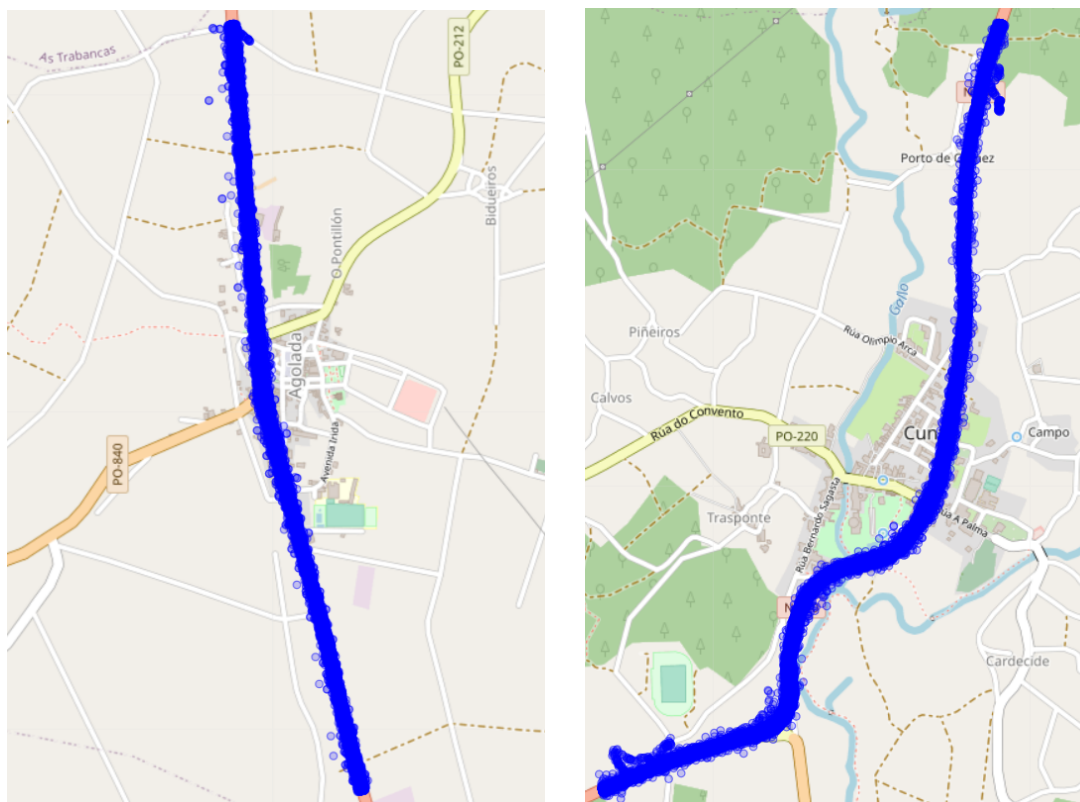
vyfiltrovat. Hrubá filtrace je provedena vnějším polygonem, jenž je pomocí GPS souřadnic umístěn kolem daného úseku silnice, viz modrý polygon na obrázku 4.2. Záznamy mimo tento polygon jsou ze souboru vyjmuty. Červené polygony na stejném obrázku identifikují startovací a cílové pozice, čímž lze určit směr jízdy.



Obrázek 4.3: FCD vyfiltrovaného úseku silnice v obci Agolada (vlevo) a Cuntis (vpravo). Červeně ohraničené oblasti zachycují odbočující jízdy.

Na obrázku 4.3 jsou znázorněny FCD vyfiltrovaného úseku silnice v obcích Agolada a Cuntis. Tyto grafy byly vytvořeny pomocí balíčku `leaflet` v R. V tuto chvíli máme k dispozici v obci Agolada 9635 jízd, z toho 5905 před instalací opatření a 2820 po instalaci (zbylé jízdy byly uskutečněny během instalace), a v obci Cuntis celkem 11196 jízd, z čehož 7901 je před instalací opatření a 2147 po instalaci. Všimněme si, že data stále obsahují záznamy mimo zkoumané úseky. Jedná se o tzv. odbočující jízdy, kde vozidla během průjezdu daným úsekem odbočila na

vedlejší komunikaci. Chceme-li z dat odbočující jízdy vyloučit, musíme pomocí Google Map manuálně určit oblasti, pokrývající nějaké záznamy těchto jízd. Na obrázku 4.3 jsou tyto oblasti vyznačeny červeně. Dále identifikujeme jízdy, které obsahují záznamy v těchto oblastech, a ty ze souboru vyloučíme. Z obce Agolada jsme takto odstranili 7 odbočujících jízd a z Cuntisu celkem 30. Výsledná očištěná data jsou znázorněna na obrázku 4.4. Body zdánlivě mimo silnici jsou již pouze chybou měření GPS. Algoritmus identifikace a následného odstranění odbočujících jízd naleznete v příloze pod názvem `odbocujici_jizdy.R`.



Obrázek 4.4: FCD vyfiltrovaného úseku silnice v obci Agolada (vlevo) a Cuntis (vpravo) očištěná od odbočujících jízd.

Zájmový úsek silnice v Agoladě měří asi 2,2 km, v Cuntisu pak asi 2,8 km, z čehož Cuntis dolní 1,4 km a Cuntis horní 1,2 km. Agregace počtu záznamů v daném úseku ukazuje tabulka 4.2. Medián v Agoladě je roven 14, v dolním Cuntisu 10 a v horním Cuntisu 8 záznamů. Povšimněme si, že průjezdy s nejmenším počtem

záznamů obsahují pouze 1 záznam, kdežto průjezd s největším počtem záznamů dokonce 58. Tato tabulka názorně poukazuje na rozdílnost jednotlivých průjezdů co se počtu bodů týče.

Úsek	Počet záznamů		
	Medián	Minimum	Maximum
Agolada	14	7	55
Cuntis dolní	10	1	58
Cuntis horní	8	1	46

Tabulka 4.2: Agregace počtu záznamů pro dané úseky.

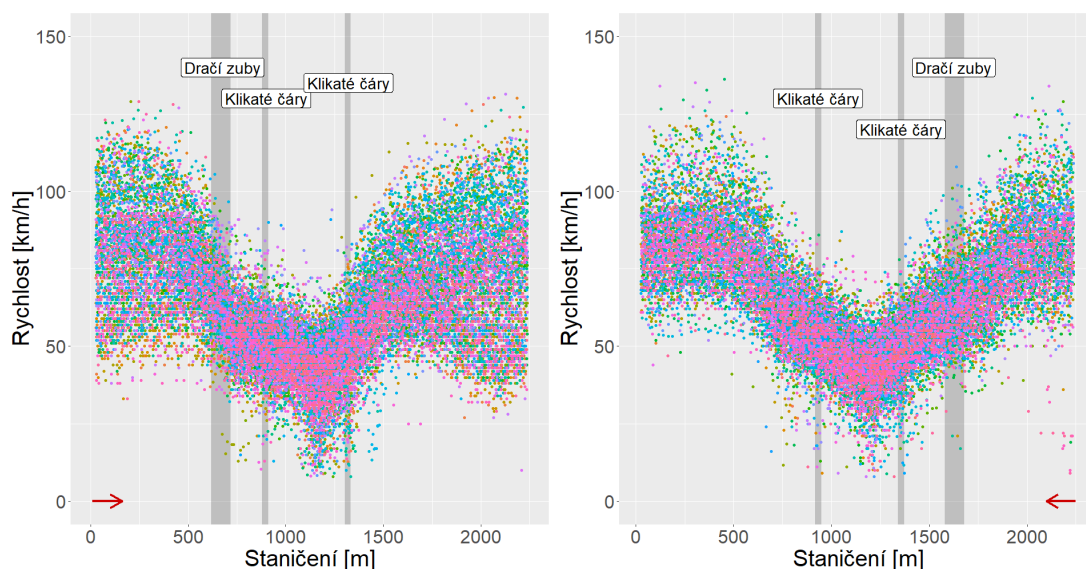
Po bližším prozkoumání jsme zjistili, že průjezd, který v dolním Cuntisu obsahuje pouze 1 záznam a v horním Cuntisu rovněž, je v datech pouze jeden a má identifikační číslo 11226. I přesto, že vyhlazování průjezdů bude v Cuntisu provedeno na celém měřeném úseku, nikoliv pro dané dva zájmové úseky separátně, je počet měření jízdy č. 11226 celkem roven 2, což považujeme za nedostatečné, a z dat proto tento průjezd vyřadíme. Ostatní průjezdy tento problém nemají.

Důležitým krokem v přípravě FCD je transformace prostorových dat do dvourozměrné podoby, kterou v této práci, jak již bylo řečeno, pouze lehce nastíníme. Pomocí GPS souřadnic se získají středové body pro danou komunikaci, které se aproximují vyhlazujícím splajnem. Každému GPS záznamu je přiřazena nejbližší hodnota ze středového splajnu. Na základě středového splajnu je poté vypočtena hodnota staničení silnice (v metrech). Jedná se o vzdálenost daného bodu od bodu výchozího. Výchozí bod je přitom pevně stanoven v místě začátku zájmového úseku. Hodnoty staničení spolu s hodnotami rychlosti vozidla dávají dvourozměrná data, s kterými budeme dále pracovat. Tento krok pre-processingu dat z Agolady a Cuntisu provedlo CDV.

Jelikož jsou v obcích přechody pro chodce, dochází v některých průjezdech k významnému přibrzdění vozidla či dokonce k zastavení. Chceme-li se zaměřit na analýzu efektu instalovaných značení, jsou pro nás podstatné spíše plynulejší

průjezdy, které nejsou něčím významně narušeny (například chodci na přechodu). Z toho důvodu z dat odstraníme jízdy, u nichž došlo k přibrždění pod 8 km/h. Zároveň jsme vyloučili jízdy přesahující rychlost 150 km/h. K problému velmi vysoké rychlosti docházelo především v horním Cuntisu, kde pár řidičů dle měření dosáhlo rychlosti až 200 km/h při výjezdu z obce. Všechny tyto extrémní jízdy by negativně ovlivňovaly další analýzu.

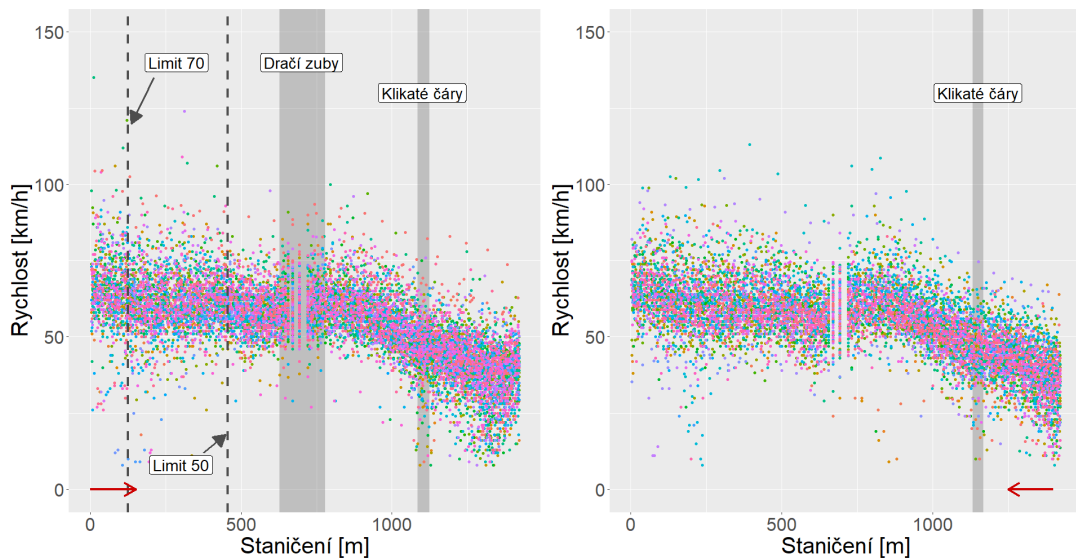
Cuntis dolní je identifikován jako úsek vozovky do hodnoty staničení 1423 m a Cuntis horní od hodnoty staničení 1600 m. Úsek dolního Cuntisu končí těsně za přechodem pro chodce a Cuntis horní končí výjezdem z obce.



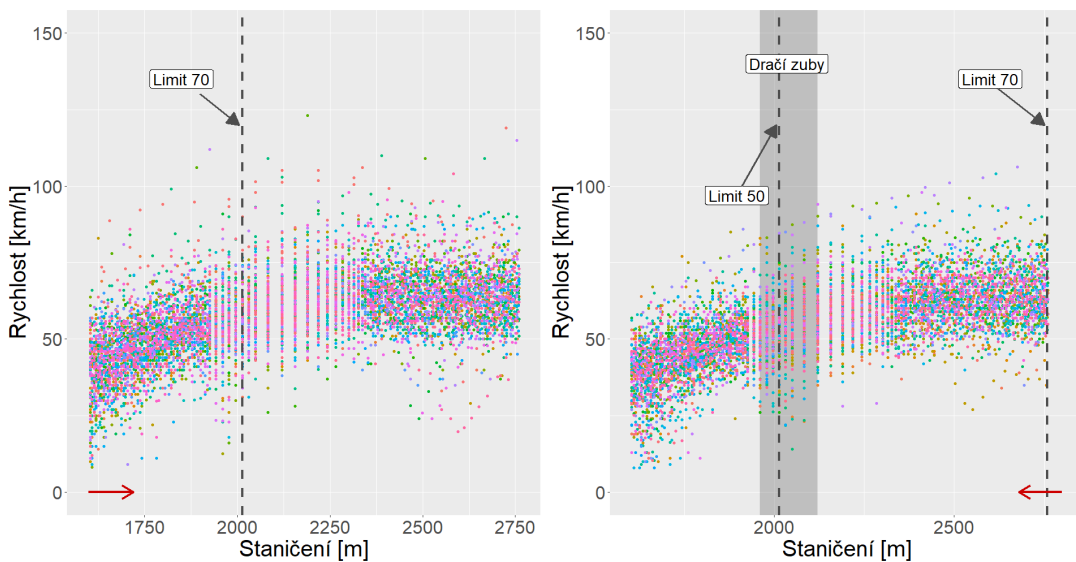
Obrázek 4.5: Nevyhlazené průjezdy Agoladou po instalaci opatření v jednotlivých směrech se zaznačenými opatřeními.

Po finálním očištění dat máme v Agoladě k dispozici celkem 9256 jízd, z čehož 5671 před a 2713 jízd po instalaci opatření. V Cuntisu je pak 8012 jízd, z toho 5697 před a 1502 po instalaci. Data z Agolady po instalaci opatření v jednotlivých směrech znázorňuje graf na obrázku 4.5. Směr průjezdu bude na každém grafu zaznačen červenou šipkou v dolním rohu. Průjezd vozidel ve směru 1 obsahuje rostoucí staničení a průjezd ve směru 2 obsahuje naopak staničení klesající. V grafech jsou zároveň zaznačeny úseky vozovky, v nichž se po instalaci nachá-

zejí daná dopravní značení a umístění přidanych značek. Úsek dolního Cuntisu je zobrazen na obrázku 4.6 a nakonec úsek horního Cuntisu na obrázku 4.7.



Obrázek 4.6: Nevyhlazené průjezdy dolním Cuntisem po instalaci opatření v jednotlivých směrech se zaznačenými opatřeními.



Obrázek 4.7: Nevyhlazené průjezdy horním Cuntisem po instalaci opatření v jednotlivých směrech se zaznačenými opatřeními.

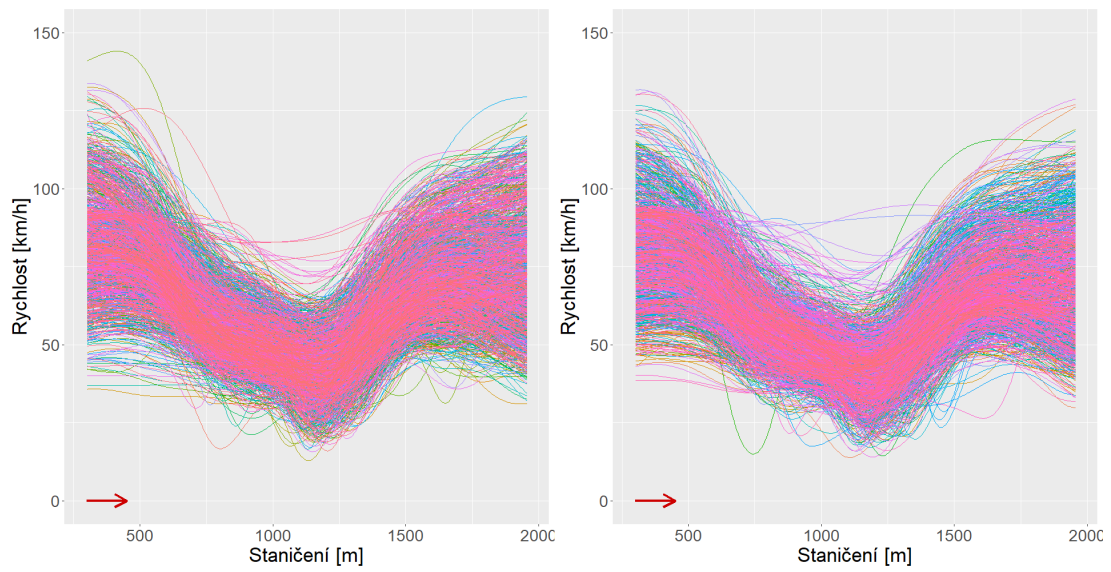
4.3. FCD jakožto funkcionální data

Data příslušných úseků vozovky již máme nachystána a očištěna, a nyní přejdeme k jejich transformaci do funkcionální podoby. Cílem bude pro každou obec, resp. zájmový úsek, vytvořit kompaktní funkcionální objekt. Pro všechny průjezdy daného úseku tak budeme užívat shodné parametry vyhlazování.

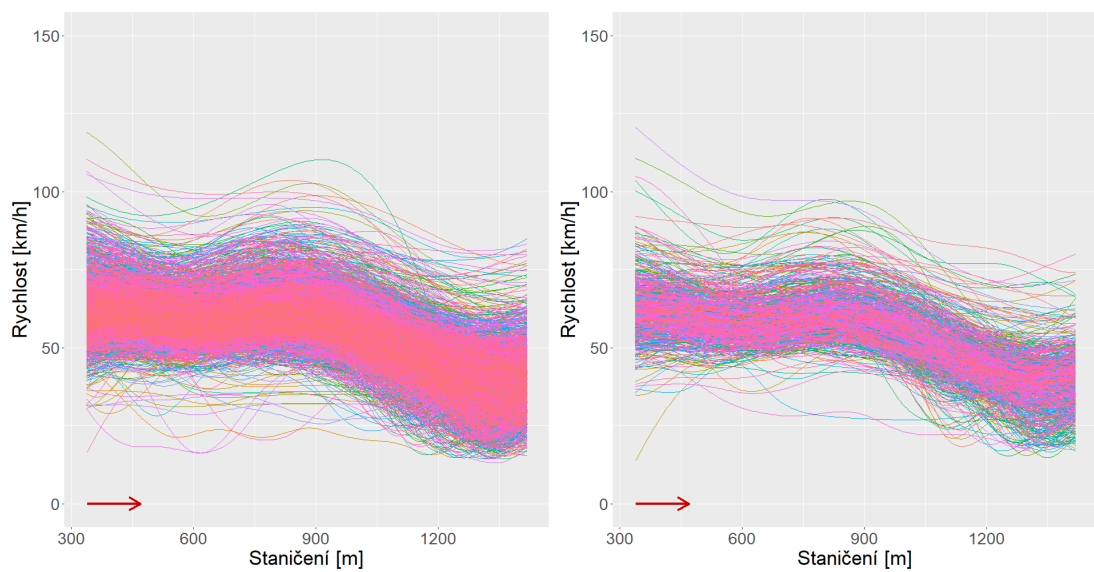
Jelikož průjezdy mají obecně různý počet záznamů a různé startovací a cílové hodnoty staničení, potřebujeme jejich definiční obory v rámci daného úseku sjednotit. Pro každý úsek tedy určíme průnik definičních oborů průjezdů a dále již budeme pracovat s průjezdy na těchto intervalech staničení. Pro obec Agolada je výsledný definiční obor [301,24; 1956,24] metrů, dolní Cuntis [338,10; 1418,10] metrů a horní Cuntis [1603,10; 2378,10] metrů.

K vyhlazení průjezdů obcemi aplikujeme metodu regularizace, přičemž pro bázeovou expanzi použijeme kubické B -splajny. Fourierovu bázi nepoužijeme z důvodu neperiodičnosti dat. Uzly s ohledem na obecnost sestaveného algoritmu pro vyhlazování volíme ekvidistantně a počet bázeových funkcí na pevnou hodnotu 100. Při práci s FCD se ukázalo, že takto nastavené parametry vyhlazování dávají uspokojivé výsledky. Stejně tak je tomu s hodnotou vyhlazovací konstanty, již jsme zvolili fixně na $\lambda = 10^5$. Výsledné vyhlazené funkce evaluujeme v bodech po 5 metrech staničení. Průjezdy obcí Cuntis budeme za účelem větší přesnosti vyhlazovat na celém měřeném úseku, tj. v hodnotě staničení [338,10; 2378,10]. Teprve až následně úsek rozdělíme na dolní a horní Cuntis.

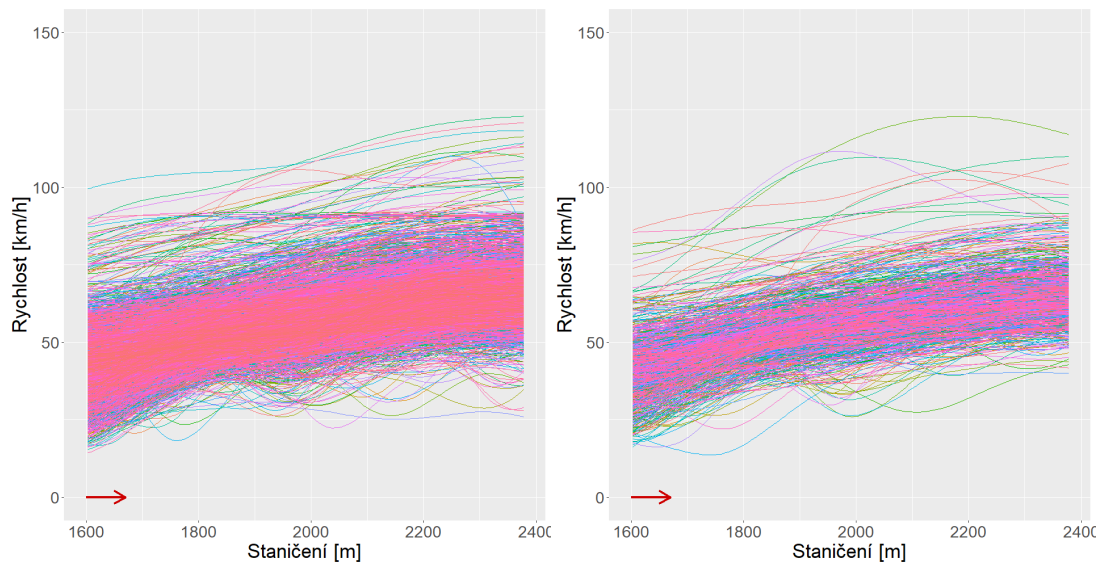
Vyhlazené průjezdy obcí Agolada před a po instalaci dopravního značení ukazuje graf na obrázku 4.8, dolním Cuntisem na obrázku 4.9 a horním Cuntisem na obrázku 4.10.



Obrázek 4.8: Vyhlazené průjezdy Agoladou ve směru 1. Vlevo: před instalací. Vpravo: po instalaci opatření.



Obrázek 4.9: Vyhlazené průjezdy dolním Cuntisem ve směru 1. Vlevo: před instalací. Vpravo: po instalaci opatření.



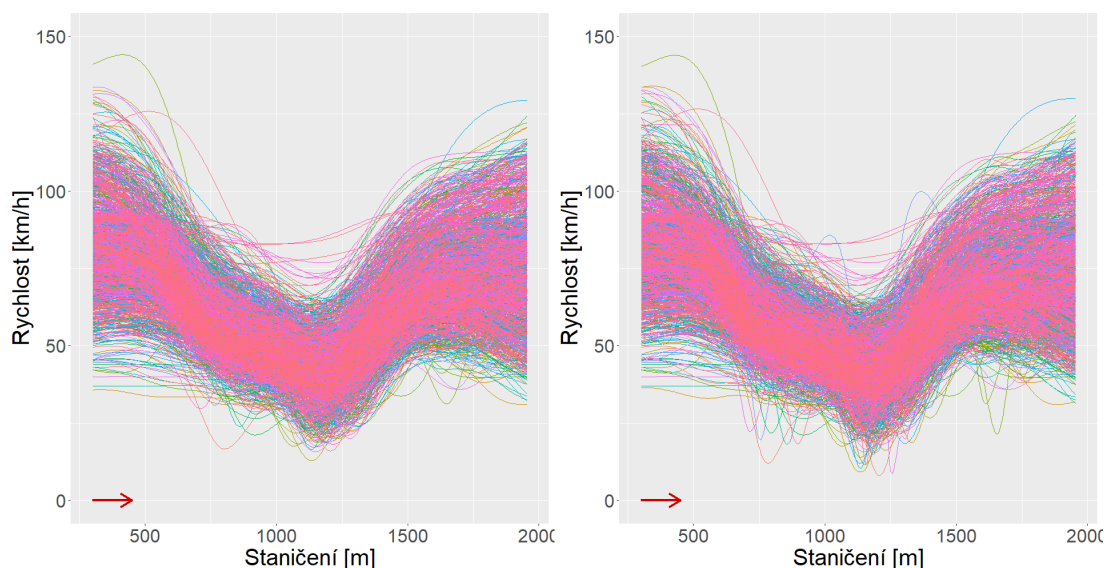
Obrázek 4.10: Vyhlazené průjezdy horním Cuntisem ve směru 1. Vlevo: před instalací. Vpravo: po instalaci opatření.

Pro optimální volbu parametru pro každý průjezd obcí bylo též vyzkoušeno použití zobecněné křížové validace. Porovnání lze vidět na obrázku 4.11, kde graf vlevo znázorňuje vyhlazení daných průjezdů Agoladou s vyhlazovací konstantou $\lambda = 10^5$ a graf vpravo pak vyhlazení stejných průjezdů s použitím GCV pro každý z nich. GCV přitom volila ze sekvence hodnot 10^{-5} až 10^5 s kroky 0,5. Při porovnání grafů je patrné, že vyhlazení pomocí GCV je menší a lépe reflektující naměřená data. Pro další analýzu bychom však žádali větší vyhlazení průjezdů. Z toho důvodu byla fixně nastavena větší hodnota $\lambda = 10^5$.

Proces vyhlazování byl proveden na MetaCentru, resp. MetaVO¹, z důvodu výpočetní náročnosti úlohy. Pro zajímavost, vyhlazení všech průjezdů danými třemi zájmovými úseky s fixní hodnotou vyhlazovacího parametru λ trvalo MetaCentru 17 minut. Pokud jsme však do algoritmu zakomponovali volbu λ zobecněnou křížovou validací, výpočet úlohy trval 5 hodin. Také v tomto případě

¹MetaVO je jedna z virtuálních organizací české Národní Gridové Iniciativy MetaCentrum NGI a je otevřená všem akademickým pracovníkům a studentům členů sdružení CESNET. Uživatelé registrovaní v MetaVO mají možnost bezplatného využití výpočetní a úložné kapacity [13].

vidíme, že ne pro každou praktickou úlohu musí být sofistikovanější přístup tím nejvhodnějším.



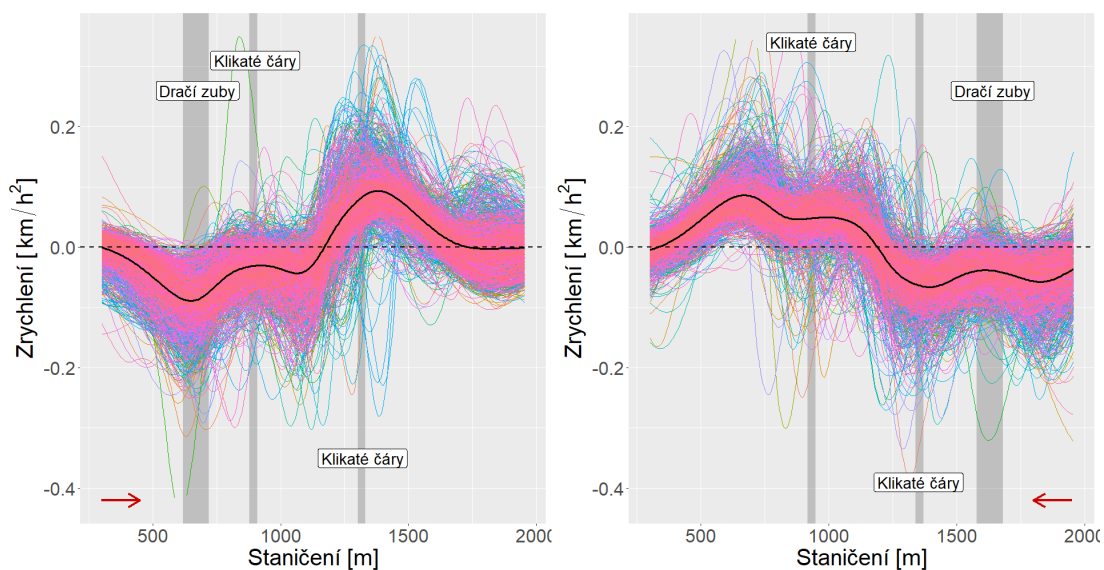
Obrázek 4.11: Porovnání vyhlazení průjezdů Agoladou ve směru 1 před instalací opatření. Vlevo: parametr $\lambda = 10^5$. Vpravo: λ pro každý průjezd zvolen pomocí GCV.

Ze softwaru R byly v zkonstruovaném algoritmu pro vyhlazování použity funkce z balíčku `fda` [17]. Například pro tvorbu B -splajnové báze jsme použili funkci `create.bspline.basis()` a pro vyhlazení `smooth.basis()`. Kód vytvořený pro vyhlazování naleznete v příloze pod názvem `vyhlazovani.R` a jeho modifikaci se zabudovanou GCV pod názvem `vyhlazovani_GCV.R`.

Pomocí první derivace rychlosti popíšeme průměrnou akceleraci vozidel při průjezdu jednotlivými úseky. Akceleraci vozidel v Agoladě v daných směrech zobrazuje graf na obrázku 4.12. Zaměříme se na průměr ve směru 1. Před vjezdem do obce Agolada řidiči postupně začali zpomalovat. Během průjezdu dračími zuby se zpomalení začíná zintenzivňovat, neboť následují klikaté čáry, a tím také přechod pro chodce. Zhruba v polovině přechodu pro chodce řidiči v průměru začínají zrychlovat, přičemž poté toto zrychlení postupně až do výjezdu z Agolady slábnou. Ve směru 2 lze pozorovat podobný trend. Při vjezdu do obce ve směru 2

však k intenzivnějšímu zpomalení vozidel v průměru docházelo již před dračími zuby.

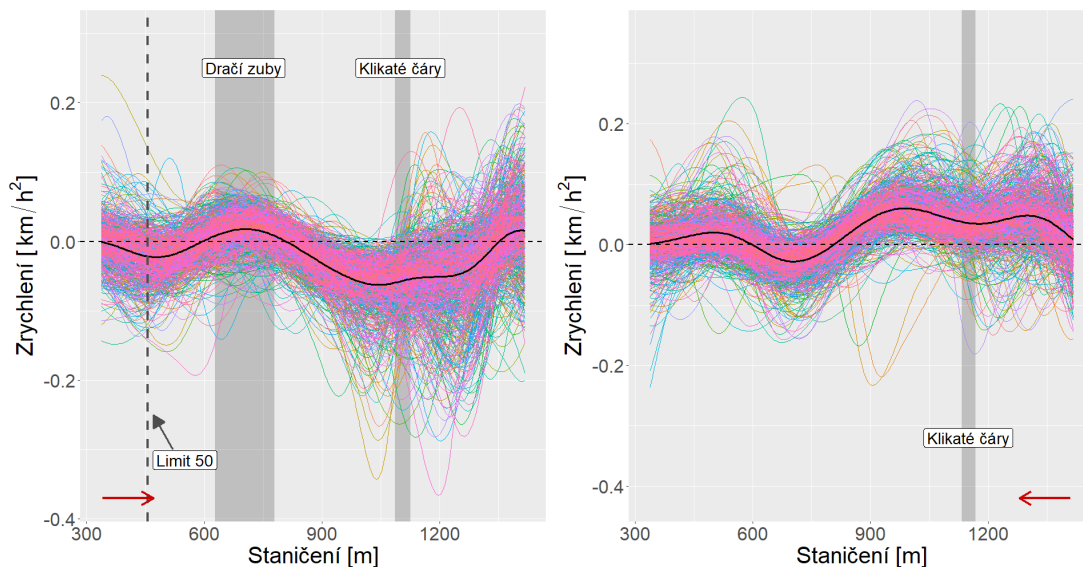
Akceleraci vozidel v dolním Cuntisu v daných směrech zobrazuje graf na obrázku 4.13. Opět se zaměříme nejprve na směr 1. Již zhruba 300 metrů před vjezdem do dolní části obce Cuntis řidiči začali mírně zpomalovat, nejspíše kvůli rychlostnímu limitu 70 km/h. Zpomalení se zintenzivňovalo po příjezdu ke značce rychlostního limitu 50 km/h. Zajímavé je, že při průjezdu dračími zuby řidiči zase začali mírně zrychlovat a až teprve po jejich přejezdu začali znovu zpomalovat. Toto zpomalení dále bylo v průměru takřka konstantní. Teprve až v polovině přechodu pro chodce dochází ke zrychlení. Akcelerace ve směru 2 má srovnatelný průběh.



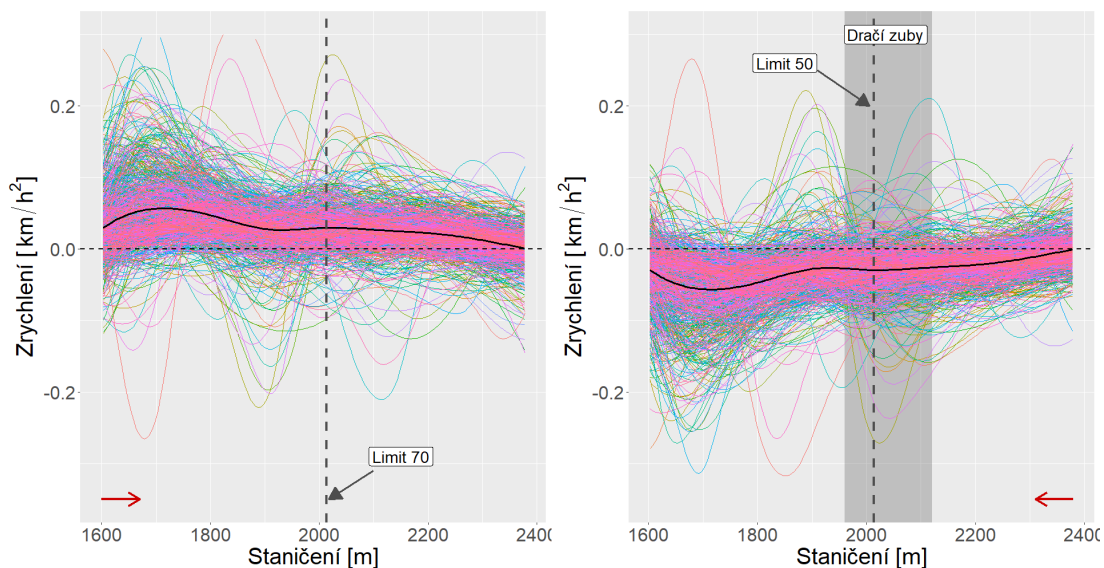
Obrázek 4.12: Akcelerace vozidel při průjezdu Agoladou po instalaci opatření v jednotlivých směrech a průměrná akcelerace (černá plná čára).

Jaký je průběh akcelerace vozidel v horním Cuntisu v jednotlivých směrech ukazuje graf na obrázku 4.14. Ve směru 1 vozidla v tomto úseku drží pozitivní akceleraci, neboť je zhruba ve třetině úseku výjezd z obce. Toto zrychlení se však od přejezdu značkou limitu 70 km/h, jež se nachází pár desítek metrů za značkou obce, mírně snižuje. Ve směru 2 je naopak průměrná akcelerace po celou dobu

průjezdu úsekem negativní, jelikož řidiči přijíždějí ke dračím zubům a rychlostnímu limitu 50 km/h. Zpomalení vozidel neustále slábne a až ke konci úseku dochází k jeho zesílení.

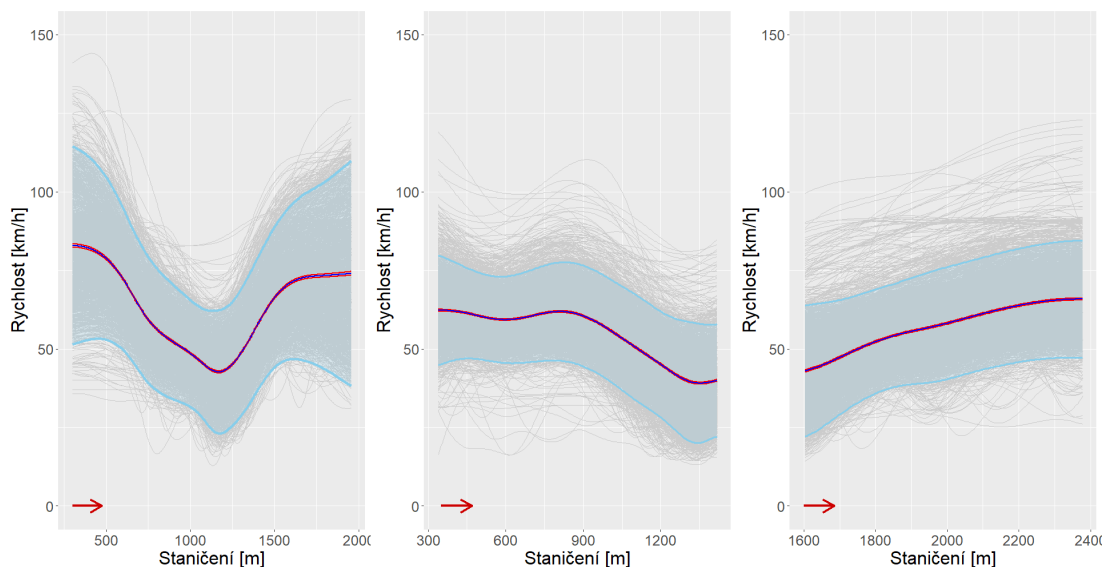


Obrázek 4.13: Akcelerace vozidel při průjezdu dolním Cuntisem po instalaci opatření v jednotlivých směrech a průměrná akcelerace (černá plná čára).



Obrázek 4.14: Akcelerace vozidel při průjezdu horním Cuntisem po instalaci opatření v jednotlivých směrech a průměrná akcelerace (černá plná čára).

Průměrné průjezdy obcemi ve směru 1 před instalací značení společně s variabilitou dat zobrazují grafy na obrázku 4.15. Variabilitu průjezdů obcemi vizualizujeme pomocí pásu kolem funkce průměru, který se bodově spočítá jako součet, resp. rozdíl, dvojnásobku směrodatné odchylky od průměru. Odhadem 95 % dat by se tak mělo v tomto pásu nacházet. Zároveň tyto grafy obsahují také bodový 95% pás spolehlivosti pro jednotlivé funkce průměru zkonstruovaný bootstrapovou metodou, kterou jsme popsali v kapitole 1.3. Lze si povšimnout, že všechny zkonstruované pásy spolehlivosti jsou velmi úzké a jejich hranice kopírují danou funkci průměru velmi těsně, až s ní splývají. To lze zdůvodnit značným počtem dat, díky němuž jsme si s odhadem funkce střední hodnoty velmi jisti. Všimněme si dále, že mezi průjezdy úsekem horního Cuntisu je značný počet extrémních jízd, které se velmi odlišují od jízdy průměrné, a nepokrývá je ani zkonstruovaný modrý pás. Někteří řidiči projeli tímto úsekem poměrně vysokou rychlostí navzdory tomu, že zhruba do třetiny úseku platí rychlostní limit 50 km/h (část úseku v obci) a hned poté limit 70 km/h.

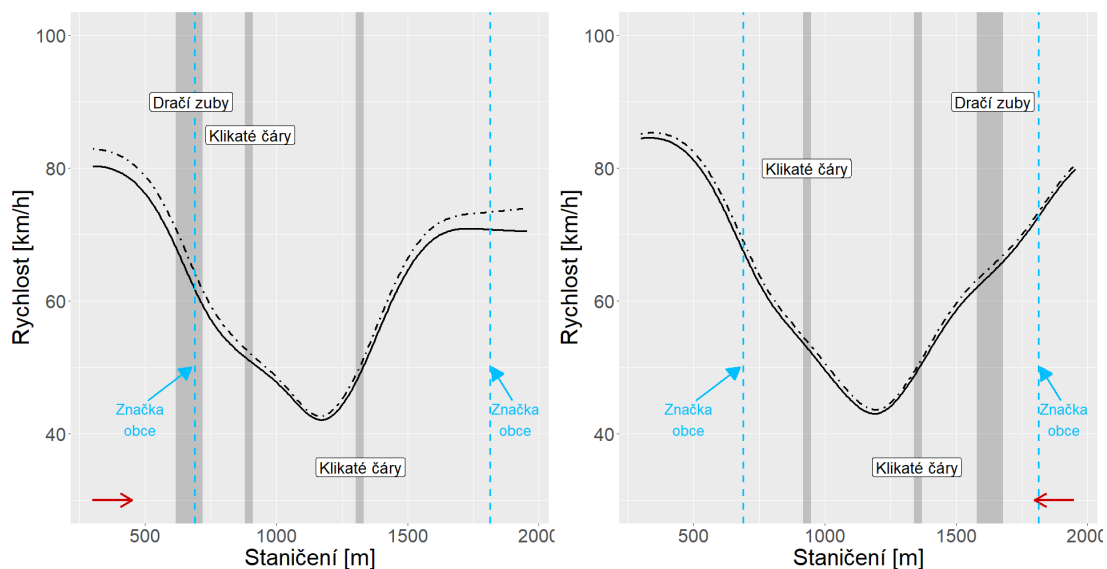


Obrázek 4.15: Výběrové charakteristiky průjezdů obcí Agolada, dolní a horní Cuntis ve směru 1. Modře: funkce průměru, růžově: 95 % bodový pás spolehlivosti funkce průměru (téměř splývá), světle modře: pás funkce průměru ± 2 směrodatné odchylky.

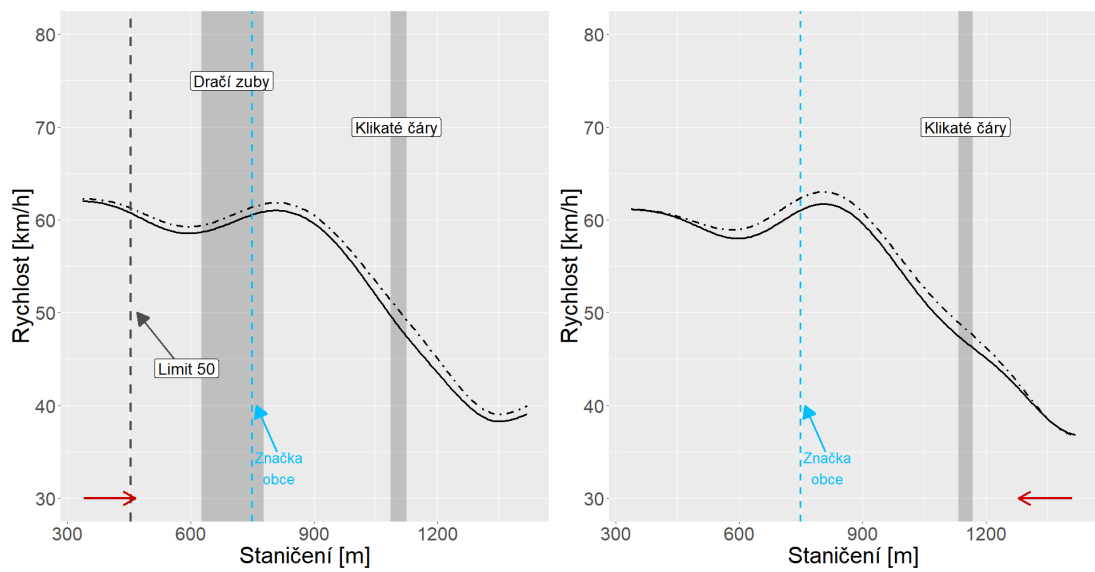
4.4. Analýza efektu instalovaných opatření

V této sekci se budeme zabývat klíčovou částí analýzy, a to samotnou analýzou efektu instalovaných značení v obcích Agolada a Cuntis. Použijeme k tomu metody uvedené v teoretické části této práce.

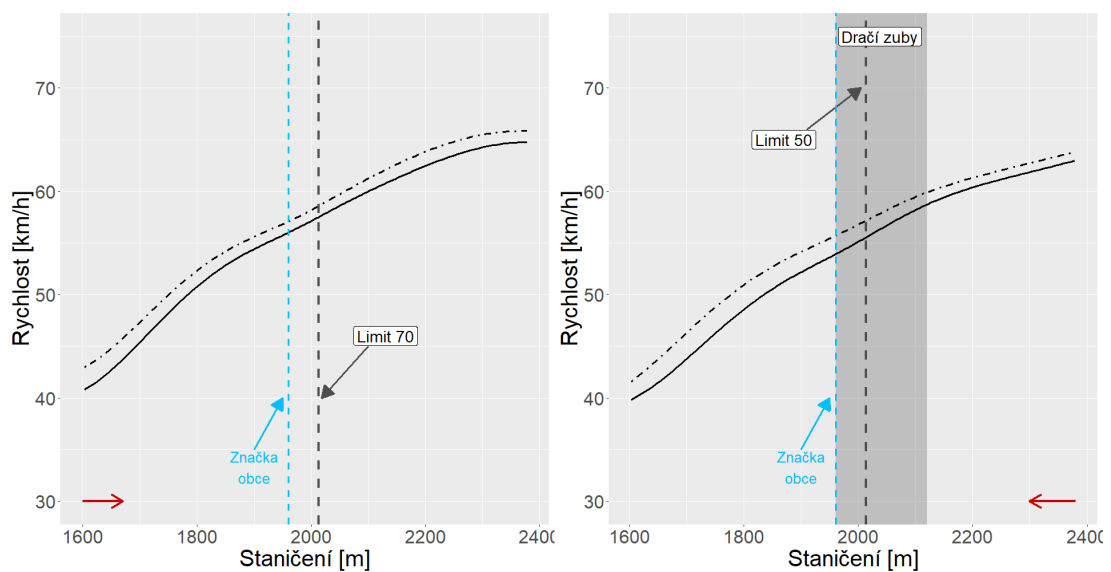
Klíčovou roli budou v tuto chvíli hrát především funkce průměru průjezdů danou obcí před a po instalaci opatření jakožto odhady funkcí středních hodnot těchto výběrů. Nejprve si funkce průměru srovnáme graficky. Pro obec Agolada a jednotlivé směry jízdy jsou průměrné průjezdy v závislosti na výskytu opatření graficky znázorněny na obrázku 4.16. Ve směru 1 můžeme po instalaci pozorovat viditelné snížení rychlosti vozidel především na krajích úseku, tedy při vjezdu a výjezdu z obce. Lze se domnívat, že ke zpomalení při příjezdu do obce došlo právě v důsledku instalovaných dračích zubů. Otázkou zůstává, proč ke snížení rychlosti vozidel došlo při výjezdu z obce, poněvadž v těchto místech žádné opatření instalováno nebylo. Mírné zpomalení řidičů pozorujeme dále na zbylém úseku. Ve směru 2 došlo k mírnému snížení rychlosti po celou dobu tohoto úseku.



Obrázek 4.16: Průměrné průjezdy Agoladou v jednotlivých směrech. Čárkované: před instalací dopravních opatření. Plně: po instalaci.



Obrázek 4.17: Průměrné průjezdy dolním Cuntisem v jednotlivých směrech. Čárkovaně: před instalací dopravních opatření. Plně: po instalaci.



Obrázek 4.18: Průměrné průjezdy horním Cuntisem v jednotlivých směrech. Čárkovaně: před instalací dopravních opatření. Plně: po instalaci.

Podobný scénář jako v obci Agolada se odehrává rovněž i na úseku dolního a horního Cuntisu. Jak ukazuje graf na obrázku 4.17, v obou směrech došlo ke snížení rychlosti vozidel prakticky na celém úseku. Stejná situace je potom také v horním Cuntisu, viz obrázek 4.18, kde lze pozorovat poměrně výrazné zpomalení

řidičů.

Pojďme se, stále zatím pouze formou popisné statistiky, podívat na příjezdy řidičů do obcí v závislosti na výskytu dopravního značení. Umístění příjezdových značek do dané obce je znázorněno rovněž na grafech funkcí průměru 4.16, 4.17 a 4.18. Značka obce se při příjezdu do Agolady ve směru 1 nachází v hodnotě staničení 688,8 m (po instalaci se v těchto místech nachází i dračí zuby). Vozidla se, dle vyhlazení, před instalací dračích zubů pohybovala v tomto místě průměrnou rychlostí 64 km/h a po instalaci 61,5 km/h, což znamená snížení rychlosti o 3,9 %. Ve směru 2 se v místě značky obce pohybovala vozidla před instalací dračích zubů rychlostí 73,2 km/h a po instalaci 72,6 km/h (nepatrné zpomalení o 0,8 %). Při příjezdu do dolního Cuntisu (směr 1) se nachází značka obce v hodnotě staničení 749 m. Řidiči jeli v těchto místech před instalací dračích zubů průměrnou rychlostí cca 61,4 km/h a po instalaci 60,7 km/h (zpomalení o 1,2 %). Při příjezdu do horního Cuntisu (směr 2) se značka obce nachází v hodnotě staničení 1960,4 m. Vozidla se pohybovala v těchto místech před instalací dračích zubů průměrnou rychlostí cca 55,6 km/h a po instalaci 53,8 km/h (zpomalení o 3,2 %). Z těchto údajů se lze domnívat, že dračí zuby, případně značky stanovující rychlostní limit, na zpomalení řidičů při vjezdu do obcí mohly zafungovat. V některých místech se ovšem jedná v průměru o velmi nepatrné zpomalení, které nemusíme považovat za statisticky významné, nýbrž náhodné.

Úsek	Směr 1			Směr 2		
	Před	Po	Rel. rozdíl	Před	Po	Rel. rozdíl
Agolada	63,48	61,53	-3,07 %	63,97	63,04	-1,45 %
Cuntis dolní	54,67	53,88	-1,45 %	54,65	53,76	-1,63 %
Cuntis horní	57,13	55,93	-2,10 %	55,38	53,74	-2,96 %

Tabulka 4.3: Srovnání průměrných rychlostí v rámci celých úseků v jednotlivých směrech před a po instalaci opatření (km/h).

Jak tato opatření spolu se zavedenými klikatými čarami uvnitř obce fungují komplexně, ukáží výsledky globálního testu, jenž jsme popsali v kapitole 3.1. V tabulce 4.3 uvádíme ještě srovnání průměrných rychlostí v rámci celých úseků před a po instalaci značení. Můžeme si tak udělat velmi hrubou představu o efektu opatření v úseku z komplexního hlediska. Všimněme si, že ve všech úsecích a obou směrech došlo ke snížení průměrné rychlosti řidičů, což jsme ostatně již pozorovali na grafech průměrných průjezdů diskutovaných výše. K největšímu celkovému zpomalení došlo v Agoladě ve směru 1 a v horním Cuntisu ve směru 2, kde vozidla snížila rychlost průměrně o asi 3 %. I to ovšem považujeme za ne příliš významný rozdíl.

Konečně se pojdme podívat na výsledky globálního testu. Přesvědčíme se, zda rozdíly, které jsme ve funkcích průměru dosud pozorovali pouze popisnou statistikou, jsou (globálně) dostatečně signifikantní. Použili jsme algoritmus globálního testu z kapitoly 3.1, a tedy již zmíněnou funkci `tperm.fd()` z knihovny `fda` v R. Testujeme nulovou hypotézu shody funkcí středních hodnot daných výběrů. V našem problému tyto výběry definujeme pomocí výskytu určitých dopravních opatření. Jelikož se jedná o permutační test, nastavili jsme počet permutací v hodnotě 1000. Test je proveden na hladině významnosti 5 %. Kód k použití testu naleznete v příloze pod názvem `testy.R`.

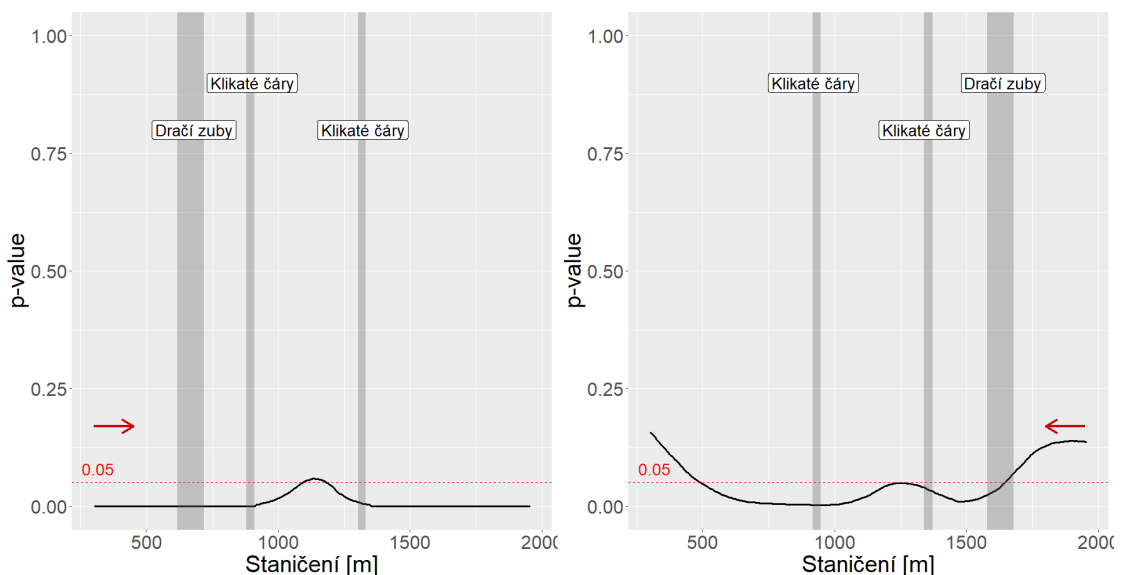
Výsledky globálního testu pro jednotlivé úseky a oba jejich směry ukazuje tabulka 4.4. Vidíme, že ve všech analyzovaných úsecích se prokázal statisticky významný globální efekt nainstalovaných opatření, a to s velmi malou p-hodnotou. Nulovou hypotézu shody funkcí středních hodnot tedy zamítáme velmi jistě. Pro některé úseky může být tento výsledek lehce překvapující, poněvadž grafy funkcí průměru v některých případech sice zaznamenaly jistý efekt, avšak na první pohled ne tak významný, viz např. ve směru 2 obce Agolada (graf na obrázku 4.16). Na základě tohoto zjištění lze globální test posoudit za silný, protože zachytil i slabší efekt opatření. Domníváme se, že síla tohoto testu plyne z velkého počtu dat, a tím také velmi jistých odhadů.

Úsek	p-hodnota	
	Směr 1	Směr 2
Agolada	< 0,001	< 0,001
Cuntis dolní	< 0,001	< 0,001
Cuntis horní	< 0,001	< 0,001

Tabulka 4.4: Výsledek globálního testu pro dané úseky v jednotlivých směrech.

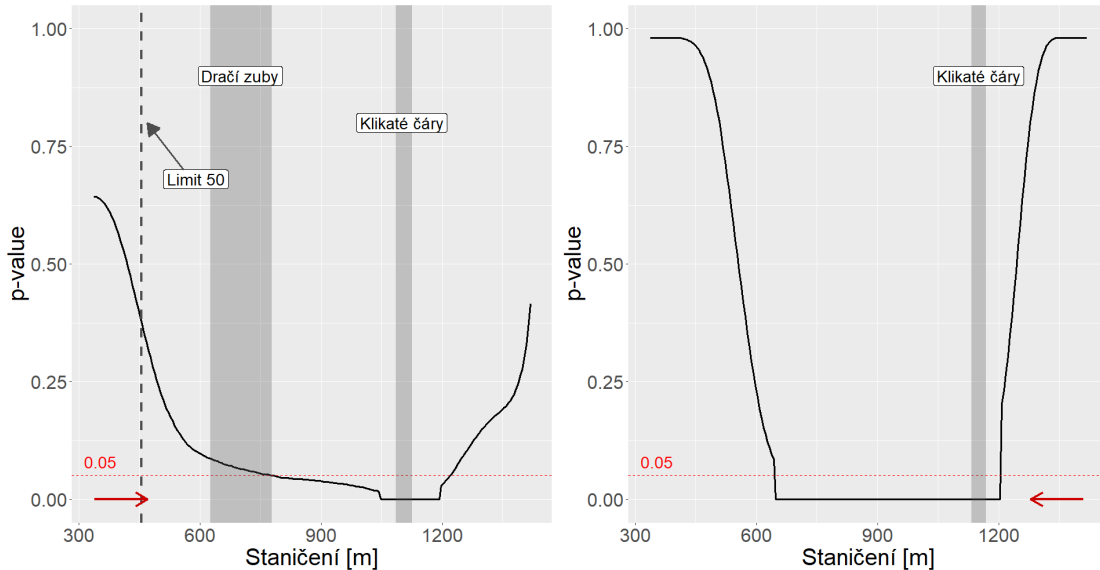
V tuto chvíli již víme, že z globálního hlediska došlo po instalaci opatření k statisticky významnému snížení rychlosti řidičů při průjezdu obcemi. Mohlo by se tak jevit, že instalovaná opatření mají význam. Jak si ale ukážeme později, toto hodnocení není tak jednoduché.

Nyní se dále podíváme na výsledky intervalového testu, v jehož případě se dozvíme, v jakých částech úseku k významnému rozdílu v průměrných průjezdech došlo. Mohli bychom tak zjistit, zda komplexně fungují všechna opatření v úseku či pouze některá z nich. Algoritmus tohoto testu jsme popsali v sekci 3.2 a pro jeho výpočet jsme použili již zmíněnou funkci `ITP2bspline()` a k tomu počet permutací roven 10000. Kód k intervalovému testu rovněž naleznete v příloze pod názvem `testy.R`.



Obrázek 4.19: Výsledek intervalového testu pro Agoladu v jednotlivých směrech.

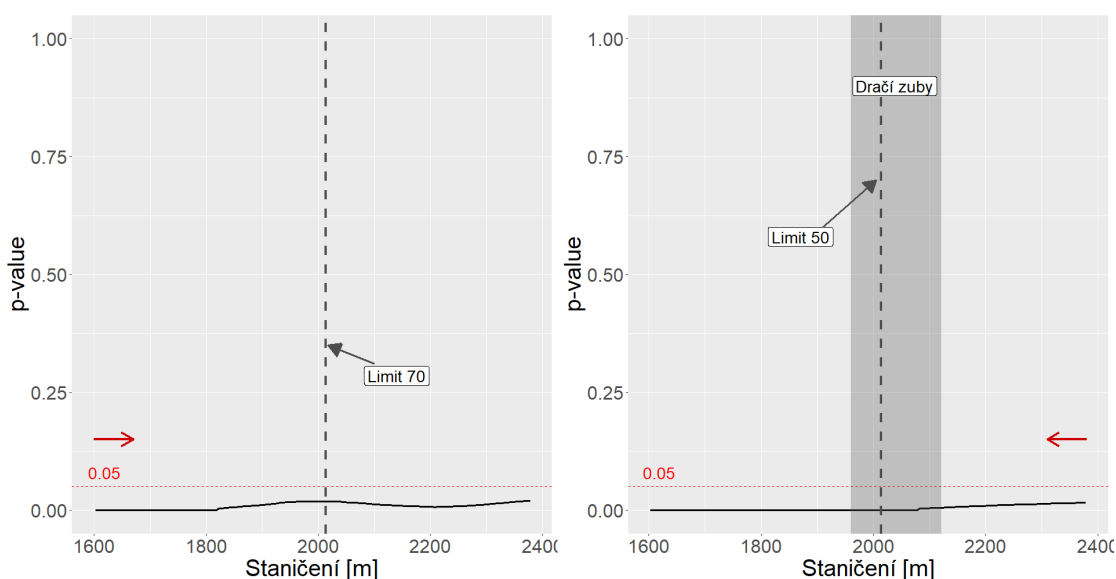
Intervalový test pro všechny analyzované úseky a jejich směry byl kvůli výpočetní náročnosti proveden na MetaCentru, kde jeho výpočet trval asi 2 hodiny. Výstupem jsou p-hodnoty dle staničení. V těch lokalitách úseku, kde p-hodnota klesla pod hodnotu 0,05, test zaznamenal signifikantní rozdíl mezi funkcemi průměru. Výsledky pro obec Agolada ukazují grafy na obrázku 4.19. Ve směru 1 byl zaznamenán významný efekt opatření téměř na celém úseku. V místě úseku mezi dvěma přechody pro chodce, tzn. mezi dvěma značeními klikatých čar, je již efekt méně významný a hraničí se signifikantností. Tento výsledek názorně reflektuje také graf funkcí průměru 4.16. Co se týče směru 2, efekt zpomalení se ukázal uvnitř obce, při vjezdu a výjezdu z obce naopak nikoliv. Shodnou interpretaci s obrázkem 4.16 v tomto případě nepozorujeme. Uvnitř obce se dle této vizualizace zdají být funkce průměru stejně odlišné jako při vjezdu a výjezdu z obce. Výsledek intervalového testu tak může odrážet větší variabilitu dat právě na krajích tohoto úseku.



Obrázek 4.20: Výsledek intervalového testu pro dolní Cuntis v jednotlivých směrech.

Pro úsek dolního Cuntisu výsledek intervalového testu znázorňují grafy na obrázku 4.20. Zaměříme se nejprve na směr 1. Při vjezdu do obce test zachy-

til významné zpomalení řidičů až při přejíždění dračími zuby. V místě začátku přechodu pro chodce, tzn. těsně před klikatými čarami a po nich, se pak ukázal efekt těchto varovných čar. Po zbytek úseku další pozitivní efekt nepozorujeme. Ve směru 2 dolního Cuntisu vidíme, že klikaté čáry před přechodem pro chodce měly vliv na významné zpomalení řidičů. Tento efekt dle testu trval odhadem dalších 500 metrů. Oba tyto výsledky odráží naše pozorování ve vizualizaci funkcí průměru na obrázku 4.17.



Obrázek 4.21: Výsledek intervalového testu pro horní Cuntis v jednotlivých směrech.

V horním Cuntisu intervalový test prokázal signifikantní rozdíl ve funkcích průměru po celý úsek, a to v obou směrech. Dle obrázku 4.18 není tento výsledek překvapující. Otázkou k zamyšlení je, zda pozorované zpomalení řidičů je důsledkem pouze instalovaných opatření. Ve směru 1 byla přidána značka rychlostního limitu 70 km/h v hodnotě staničení cca 2000 m, snížení rychlosti průjezdů ovšem pozorujeme již od začátku úseku, tedy od staničení 1600 m, což je 400 metrů před značkou. Z tohoto důvodu se můžeme domnívat, že do analýzy vstupuje kromě přidaného opatření nějaký další faktor, který ke zpomalení řidičů vedl. Podobnou situaci můžeme vidět i v druhém směru. Dračí zuby začínají na hodnotě

staničení cca 2110 m, což je asi o 270 metrů dále, než odkud test zaznamenává významné snížení rychlosti. V tomto případě to však můžeme vysvětlit jako pozůstatek efektu přidané dopravní značky rychlostního limitu 70 km/h zhruba 800 metrů před vjezd do obce (640 metrů před začátkem zubů).

Z výsledků testů a grafických vizualizací můžeme říci, že dle naměřených FCD došlo na základě instalace daných dopravních značení k prokazatelnému snížení rychlosti vozidel ve všech analyzovaných úsecích.

Současně, především výsledek intervalového testu pro FCD z úseku horního Cuntisu nás vede k zamyšlení nad výskytem dalších vnějších faktorů, které analýzu mohou nežádoucně intervenovat a pozorované snížení rychlosti vozidel tak nemusí být nutně pouze důsledkem instalace opatření. Takovým významným faktorem může být sezónnost.

4.5. Problematika sezónnosti FCD

Během analýzy efektu opatření se nabídl myšlenka vlivu sezónnosti dat z plovcích vozidel na výsledky analýzy. Předpokládáme, že v praxi je rychlost průjezdu vozidel ovlivněna mimo jiné počasím, s čímž souvisí roční období. Z hlediska silničního provozu lze rozdělit období roku na tyto čtyři sezóny

- zimní (Z) – prosinec, leden, únor, březen,
- jarní (J) – duben, květen, červen,
- prázdninová (Pr) – červenec, srpen,
- podzimní (Po) – září, říjen, listopad.

Například můžeme říci, že řidiči se na vozovce chovají jinak v zimním období a v období jarním (zvláště pozdním). V zimě ovlivňuje dopravu především stav vozovky (např. náledí a sníh) a řidiči tomu přizpůsobují svoji rychlost. Specifická je potom sezóna prázdnin, kdy obvykle bývá slunečné počasí a slabší provoz, tudíž

by se především mimo obce dalo očekávat zvýšení rychlosti. Z toho můžeme usoudit, že pokud bychom, hypoteticky, sbírali FCD před instalací opatření například jen v podzimním období, a po instalaci naopak pouze v období zimním, testem prokázané signifikantní snížení rychlosti by mohlo být způsobeno pouze změnou sezóny (horším počasím) a nikoliv samotným efektem opatření. Tuto skutečnost je proto potřeba při analýze zohlednit.

Připomeňme, že před instalací dopravních opatření máme k dispozici data za období leden až srpen 2022 a po instalaci data z října 2022 až března 2023 (včetně). Z období před instalací máme tedy pokrytu sezónu zimní (částečně), jarní a prázdninovou, po instalaci však pouze podzimní a zimní. Vzhledem k tomu, že v těchto dvou výběrech dat nemáme pokryté shodné sezóny, může být výsledek zkreslen.

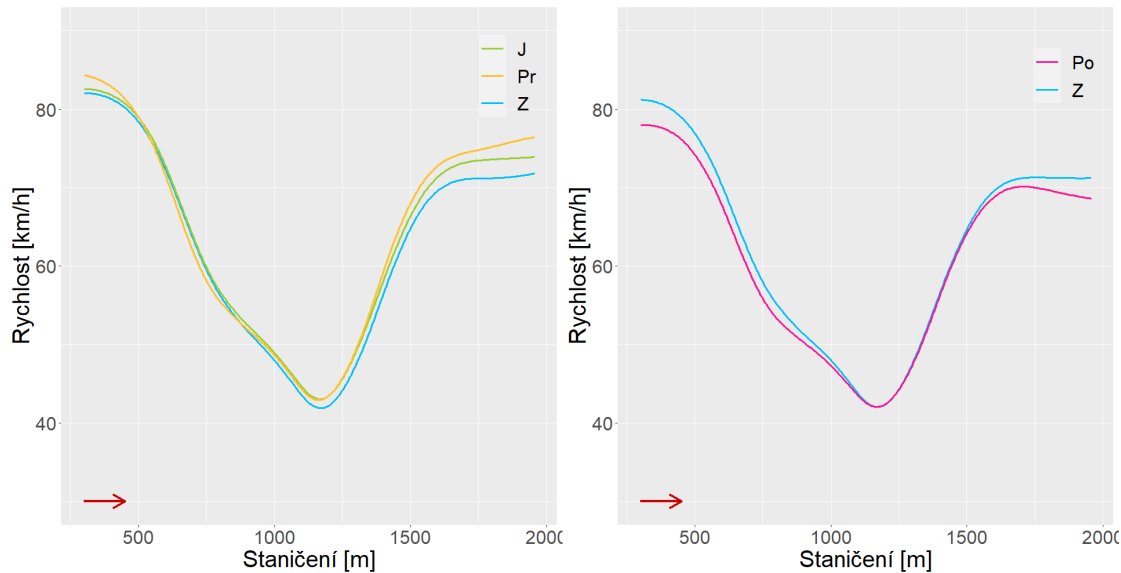
Sezóna	Agolada				Cuntis			
	Směr 1		Směr 2		Směr 1		Směr 2	
	Před	Po	Před	Po	Před	Po	Před	Po
Z	841	909	831	846	998	7	826	0
J	1328	0	1244	0	1295	0	1062	0
Po	0	498	0	460	0	784	0	711
Pr	814	0	613	0	820	0	696	0

Tabulka 4.5: Počty průjezdů obcemi v jednotlivých sezónách při daném období a směru.

Zdali je mezi sezónními provozny natolik signifikantní rozdíl, aby byla z tohoto důvodu analýza efektu opatření zkreslena, se přesvědčíme globálním testem. V datech, které máme k dispozici, můžeme otestovat odlišnost provozů v zimě vs. na jaře, v zimě vs. o prázdninách, na jaře vs. o prázdninách a v zimě vs. na podzim. Ostatní varianty otestovat vzhledem k absenci dat nemůžeme. Aby analýzu sezónnosti neovlivňovaly jiné faktory, provedeme ji na průjezdech týkající se pouze daného směru a daného období z hlediska instalace opatření. Počty průjezdů obcemi pro jednotlivé sezóny v daných kategoriích uvádíme v tabulce 4.5. Všimněme si, že v Cuntisu pozorujeme v zimním období po instalaci opatření nedostatek

dat (v obou směrech). Období po instalaci tak v obci Cuntis k analýze sezónnosti nepoužijeme.

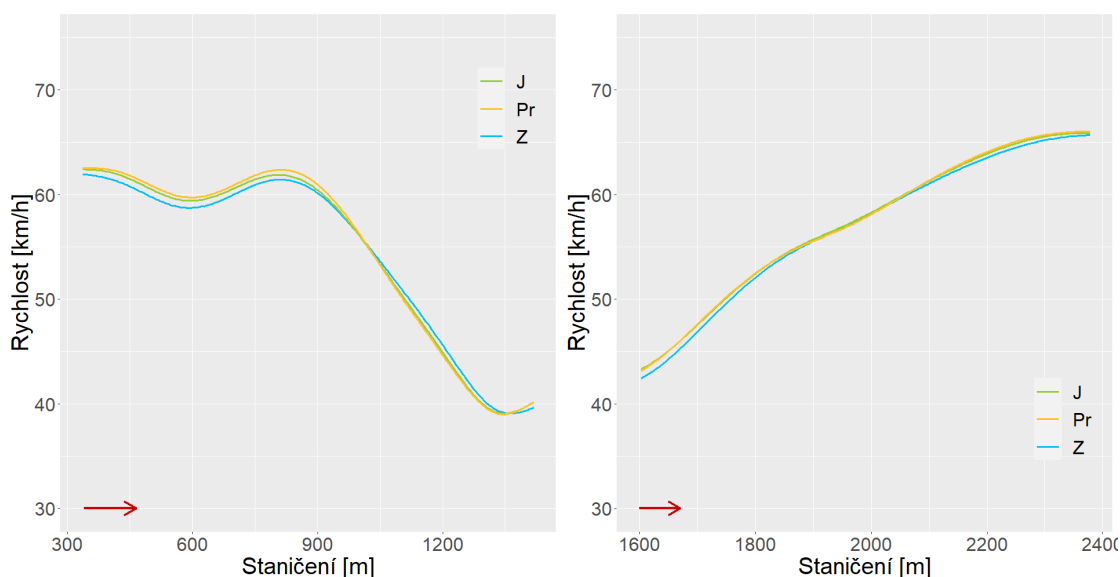
Shodu výše zmíněných dvojic sezónních provozů otestujeme opět globálním testem. Nejprve se však podívejme na vizuální srovnání. Grafy na obrázku 4.22 znázorňují průměrné sezónní průjezdy obcí Agolada ve směru 1 v příslušném období. V období před zavedením opatření skutečně pozorujeme snížení rychlosti vozidel v zimním období v porovnání s ostatními ročními obdobími. Naopak, během prázdnin bylo ve srovnání s ostatními sezónami zaznamenáno zvýšení rychlosti vozidel. Dle této vizualizace sezónnost hraje roli především při výjezdu z obce. Uvnitř Agolady k výrazným rozdílům nedochází. V zimní sezóně lze pozorovat jen menší snížení rychlosti i na části úseku uvnitř obce. Toto chování se dá vysvětlit. Při lepším počasí a lepším stavu vozovky se zvýšení rychlosti řidičů dá očekávat spíše na úsecích mezi obcemi, poněvadž uvnitř obce jsou stále omezeni rychlostním limitem 50 km/h, který se hlídá radary téměř v každé obci.



Obrázek 4.22: Srovnání sezónních průměrných průjezdů Agoladou ve směru 1. Vlevo: před instalací. Vpravo: po instalaci opatření.

Co se týče druhého grafu na obrázku 4.22, jenž se týká porovnání sezónních průjezdů v období po instalaci značení v Agoladě, může se zdát lehce překvapu-

jící. V zimní sezóně zde totiž vozidla oproti podzimu svou rychlost zvýšila. To je proti tomu, co bychom očekávali. To však může mít jednoduché vysvětlení. Jelikož se jedná o období po instalaci opatření, která probíhala v září, tyto podzimní průjezdy zaznamenávají období ihned po zavedení opatření, kdy mohl být efekt opatření nejintenzivnější, což způsobilo, že v nadcházející zimě pozorujeme rychlejší průměrný průjezd.



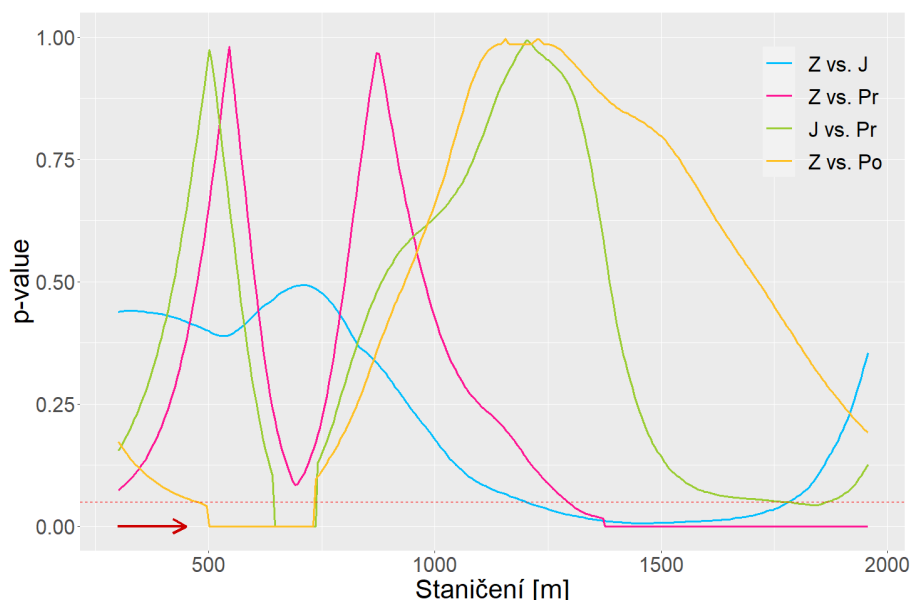
Obrázek 4.23: Srovnání sezónních průměrných průjezdů dolním (vlevo) a horním Cuntisem (vpravo) po instalaci opatření ve směru 1.

Porovnání sezón v dolním a horním Cuntisu v období před instalací opatření ve směru 1 znázorňují grafy na obrázku 4.23. Jarní a prázdninová sezóna téměř splývají, a zimní sezóna se mírně odchyľuje jen na některých místech úseků.

Testované sezóny	Agolada		Cuntis dolní		Cuntis horní	
	Směr 1	Směr 2	Směr 1	Směr 2	Směr 1	Směr 2
Z vs. J	< 0,001	< 0,001	0,045	0,212	0,117	0,612
Z vs. Pr	< 0,001	< 0,001	0,004	< 0,001	0,238	0,001
J vs. Pr	< 0,001	0,001	0,514	< 0,001	0,966	< 0,001
Z vs. Po	< 0,001	0,025	–	–	–	–

Tabulka 4.6: Výsledky globálního testu pro dané dvojice sezónních provozů dle jednotlivých úseků v daných směrech.

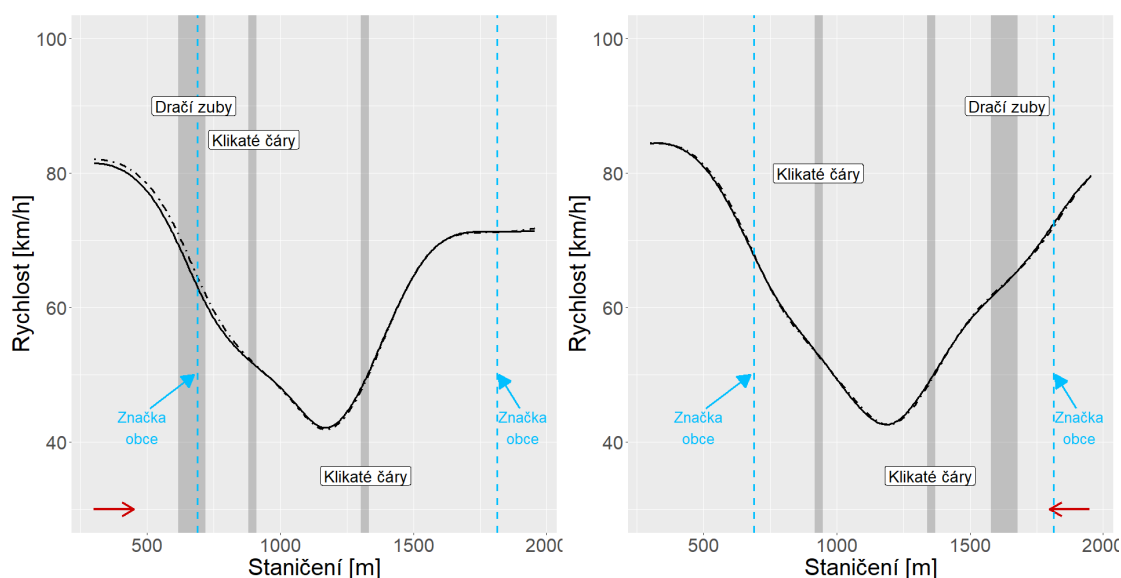
Nyní se podívejme na výsledky globálního testu pro testované sezóny, viz tabulka 4.6. V obci Agolada se projevila signifikantní sezónnost. Ve všech testovaných dvojicích sezónních provozů jsou statisticky významné rozdíly. V dolním Cuntisu už nejsou výsledky tak jednoznačné. Odlišnost zimního a prázdninového provozu se prokázala v obou směrech. Lze obecně očekávat, že tyto sezónní provozy se budou od sebe lišit nejvýznamněji. Co se týče zimní a jarní sezóny, v dolním Cuntisu globální test nezachytil žádný rozdíl ve směru 2 a hraničně signifikantní rozdíl ve směru 1. Mezi jarním a prázdninovým provozem byl významný rozdíl v průměrných průjezdech zaznamenán pouze ve směru 2. V úseku horního Cuntisu test zachytil efekt sezónnosti pouze ve dvou případech, a to u provozu zimního a prázdninového ve směru 2, a v tomtéž směru u provozu jarního a prázdninového. Navzdory tomu, že v některých případech jsme významnou závislost průměrné rychlosti vozidel při průjezdu obcemi na ročním období neidentifikovali, je zcela očividné, že problém sezónnosti FCD do analýzy zasahuje. V každém námi analyzovaném úseku se, ve větší či menší míře, projevila. Musíme proto tuto skutečnost při analýze vlivu opatření na rychlost vozidel brát v potaz.



Obrázek 4.24: Výsledky intervalového testu pro dané dvojice sezónních provozů v Agoladě ve směru 1.

Pro zajímavost uvádíme na obrázku 4.24 graf výsledku intervalového testu pro analyzované dvojice sezón v Agoladě ve směru 1. Vidíme, že rovněž intervalový test ve všech případech ukázal statisticky významnou odlišnost v sezónních provozech. Nyní se ovšem můžeme detailněji zaměřit na lokality úseku, kde průjezdy řidičů byly ovlivněny změnou ročního období.

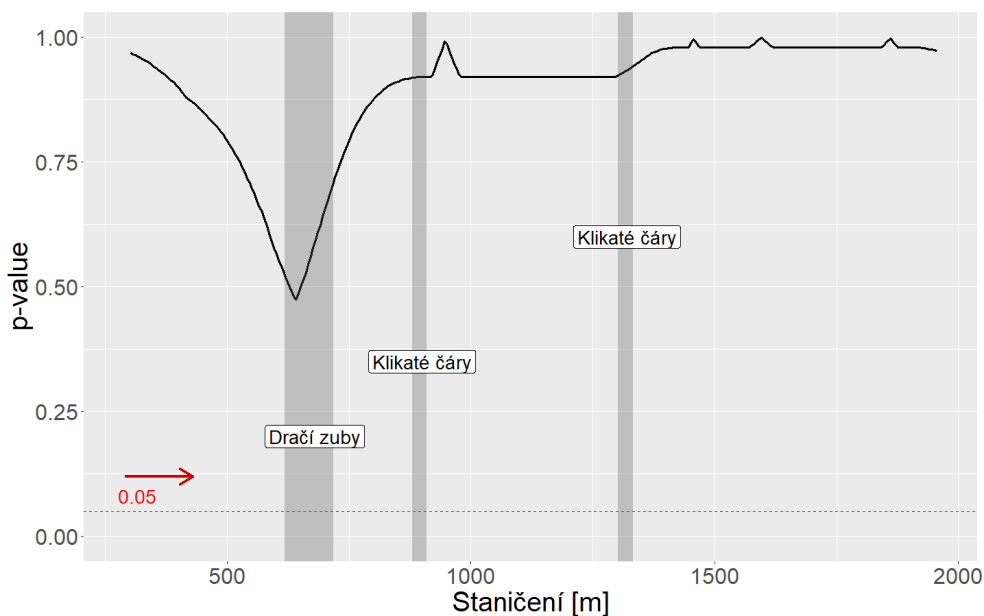
Uvedenou analýzu sezónnosti FCD v případě zájmu naleznete v příloze pod názvem `sezonnost.R`.



Obrázek 4.25: Průměrné průjezdy Agoladou v jednotlivých směrech v měsících leden až březen. Čárkovaně: před instalací dopravních opatření. Plně: po instalaci.

V tuto chvíli již máme testem prokázaný výskyt sezónnosti FCD. Jak bylo řečeno, v období před instalací a po instalaci dopravních opatření nemáme prokryté shodné sezóny, čímž může být vyhodnocení efektu opatření zkresleno. Zkusme proto analýzu znovu provést na shodném období, kterým je zimní sezóna (bez prosince). V Agoladě máme v těchto měsících k dispozici dostatek dat, v Cuntisu však nikoliv, jak jsme již uváděli při popisu tabulky 4.5. Z toho důvodu můžeme analýzu funkčnosti opatření pro shodné období provést pouze v Agoladě, kde máme za leden až březen ve směru 1 v období před instalací opatření 841 jízd a v období po instalaci 719 jízd, ve směru 2 potom 831 a 683 jízd. Nej-

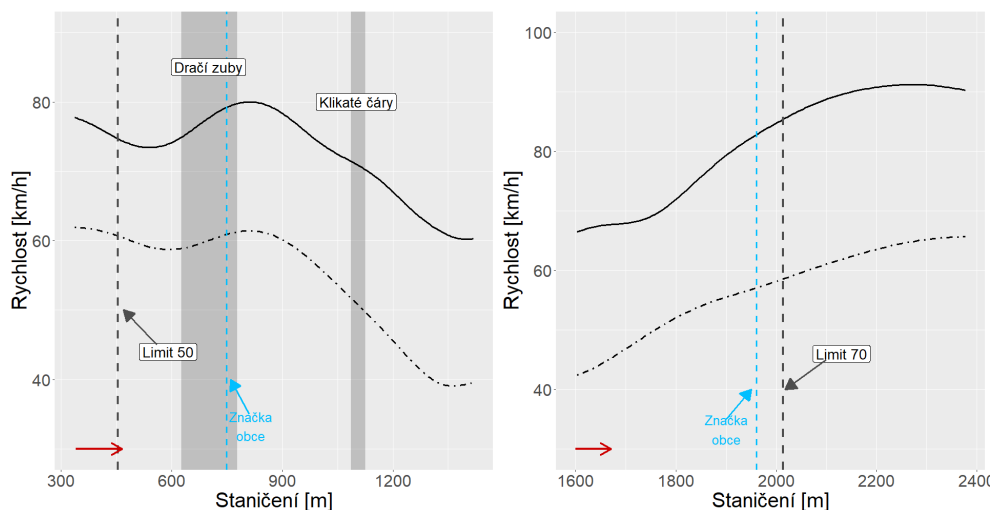
prve si ukažme vizualizaci průměrných rychlostních profilů pro toto roční období. Z grafu na obrázku 4.25 pozorujeme ve směru 1 mírné zpomalení vozidel v první třetině úseku, tj. od příjezdu do obce, kde byly instalovány dračí zuby, až do příjezdu k prvním klikatým čárám. V místě značky obce řidiči zpomalili v průměru o 2,3 %. Můžeme ovšem pochybovat o tom, zda pozorované zpomalení přisuzovat instalovaným dračím zubům. Řidiči začali zpomalovat již zhruba 300 metrů před tímto značením, kde nemohlo být ani spatřeno. Ve směru 2 nelze pozorovat žádnou změnu. Ani globální test ve směru 2 skutečně žádný efekt opatření neprokázal. Ve směru 1 test vyhodnotil p-hodnotu na 0,021, což na 5% hladině významnosti prokazuje signifikantní změnu průměrných rychlostních profilů po zavedení opatření. Intervalový test pro tento směr už však nedetekoval žádnou lokalitu významné změny, viz graf na obrázku 4.26. Je tak evidentní, že globálním testem detekovaný efekt je tak malý, že pro intervalový test není dostatečně průkazný.



Obrázek 4.26: Výsledek intervalového testu pro Agoladu ve směru 1 v měsících leden až březen.

Závěrem této analýzy lze konstatovat, že při minimalizaci vlivu sezónnosti dat již nelze nadále pozorovat snížení rychlosti vozidel při průjezdu obcí Agolada, a tím ani žádný pozitivní efekt instalovaných dračích zubů a klikatých čar. Z provedené analýzy se lze domnívat, že tato opatření měla vliv na rychlost řidičů čerstvě po zavedení opatření, tj. od října do prosince 2022, a později došlo k adaptaci řidičů k těmto novým opatřením.

Již víme, že v období leden až březen po instalaci opatření máme v Cuntisu měřených pouze 7 jízd ve směru 1, ve směru 2 dokonce žádné. Pro zajímavost se podívejme alespoň na průměrné rychlostní profily. Grafy na obrázku 4.27 znázorňují průměrné průjezdy dolním a horním Cuntisem ve směru 1 v období před a po instalaci opatření za měsíce leden až březen. Vidíme opravdu značný nárůst rychlosti na celých úsecích. Ovšem vzhledem k nedostatku dat musíme tato pozorování brát s velkou rezervou.



Obrázek 4.27: Průměrné průjezdy dolním (vlevo) a horním Cuntisem (vpravo) ve směru 1 v měsících leden až březen. Čárkovaně: před instalací dopravních opatření. Plně: po instalaci.

Z celkového hlediska jsou FCD Cuntisu problematická kvůli nedostatečně dlouhému období sběru dat po instalaci dopravních opatření. Tato skutečnost

způsobuje nejistotu v ohodnocení skutečného efektu opatření, a to ať už z hlediska nekonzistence dat v kontextu sezónnosti či k nedostatečnému posouzení vlivu opatření na rychlost vozidel po delší době od zavedení. Nejsme schopni říci, zda efekt pozorovaný v měsících ihned po zavedení opatření (tj. říjen, listopad a prosinec) přetrvává nadále. V praxi může nastat situace, kdy účinnost nově zavedených opatření postupně slábne po určitém období, což je způsobeno adaptací řidičů k těmto novým zavedením.

Tudíž nejen sezónnost, ale také další řada faktorů vstupuje do analýzy, aniž bychom chtěli, a proto je nutné s nimi pracovat. Vliv některých z nich můžeme potlačit díky komunikaci tohoto problému s daným poskytovatelem dat už před samotným sběrem FCD, a to například doporučením období sběru dat. Jiné faktory ovlivňující výsledky analýzy můžeme zase potlačit během pre-processingu dat. Jistě se ale rovněž najdou takové faktory, jejichž vliv nemůže redukovat ani poskytovatel dat, ani analytik (např. události na vozovce dlouhodoběji ovlivňující dopravu). To jen potvrzuje, že správné ohodnocení efektu instalovaných dopravních opatření není problematikou zdaleka triviální.

Závěr

V této práci jsme představili funkcionální data, jejich teoretické pozadí a jak s nimi pracovat v praktických úlohách. Stěžejní částí teoretického výkladu přitom byly vybrané statistické testy shody dvou funkcionálních výběrů, které jsme se rozhodli použít pro následnou analýzu vlivu dopravních opatření na rychlost vozidel v obcích Agolada a Cuntis ve Španělsku. Představili jsme algoritmus globálního testu, jenž slouží zejména k prvotní identifikaci změn v průměrných rychlostních profilech před a po instalaci dopravních opatření na daném úseku. Globální test se v praktické části vzhledem k velkému počtu funkcionálních pozorování ukázal jako velmi citlivý i na malou změnu. Dále byl detailně popsán komplikovaný algoritmus intervalového testu dle Pini a Vantiniho. Intervalový test slouží k identifikaci klíčových oblastí, v nichž k signifikantnímu rozdílu ve funkcích průměru výběrů dochází. Také tento test tak byl použit v analýze efektu dopravních opatření v obcích Agolada a Cuntis, a prostřednictvím výsledků jsme zodpověděli na otázku, která opatření daného úseku nevedla k žádnému významnému zpomalení vozidel, ve kterých místech úseku byla změna jednoznačně signifikantní a kde rozporuplná.

Než jsme se však v rámci praktické části dostali k samotnému hodnocení vlivu opatření na rychlost vozidel, podrobně jsme popsali analyzovaná data z plovoucích vozidel, analyzované úseky a instalovaná dopravní opatření. Rozebrány byly rovněž všechny nutné kroky pre-processingu dat tohoto typu a vyhlazování průjezdů obcemi. Veškeré kroky pre-processingu stejně jako volbu parametrů vyhlazování

jsme dostatečně odůvodnili.

Jak v obci Agolada, tak v dolním a horním Cuntisu jsme pomocí zmíněných statistických testů zjistili signifikantní změny v průměrných rychlostních profi-
lech před a po instalaci dopravních opatření, a v jakých oblastech k nim došlo. Při diskusi nad výsledky vyvstala myšlenka dalších vnějších faktorů, které do analýzy zřejmě vstupují, a výsledky tak nezachycují pouze efekt instalovaných opatření. Takovým významným faktorem se ukázala sezónnost. Na základě rozdělení průjezdů dle jednotlivých sezónních provozů jsme stejnými statistickými testy vyhodnotili sezónnost v těchto datech jako významnou. Navrhli jsme proto triviální řešení, jakým způsobem minimalizovat vliv sezónnosti, a sice analyzovat efekt opatření na datech, která jsou z hlediska sezónnosti konzistentní. Tuto analýzu jsme provedli pro obec Agolada, jen pro tuto obec jsme měli k dispozici dostatečný počet jízd v příslušném období, a zjistili jsme, že žádný efekt instalovaných dopravních opatření nadále již nelze pozorovat.

Jsme přesvědčeni, že náš výzkum poskytuje solidní základ pro další zkoumání v oblasti analýzy vlivu dopravních opatření na rychlost vozidel. Budoucí směřování výzkumu by se mohlo zaměřit na rozšíření analýzy na další oblasti pro upevnění a zpřesnění metodologie, a na hlubší rozbor vnějších faktorů, které v této úloze hrají významnou roli.

Literatura

- [1] Altintasi, O.; Tuydes-Yaman, H. a Tuncay, K.: Detection of urban traffic patterns from Floating Car Data (FCD) [online]. *Transportation Research Procedia*. 2017, VOL. 22, 382-391. ISSN 23521465. Dostupné z: <https://doi.org/10.1016/j.trpro.2017.03.057>. [cit. 2024-01-01].
- [2] Corain, L. a Salmaso, L.: Nonparametric permutation and combination-based multivariate control charts with applications in microelectronics [online]. *Applied Stochastic Models in Business and Industry*. 2013, VOL. 29, NO. 4, 334-349. ISSN 1524-1904. Dostupné z: <https://doi.org/10.1002/asmb.1976>. [cit. 2023-12-26].
- [3] De Boor, C.: *A Practical Guide to Splines*. Springer, New York, 1978.
- [4] Dierckx, P.: *Curve and surface fitting with splines*. New York: Clarendon, 1993. ISBN 0198534418.
- [5] Ferraty, F. a Vieu, P.: *Nonparametric functional data analysis: theory and practice*. Springer series in statistics. New York: Springer, 2006. ISBN 0-387-30369-3.
- [6] Giancristofaro, R. A. a Brombin, Ch.: Overview of nonparametric combination-based permutation tests for multivariate multi-sample problems [online]. *Statistica*. 2014, VOL. 74, NO. 3, 233-246.

- [7] Górecki, T.; Horváth, L. a Kokoszka, P.: Tests of Normality of Functional Data [online]. *International Statistical Review*. 2020, VOL. 88, NO. 3, 677-697. ISSN 0306-7734. Dostupné z: <https://doi.org/10.1111/insr.12362>. [cit. 2023-10-20].
- [8] Hron, K.; Kunderová, P. a Vencálek, O.: *Základy počtu pravděpodobnosti a metod matematické statistiky*. 4. doplněné vydání. Olomouc: Univerzita Palackého v Olomouci, Přírodovědecká fakulta, 2021. ISBN 978-80-244-5990-5.
- [9] James, G.; Witten, D.; Hastie, T. a Tibshirani, R.: *An introduction to statistical learning: with applications in R*. Second edition. Springer texts in statistics. New York: Springer, 2021. ISBN 978-1-0716-1417-4.
- [10] Kokoszka, P. a Reimherr, M.: *Introduction to functional data analysis*. Boca Raton: CRC Press, Taylor & Francis Group, 2017. ISBN 978-1-498-74634-2.
- [11] Kolářček, J.: *Text prezentací k předmětu M7777 Aplikovaná analýza funkcionálních dat*. Brno: Masarykova univerzita, 2019.
- [12] Maharani, M. a Saputro, D. R. S.: Generalized Cross Validation (GCV) in Smoothing Spline Nonparametric Regression Models [online]. *Journal of Physics: Conference Series*. 2021, VOL. 1808, NO. 1. ISSN 1742-6588. Dostupné z: <https://doi.org/10.1088/1742-6596/1808/1/012053>. [cit. 2023-10-07].
- [13] MetaCentrum (MetaVO) - virtuální organizace pro celou akademickou obec. (n.d.). MetaCentrum (MetaVO) - virtuální organizace pro celou akademickou obec. <https://metavo.metacentrum.cz/>. [cit. 2024-01-14].
- [14] Pini, A. a Vantini, S.: The interval testing procedure: A general framework for inference in functional data analysis [online]. *Biomet-*

- rics*. 2016, VOL. 72, NO. 3, 835-845. ISSN 0006-341X. Dostupné z: <https://doi.org/10.1111/biom.12476>. [cit. 2023-11-21].
- [15] Pini, A. a Vantini, S.: The interval testing procedure: Inference for functional data controlling the family wise error rate on intervals [online]. *MOX – Report*. Politecnico di Milano, 2013, NO. 13. [cit. 2023-11-28].
- [16] Pini, A.; Vantini, S. a Pini, M. A.: *fdatest: Interval Testing Procedure for Functional Data*. R package version 2.1.1, 2015. Dostupné z: <https://CRAN.R-project.org/package=fdatest>. [cit. 2024-01-15].
- [17] Ramsay, J. O.; Hooker, G. a Graves, S.: *fda: Functional Data Analysis*. R package version 6.0.5, 2022. Dostupné z: <https://CRAN.R-project.org/package=fda>. [cit. 2023-11-23].
- [18] Ramsay, J. O.; Hooker, G. a Graves, S.: *Functional data analysis with R and MATLAB*. Use R!. New York: Springer, 2009. ISBN 978-0-387-98184-0.
- [19] Ramsay, J. O. a Silverman, B. W.: *Applied functional data analysis: methods and case studies*. New York: Springer, 2002. ISBN 0-387-95414-7.
- [20] Ramsay, J. O. a Silverman, B. W.: *Functional data analysis*. 2nd ed. Springer series in statistics. New York: Springer, 2005. ISBN 978-0387-40080-8.
- [21] Ředitelství služby dopravní policie: *Informace o nehodovosti na pozemních komunikacích v České republice v období leden – červen 2023*. Policejní prezidium České republiky, 2023. [cit. 2024-03-23].

Seznam kódů v příloze

- `odboujici_jizdy.R` – identifikace a odstranění odbočujících jízd,
- `vyhlazovani.R` – vyhlazení průjezdů s kubickou B -splajnovou bází a fixní vyhlazovací konstantou,
- `vyhlazovani_GCV.R` – vyhlazení průjezdů s kubickou B -splajnovou bází a volbou vyhlazovací konstanty pomocí GCV,
- `testy.R` – použití globálního a intervalového testu pro analýzu vlivu dopravních opatření na rychlost vozidel,
- `sezonnost.R` – analýza sezónnosti FCD.