



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

ÚSTAV SOUDNÍHO INŽENÝRSTVÍ

INSTITUTE OF FORENSIC ENGINEERING

IMPLEMENTACE STATISTICKÉ METODY KDE+

IMPLEMENTATION OF KDE+ STATISTIC METHOD

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Mgr. Tomáš Svoboda

VEDOUCÍ PRÁCE

SUPERVISOR

RNDr. Michal Bíl, Ph.D.

BRNO 2016

Zadání diplomové práce

Ústav:	Ústav soudního inženýrství
Student:	Mgr. Tomáš Svoboda
Studijní program:	Rizikové inženýrství
Studijní obor:	Řízení rizik v informačních systémech
Vedoucí práce:	RNDr. Michal Bíl, Ph.D.
Akademický rok:	2015/16

Ředitel ústavu Vám v souladu se zákonem č.111/1998 o vysokých školách a se Studijním a zkušebním řádem VUT v Brně určuje následující téma diplomové práce:

Implementace statistické metody KDE+

Stručná charakteristika problematiky úkolu:

Metoda KDE+ (Kernel Density Estimation plus) je nová statistická metoda, pomocí které lze detekovat místa koncentrací bodových objektů nebo jevů v jednorozměrné dimenzi. Typickou aplikací metody KDE+ je analýza shlukování dopravních nehod v rámci jednotlivých úseků pozemních komunikací. V rámci diplomové práce bude vytvořena uživatelsky snadno použitelná a přenosná implementace této metody. Bude proveden návrh vhodných datových struktur a rozhraní. Samotnému technickému popisu aplikace bude předcházet analýza jiných obdobných statistických metod.

Cíle diplomové práce:

Teoretické představení metody KDE+, analýza obdobných metod.

Vytvoření počítačové desktopové aplikace s grafickým uživatelským rozhraním, která bude uživatelům umožňovat užití metody KDE+ na vlastních datech.

Algoritmus výpočtu bude optimalizován pro paralelní běh s možností dávkového spuštění v textovém režimu.

Definice vhodných formátů vstupních a výstupních souborů.

Seznam literatury:

BÍL, Michal; ANDRÁŠIK, Richard; JANOŠKA, Zbyněk; SEDONÍK, Jiří; VALENTOVÁ, Veronika. Identifikace kritických nehodových lokalit pomocí GIS analýzy polohy dopravních nehod: metodika. Brno: Centrum dopravního výzkumu, 2014, 44 s. ISBN 978-80-86502-76-2.

BÍL, Michal; ANDRÁŠIK, Richard; JANOŠKA, Zbyněk. Identification of hazardous road locations of traffic accidents by means of kernel density estimation and cluster significance evaluation. 2013. Accident Analysis and Prevention 55, 265 – 273.

SCHILDT, Herbert. Mistrovství - Java. 1. vyd. Brno: Computer Press, 2014, 1224 s. ISBN 978-80-2-1-4145-8.

BÍL, Michal; ANDRÁŠIK, Richard; SVOBODA, Tomáš; SEDONÍK, Jiří. The KDE+ software: a tool for effective identification and ranking of animal-vehicle collision hotspots along networks. 2015. Landscape Ecology. DOI 10.1007/s10980-015-0265-6.

Termín odevzdání diplomové práce je stanoven časovým plánem akademického roku 2015/16

V Brně, dne

L. S.

doc. Ing. Aleš Vémola, Ph.D.
ředitel

Abstrakt

V rámci diplomové práce jsem představil novou statistickou metodu KDE+ (Kernel Density Estimation plus), pomocí níž lze detekovat shluky bodů na liniových datech. Vytvořil jsem samostatně běžící aplikaci, která umožní případným zájemcům si zmíněnou metodu vyzkoušet a aplikovat ji na svá vlastní data. Metodu i aplikaci lze použít k detekci míst, v nichž se koncentrují dopravní nehody. Tvorba aplikace zahrnovala analýzu statistické metody KDE+, návrh vhodných programových struktur a samotnou implementaci. Po vytvoření funkčního prototypu byly prováděny optimalizace za účelem dosažení vyššího výkonu. Následně proběhlo testování aplikace na datech o nehodách v silniční síti České republiky.

Abstract

In this master's thesis I presented a new statistical method KDE+ (Kernel Density Estimation plus) that allows detecting clusters of points on the linear data. I created a self-standing application that enables anybody to try the method and apply it on their own data. One possible usage of the method and application is for the detection of critical roads sections with a high concentration of traffic accidents. Development of the application includes analysis of KDE+ statistical method, design of appropriate program structures and the implementation. Optimization were carried out to achieve higher performance after creating the prototype. At the end the software was validated by analysing vehicle collision data from the police database of the Czech Republic.

Klíčová slova

KDE+, shlukování, pravděpodobnost, jádrový odhad hustoty, Monte Carlo, doprava, kritická místa na komunikacích, dopravní nehody

Keyword

KDE+, clustering, probability, kernel density estimation, Monte Carlo, transport, hazardous road location, transport accidents

Bibliografická citace

SVOBODA, T. Implementace statistické metody KDE+. Brno: Vysoké učení technické v Brně, Ústav soudního inženýrství, 2016. 75 s. Vedoucí diplomové práce
RNDr. Michal Bíl, Ph.D.

Prohlášení

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

V Brně dne 26. 5. 2016

.....
Tomáš Svoboda

Poděkování

Na tomto místě chci poděkovat svému vedoucímu RNDr. Michalu Bílovi, Ph.D. z Centra dopravního výzkumu, v. v. i. za četné náměty a připomínky k vytvářené aplikaci, a také za poskytnutí možnosti práce na daném tématu. Dále děkuji Mgr. Richardu Andrášikovi, spoluautorovi popisované statistické metody, za časté a podnětné konzultace z různých oblastí matematiky.

Za poskytnuté výpočetní a úložné zdroje děkuji centru CERIT Scientific Cloud, působícímu v rámci programu „Projekty velkých infrastruktur pro výzkum, vývoj a inovace“ (LM2015085).

Obsah

Obsah	9
1 Úvod.....	11
2 Rizika v dopravě.....	14
2.1 Definice rizika.....	14
2.2 Druhy dopravních rizik	14
2.3 Statistiky nehodovosti	17
2.4 Dopravní nehody.....	19
2.5 Sběr dat o nehodách	20
2.6 Identifikace kritických lokalit	21
2.7 Možnosti řešení problémových míst	23
2.8 Ekonomické důsledky nehod	24
3 Statistické metody.....	26
3.1 Základní pojmy	26
3.2 Metoda KDE+	27
3.3 Další metody	31
4 Implementace	36
4.1 Použité technologie a nástroje.....	36
4.2 Návrh aplikace	38
4.3 Definice souborových formátů.....	40
4.4 Implementace statistického algoritmu.....	42
4.5 Paralelizace	43
4.6 Internacionalizace	44
4.7 Grafické rozhraní	44
4.8 Automatizované testování.....	48
4.9 Měření rychlosti výpočtu	49
5 Použití a vyhodnocení.....	51

5.1	Příprava vstupních dat.....	51
5.2	Postup spuštění a práce s aplikací	53
5.3	Prezentace vypočtených výsledků.....	55
5.4	Aplikace metody KDE+ na nehody v silniční síti ČR	56
5.5	Příklady použití	56
6	Závěr.....	60
6.1	Možnosti dalšího rozvoje aplikace.....	61
	Použitá literatura	63
	Použité zkratky	67
	Seznam obrázků.....	68
	Seznam tabulek	69
	Seznam grafů.....	69
	Přílohy.....	70
	Příloha A – Seznam elektronických příloh	70
	Příloha B – Příručka pro práci s aplikací	71
	Příloha C – Snímky obrazovek vytvořené aplikace	73

1 Úvod

Doprava patří ke každodenním potřebám civilizace. Již několik století před počátkem našeho letopočtu hrály významnou společenskou roli starověké obchodní stezky. Od 18. do počátku 20. století docházelo k výrazným změnám existující dopravní infrastruktury. Nově vznikající komunikační síť se vyznačovala mnohem větší mírou systematickosti a spolehlivosti, neboť začala být méně závislá na výkyvech počasí a ročním období. Současně s tím se také začaly dramaticky snižovat náklady na přepravu osob a zboží na velké vzdálenosti. Pokles nákladů vedl k významné změně mobility obyvatelstva a tento posun se následně stal katalyzátorem zásadní společenské a ekonomické transformace. V souvislosti s prudkou modernizací dopravy dochází ke změně ve vnímání prostoru a času. [1]

Moderní doprava se vyznačuje dvěma základními charakteristikami. Za prvé je založena na často velmi finančně náročné infrastruktuře. Za druhé pak již není jako hnací síla vozidel použita síla člověka či zvířete, ale je užito mechanického zařízení (parního stroje, spalovacího, vznětového či v současné době na trh stále výrazněji pronikajícího elektrického motoru).

Modernizace dopravy v 18. století je spojena s rozšiřováním sítě kamenných silnic. Ty nejkvalitnější z nich se vyznačovaly kamenným povrchem po celé ploše, přímým směrem a postranními příkopy. V 19. století se pak objevují i první technické stavební předpisy kladoucí na vznikající silnice základní požadavky. Budování hlavních silničních spojů bylo motivováno posílením integrity státního celku a přirozeně také podporou akceschopnosti vojska. Význam silniční infrastruktury se začíná měnit v polovině 19. století, kdy dochází k rozmachu železniční dopravy. Tento trend je převážně patrný v dálkové přepravě. Silniční síť přijímá svůj nový, spíše regionální význam. Zdokonalující se silniční síť pak byla předpokladem pro nástup automobilové dopravy počátkem 20. století.

Mohlo by se zdát, že s rozvojem informačních technologií, které umožňují rychlý přenos informací bez nutnosti přepravy fyzického nosiče, bude objem přepravovaného zboží a osob klesat, avšak dostupné statistické informace, např. o počtu registrovaných automobilů v České republice [2], hovoří o jednoznačném růstu pozemní silniční dopravy. A právě pozemní silniční dopravou se budu v této práci dále zabývat. Představené metody by ale bylo možno s úpravami aplikovat i na jiné druhy pozemní dopravy, jakými je např. doprava železniční nebo vodní. Popřípadě by bylo možno popisovanou statistickou metodu nebo vytvořenou počítačovou aplikaci použít i na řešení problémů s dopravou nijak nesouvisejících, např. pro vyhledávání významných událostí v časových řadách.

Jako společnost čelíme mnohým pozitivním i negativním dopadům dopravy. Tyto dopady je možno z velké části ovlivnit, o což by se příslušné subjekty měly při plánování, konstrukci a provozu dopravních sítí snažit.

Mezi pozitivní přínosy dopravy patří zjevně možnost individuálního či hromadného cestování osob, např. do zaměstnání, za nákupy, za odpočinkem apod. Nákladní doprava umožňuje společnosti život v luxusu, který lidem poskytuje dostupnost široké nabídky zboží kdekoli na světě bez závislosti na místě jeho původu.

Do negativních dopadů dopravy můžeme řadit čerpání nerostných zdrojů (ropy, plynu,...) a znečištění životního prostředí plynoucího z jejich těžby, z dopravy na místo určení a samotného spalování v motorech dopravních prostředků. Ke znečišťování životního prostředí dochází ve více rovinách. Ovzduší je znečišťováno spalinami vznikajícími při provozu motorů, prostředí je ovlivněno hlukem či vibracemi z dopravy. Dopravní prostředky mohou také přispívat ke světelnému smogu. Jakkoli jsou ekologické dopady dopravy závažné pro udržitelný rozvoj, bude se naše práce zabývat jiným podstatným negativem souvisejícím s dopravou, a to dopravními nehodami. Samozřejmě i dopravní nehody mohou mít velmi negativní ekologické následky ve formě znečištění vodních toků či půdy, nebo v podobě usmrcení volně žijících zvířat.

Dopravní nehody integrálně patří k dopravě samotné, jsou jejím projevem, který nejspíš nikdy nebude vzhledem k množství faktorů, jež v rámci dopravní soustavy vystupují a ovlivňují její provoz, možno zcela eliminovat. Mělo by však být obecnou snahou tento negativní jev v co nejvyšší míře minimalizovat. Výskyt dopravních nehod s sebou nese relativně velký celospolečenský dopad. Jedná se především o finanční ztráty na majetku a zdraví osob. Kromě finančních ztrát nelze opomenout ani sociální důsledky plynoucí ze snížení životního standardu osob, které jako účastníci nehod utrpěly zdravotní újmu, či přišly během dopravní nehody o život. Dle statistik Evropské unie [3] přišlo v roce 2015 na jejím území o život při dopravních nehodách 26 tisíc osob. Dalších 135 tisíc lidí utrpělo vážné zranění. Ačkoli mají meziroční srovnání relativně příznivou tendenci, je zřejmé, že je nutno se touto problematikou cíleně zabývat.

Metodickým omezením výskytu dopravních nehod v silničních i jiných sítích se spolu s dalšími institucemi v České republice zabývá Centrum dopravního výzkumu, v. v. i.¹, které je vědeckovýzkumnou institucí působící pod Ministerstvem dopravy České republiky. Ve spolupráci s výzkumníky z této instituce byla vypracována i tato práce.

¹ Centrum dopravního výzkumu, v. v. i. (CDV) – www.cdv.cz

V následující kapitole představím rizikologická a dopravní východiska hledání nehodově kritických míst v silničních sítích. Ve třetí kapitole se budu zabývat popisem statistické metody KDE+, přehledem obdobných metod a představením statistických pojmů, které budou v této práci dále použity. Ve čtvrté kapitole popíši průběh tvorby a výslednou implementovanou podobu počítačové aplikace. V páté kapitole popíši přípravu zdrojových dat, představím modelový případ užití a nastíním některé možnosti využití aplikace. V závěru shrnu dosažené výsledky a zmíním několik dalších možných směrů rozvoje vytvořeného software.

2 Rizika v dopravě

V této kapitole se budu zabývat problematikou analýzy rizik souvisejících s dopravními sítěmi a dopravou samotnou. Mezi ostatními vyskytujícími se riziky v dopravě zde bude kladen důraz především na skutečnosti týkající se dopravních nehod.

Mnoho oblastí naší společnosti je na dopravních sítích, ať už se jedná o sítě silniční, železniční či letecké koridory, velkou měrou závislých. Pokud dojde byť jen k částečnému porušení funkcionality jakékoli z nich, lze často pozorovat značné snížení životního standardu obyvatel. Jedná se například o zhoršenou možnost získání rychlé zdravotní péče a dalších kritických služeb, nebo omezení běžných každodenních aktivit, jakými jsou např. dojíždění za prací nebo nakupování. [4]

2.1 Definice rizika

Dle Janíčka [5] je riziko „*pravděpodobnost, že na konkrétní, tzv. rizikové entitě se zrealizuje událost, která bude mít negativní důsledky pro sebe nebo své okolí, tedy, že se zrealizuje nebezpečí*“. Nebezpečí stejný autor definuje jako „*reálnou možnost (hrozbu), že na určité (rizikové) entitě vznikne negativní událost.*“ V bezpečnostní praxi se obvykle vyjadřuje riziko jako součin pravděpodobnosti vzniku negativní události a kvantifikace důsledků (míry dopadu) uplatnění události. Vypočtené hodnoty rizik je možno vzájemně porovnávat a určovat jejich pořadí. Tímto způsobem lze určit rizika s nejvyšší závažností.

Frič ve své publikaci [6] hovoří o rizikovém místě v dopravním kontextu jako o „*místu na pozemní komunikaci, kde je vyšší pravděpodobnost vzniku dopravních nehod (které se ale zatím nemusely stát)*“. Oproti tomu za nehodové místo se označuje taková lokalita, na které nastala vyšší koncentrace nehod určitého typu za stanovené časové období (definice dle normy ČSN 73 6100-2). Riziková a nehodová místa pak tvoří dohromady kritická místa, která je potřeba identifikovat a následně na nich realizovat opatření snižující pravděpodobnost vzniku nebezpečí či míru jeho dopadu. [7]

2.2 Druhy dopravních rizik

Rizika ve vztahu k dopravní infrastruktuře je možno rozdělit do několika hlavních kategorií a podkategorií.

1. Rizika vnitřní
 - 1.1. Infrastrukturální
 - 1.2. Provozní
2. Rizika vnější
 - 2.1. Přírodní (živly)
 - 2.2. Umělá

2.2.1 Rizika vnitřní

Vnitřní rizika mají svého původce uvnitř dopravního systému a lze je v základním členění rozdělit na rizika infrastrukturální a provozní.

Rizika infrastrukturální mají svou příčinu v infrastruktuře dopravní sítě. Do této kategorie lze také zahrnout rizika, která neleží přímo uvnitř dopravního systému, ale mohou se nacházet na pomezí tohoto systému, avšak s tím, že jejich původ je úzce spjatý s provozováním dopravy. Jako příklad lze uvést pád kamení na železniční těleso, které bylo zapříčiněno dlouhodobým působením vibrací přenášených z kolejí na okolní skálu. Dále se může jednat o stavební a technická selhání, která mohou být zapříčiněna špatnou konstrukcí či používáním dopravního objektu. Takovými selháními mohou být např. pád mostu, utržení krajnice vozovky či vytvoření výraznějších nerovností na jejím povrchu. [8]

Další kategorie tvoří rizika provozní, které mají svůj původ v provozu probíhajícím na silniční síti. Jsou ovlivněna množstvím faktorů, kterými mohou být např.:

1. Hustota provozu
2. Podnebí a počasí
3. Technický stav vozidla
4. Stav vozovky
5. Dopravní předpisy
6. Lidský faktor

Nejvýraznějším zástupcem této kategorie jsou dopravní nehody. Dalšími riziky spadajícími do této kategorie mohou být nejrůznější negativní ekologické projevy, kterými jsou například srážky se zvěří. Při výskytu dopravních nehod dochází často k újmě na zdraví, škodám na majetku a k částečnému nebo úplnému omezení silničního provozu, což může přeneseně působit další škody. Dopravní nehodu vzniklou především kvůli lidskému faktoru lze vidět na obrázku 2.1. Lidský faktor spočívá v nedodržení bezpečné rychlosti při řízení nákladního vozidla při jízdě v oblouku. Vlivem působení setrvačných sil na náklad vozidla s výše položeným těžištěm (starý papír) došlo při změně směru jízdy k překročení mezní hodnoty a převrácení kamionu. Uvedená nehoda se obešla bez vážnějšího zranění.



Obrázek 2.1 – Dopravní nehoda – překlopení tahače s návěsem
(Zdroj: fotoarchiv autora práce)

Jako příklad selhání způsobeného stavem komunikace, i když v tomto případě byla doprava na vozovce zastavena, lze uvést zřícení opravované mostní konstrukce ve Vilémově, při kterém zemřeli čtyři dělníci. Dle závěrů policejního prošetřování a znaleckých posudků z oboru stavebnictví byl během konstrukčních prací porušen technologický postup a vlivem špatné koordinace prací subdodavatelských firem došlo k výraznému oslabení statiky mostu vedoucí k jeho pádu. [9]

2.2.2 Rizika vnější

Další kategorií nepříznivých událostí v silniční síti tvoří poruchy způsobené vnějšími vlivy, jež mají svůj původ mimo silniční síť. Tyto jevy není možno zpravidla z pozice správce komunikace předem plánovat a ani předpovídat. Lze pouze vytipovat úseky, kde s relativně vysokou pravděpodobností mohou nastat.

V případě přírodních rizik se jedná např. o povodně, sesuvy půdy a skalní řícení, požáry, výskyt zvěře na vozovce, intenzivní sněžení, zemětřesení a další katastrofy přírodního charakteru. Pro výskyt některých přírodních vlivů lze sestavit predikční modely, např. lze predikovat regionální ohrožení sesuvy půdy. Významné ohrožení provozu působí větší množství nově napadlého sněhu, zvláště pokud se jedná o intenzivní sněžení v krátkém časovém intervalu. [10]

Umělá rizika mohou tvořit nebezpečí, která mají svůj původ nezávislý na přírodních vlivech. Jako příklad lze uvést pád předmětu na vozovku či teroristický útok.

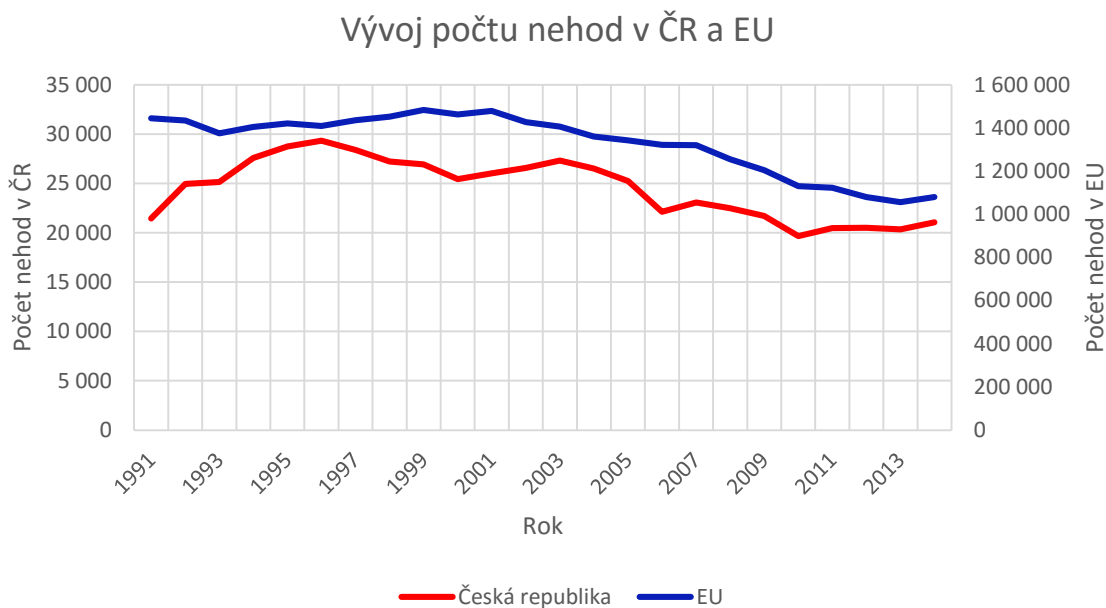
2.3 Statistiky nehodovosti

Dle rozhodnutí Rady Evropské unie 93/704/ES ze dne 30. listopadu 1993 (o vytvoření databáze Společenství týkající se dopravních nehod v provozu na pozemních komunikacích) jsou na území všech členských států sbírány statistické údaje o dopravních nehodách. Národní statistiky jsou z jednotlivých států předávány Evropskému statistickému úřadu, který z pověření Komise tyto údaje agreguje a zveřejňuje souborné výsledky. [11]

Pro účely zmíněné směrnice jsou státy povinny vést evidenci všech dopravních nehod, jejichž následkem je zranění nebo smrt alespoň jednoho účastníka silničního provozu, a které zapříčinilo pohybující se vozidlo jedoucí po veřejně přístupné komunikaci.

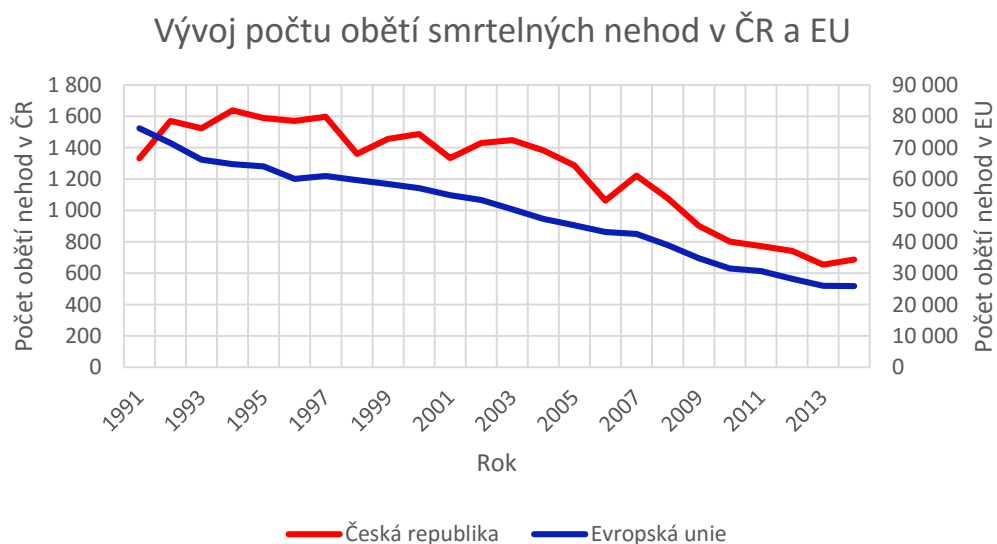
V Evropské unii se v roce 2014 událo 1 079 800 dopravních nehod. Při těchto nehodách přišlo o život 25 900 osob. Jako osoba usmrcená při nehodě se započítává každý člověk, který zemřel následkem dopravní nehody přímo na místě jejího výskytu, nebo nejdéle 30 dní poté. Z uvedených čísel je zřejmé, jak nutná je existence preventivních opatření zabráňujících vzniku dopravních nehod. Tato potřeba je zakotvena v usnesení (91/C 178/01) Rady a zástupců vlád členských států zasedajících v Radě. V usnesení je konstatováno, že tehdejší počty zraněných nejsou přijatelné z morálního, politického, ekonomického ani sociálního hlediska. Při předpokladu stálého růstu počtu vozidel pohybujících se na silnicích mají členské státy za povinnost vynaložit významné úsilí pro zvýšení silniční bezpečnosti, a to především pomocí preventivních opatření. [12]

Níže uvedené grafy byly vypracovány na základě tabulkových dat zveřejněných Evropskou komisí. V datech, zveřejněných v únoru 2016, se zatím nenacházejí statistiky za všechny členské státy za rok 2015, z toho důvodu jsou grafy vytvořeny po rok 2014. [13]



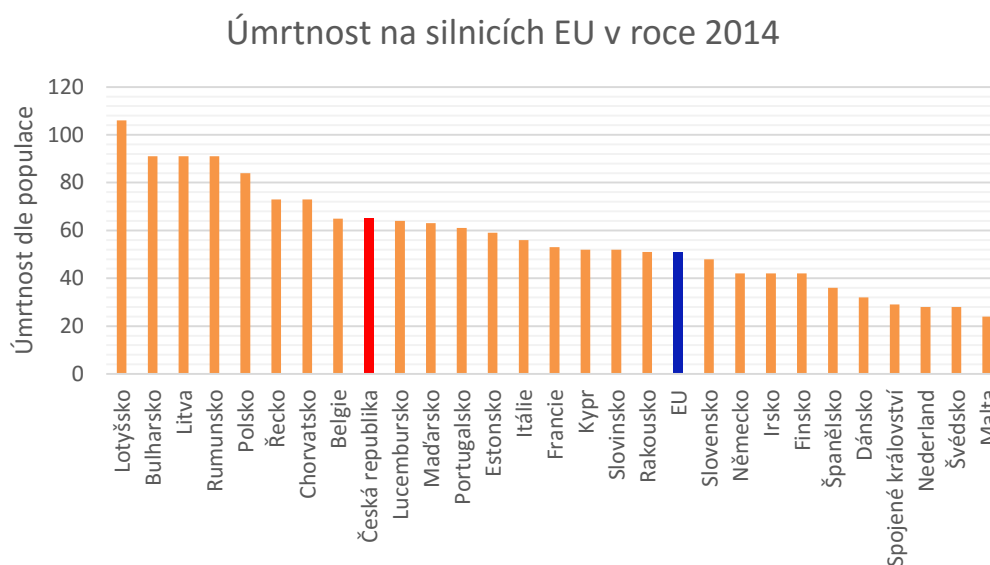
Graf 2.1 – Vývoj počtu nehod v ČR a EU

Na grafu 2.1 lze vidět vývoj počtu dopravních nehod, ke kterým došlo na území České republiky a Evropské unie od roku 1991. Z průběhu diagramu je vidět, že obě dvě křivky dodržují jistou společnou tendenci. Z dlouhodobého měřítka došlo ke snížení počtu nehod jak v České republice, tak v Evropské unii. Avšak od roku 2009 mají počty dopravních nehod stagnující charakter.



Graf 2.2 – Vývoj počtu obětí smrtelných nehod v ČR a EU

Z grafu 2.2 lze získat pozitivní informaci, že vývoj počtu nehod, při kterých došlo k usmrcení jedné nebo více osob, má jak v České republice, tak v Evropské unii, klesající charakter. Mezi roky 1995 a 2014 došlo k poklesu počtu smrtelných dopravních nehod přibližně na polovinu. Příbyl ve své monografii [14] uvádí morbidní příměr o počtu při nehodách usmrčených osob v České republice v posledním dvacetiletí. Toto číslo podle něj představuje vy-mření (doslovně hovoří o vyvraždění) jednoho menšího okresního města.



Graf 2.3 – Úmrtnost na silnicích EU v roce 2014

Na grafu 2.3 je uvedeno porovnání počtu zemřelých mezi jednotlivými státy. Hodnoty zemřelých jsou přepočítány vzhledem k celkovému počtu obyvatel států, lze tedy provést jejich vzájemné porovnání. Číslo na vertikální ose odpovídá počtu zemřelých osob na jeden milion obyvatel. Česká republika (zvýrazněna červeně) zastává v tomto relativním pořadí deváté místo. Nachází se tak na rozmezí první a druhé třetiny států s nejvyšší úmrtností. Modrým sloupcem je vyveden průměr celé Evropské unie.

2.4 Dopravní nehody

Jedna z několika možných definic dopravní nehody zní následovně:

„Dopravní nehoda je nepředvídaná, ale zpravidla předvídatelná událost, která vznikla během provozu na dopravní cestě a měla za následek škodu na životě, zdraví, majetku či jiný, zvláště závažný následek.“ [15]

Dle uvedené definice je jako dopravní nehoda označena událost charakterizovaná výskytem všech tří výše uvedených skutečností.

Prvním znakem je předvídatelnost. Dopravní nehoda je náhlá událost zahrnující v sobě jistý prvek překvapení. Míra překvapení je subjektivní vlastnost, která je závislá na psychických dispozicích účastníků nehody a ostatních vlivech (např. celkové dopravní situaci, oslnění, ...). Existují ale i výjimky, v rámci kterých není prvek překvapení přítomen, jako jsou např. sebevraždy či pojistné podvody. I přes její nenadálouost mohou účastníci nehody při vynaložení jistého množství úsilí před jejím vznikem důvodně předvídat vznik konkrétních událostí, které s nezanedbatelnou mírou pravděpodobnosti mohou vést k nehodě.

Další vlastností nehody je místo jejího vzniku. K nehodě musí dojít na dopravní cestě. Z hlediska automobilové dopravy může být touto cestou jakákoli silniční komunikace včetně účelové. Pokud tedy dojde k poškození vozidla či újmě na zdraví člověka v garáži, tak se o dopravní nehodu nejedná.

Za třetí je zde požadavek na následek, tedy že musí dojít k poškození majetku či zdraví. Škoda způsobená na majetku může být jak na samotném vozidle, tak na dopravní infrastruktuře, nebo i mimo ni.

Pokud nedošlo při dopravní události k naplnění všech pojmových znaků, ale pouze k výskytu nepředvídané negativní události na dopravní cestě bez vzniku škody, hovoříme o dopravním incidentu. Výskyt mnoha dopravních incidentů na krátkém časovém či prostorovém úseku může indikovat vysokou pravděpodobnost vzniku dopravní nehody. Dopravní incidenty však nejsou nijak hromadně evidovány.

2.5 Sběr dat o nehodách

V posledních letech se markantně zlepšila kvalita dat, která jsou dopravní inženýři o dopravních nehodách schopni získat. Dříve obsahovala policejní hlášení pouze přibližnou lokalizaci nehody určenou staničením. Jako staničení se označuje využití systému lineárního referencování, pomocí kterého lze určit absolutní vzdálenost každého bodu na linii vzhledem k danému referenčnímu bodu, kterým je obvykle počátek kilometráže. Hodnota staničení 700 m tedy odpovídá bodu vzdálenému 700 m od počátku konkrétní silnice.

Staničení bylo často odebráno velmi přibližně a se značnou chybou, jež mohla činit v závislosti na periodicitě umístění kilometrovníků i stovky metrů. Taková přesnost byla zvláště na kratších úsecích nedostatečná. Díky rozšíření satelitních lokalizačních systémů

(především systému GPS²) se od roku 2006 stanovuje poloha každé nehody s přesností nejvýše na několik jednotek metrů. Navýšení přesnosti umožňuje využití dat o nehodách při statistickém zpracování. [16]

Vznik nepřesností se objevuje dále při měření odlišných geografických bodů souvisejících s nehodou. Můžeme určit bod vzniku nehody, tedy takový bod, ve kterém začal nehodový děj, či můžeme určit také koncový bod nehody, tedy místo, kde se vozidlo po nehodě zastavilo. S vyšší rychlostí rozdíl vzdálenosti význačných nehodových bodů dále narůstá. Ještě komplikovanější situace je při hromadných nehodách, u kterých se může nehodových bodů vyskytovat ještě více. [17]

Dále má na kvalitu získaných dat o nehodách vliv podíl nahlášených a zaevidovaných dopravních nehod vůči celkovému počtu nehod. Na tento podíl má zásadní vliv finanční hranice vzniklé škody na některém zúčastněném vozidle, od které je povinnost k nehodě přivolat Policii ČR. Tato hranice naposled vzrostla v roce 2009 a aktuálně činí 100 000 Kč. Policii je třeba k vyšetření dopravní nehody přivolat také v případě, kdy dojde k poškození zdraví osob, vznikne škoda na majetku třetí osoby (poškození pozemní komunikace, třetího vozidla, plotu k silnici přilehlé zahrady), nebo nejsou účastníci vlastními silami schopni obnovit plynulost silničního provozu. [7]

Kromě lokalizačních informací se o nehodách evidují také následující skutečnosti: [6]

- Čas výskytu (lze stanovit závislost výskytu nehody v denním, měsíčním či ročním měřítku)
- Povětrnostní podmínky (např. sucho, déšť, namrzlá vozovka)
- Směr jízdy vozidla
- Druh srážky (např. čelní, boční, zadní náraz)
- Příčina dle policie (např. nepřiměřená rychlost, alkohol za volantem)
- Následky dopravní nehody (hmotná škoda, újma na zdraví či životě)

2.6 Identifikace kritických lokalit

Metody popisující identifikaci kritických míst na dopravních komunikacích lze rozdělit na takové, které se zabývají silničními úseky v intravilánu obcí, nebo mimo obce – v extravilánu. Silniční úseky v obcích se vyznačují častým výskytem kratších úseků silnic a úrovňového křížení – křižovatek. Jiné metody se zabývají odhalováním rizikových lokalit

² GPS – Global Positioning System (Globální polohový systém)

v extravilánu, kde se nachází typicky delší úseky, s vyšší maximální povolenou rychlostí a s méně častým úrovnovým křížením.

Dopravní nehoda, ke které došlo na silnici v extravilánu, je typicky z pohledu zdravotních následků závažnější než nehoda, ke které došlo v intravilánu. Riziko usmrcení je podle Friče [6] dokonce na extravilánové silnici šestkrát vyšší. Dvakrát vyšší je riziko usmrcení na směrově nerozdělené komunikaci oproti směrově rozdělené komunikaci.

2.6.1 Metody zaměřené na numerické definici

U těchto metod je rozhodným kritériem počet dopravních nehod v rámci stanoveného časového období. Numerické metody jsou založeny na principu plovoucího okna, kterým je posouváno po komunikaci, přičemž je vyhledávána lokalita, na které počet nehod překročí stanovenou mez.

V roce 2001 aktualizovaná metodika [18] používaná v České republice stanoví jako častá místa dopravních nehod křižovatky nebo úseky o délce 250 m, jestliže se na nich staly:

- Alespoň tři dopravní nehody osobního vozidla během jednoho roku
- Alespoň tři nehody stejného typu s osobními následky za tři roky
- Alespoň pět nehod stejného typu za jeden rok

V roce 2006 byl v České republice zaveden systém INFOBESI, který používá k agregaci dat numerickou metodiku. Informace o kritických místech zpracované tímto systémem jsou elektronicky dostupné.³

2.6.2 Metody hodnocení silnic

Metody náležející do této kategorie vycházejí z expertního posouzení silničního úseku odborníkem, který pro své rozhodování použije dostupné údaje o komunikaci, jakými jsou např. popisné charakteristiky komunikace získané při její pasportizaci, data o dopravní intenzitě a nehodovosti. Konkrétní programy této metodiky představují např. EuroRAP (European Road Assessment Programme), nebo usRAP (U.S. Road Assessment Programme). Programy se soustředí především na předcházení nejzávažnějším nehodám, jakými jsou čelní srážky nebo střet s pevnou překážkou mimo vozovku. [7]

³ INFOBESI - <http://infobesi.dopravniinfo.cz>

2.6.3 Predikční modely nehodovosti

Predikční modely nehodovosti používají statistických postupů, za pomoci kterých je předpověděn počet očekávaných nehod v závislosti na zvolených vstupních parametrech. Vstupními parametry často bývají proměnné vyjadřující:

- Dopravní zatížení komunikace
- Typ komunikace (např. dálnice, silnice I. třídy)
- Šířkové uspořádání komunikace (např. počet pruhů a jejich šířka)
- Způsob řízení dopravy (např. nejvyšší povolená rychlost)

Obecný předpis modelu pro mezikřižovatkové úseky lze vyjádřit vztahem: [19]

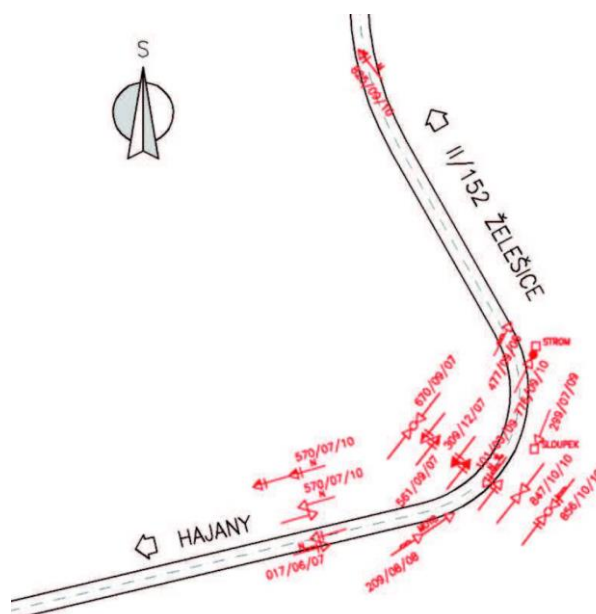
$$E(\lambda) = \alpha \times Q^\beta \times e^{\sum \gamma_i x_i} \quad (2.1)$$

Kde $E(\lambda)$ představuje očekávaný počet silničních nehod při intenzitě dopravy Q a vysvětlujících proměnných x_i . Model by měl zahrnovat takové proměnné, jejichž vztah je k nehodovosti příčinný. Proměnné α, β, γ jsou regresní koeficienty. Cílem modelování je odhad hodnot těchto koeficientů.

Oproti předchozím postupům nejsou vstupem predikčních metod historické počty nehod. Historické údaje o nehodách totiž podléhají zkreslujícím vlivům, zejména regresi k průměru, což je jev, při kterém se při opakování jistého měření zprvu naměřené extrémní hodnoty blíží průměrným hodnotám. [19]

2.7 Možnosti řešení problémových míst

Po procesu identifikace problémových míst lze přistoupit k bezpečnostní inspekci (audit) konkrétních nehodových lokalit a podrobnějšímu zkoumání nehod, které se na daném úseku udály. K popisu souboru nehod se častou používá kolizní diagram, na kterém jsou souhrnně graficky znázorněny veškeré nehody na zvoleném místě. Pro dosažení kvalitních výsledků je doporučeno využít data za delší časové období. Minimální délka tohoto období by měla činit alespoň tři roky. [6]



Obrázek 2.2 – Kolizní diagram
(Zdroj: Metodika [6])

Dopravní nehodovost lze návrhem a úpravou stávajících silnic do značné míry ovlivnit jak v kladném, tak i v záporném smyslu. Výsledkem bezpečnostní inspekce bývají na úsecích mezi křižovatkami zpravidla následující opatření: [14]

- Odstranění ohrožujících prvků (překážky ve výhledu, nevyhovující stromy, reklamní plochy, zbytečné pevné překážky)
- Dosažení unifikace (dodržení stejného označení obdobných lokalit)
- Modifikace šířky krajnice a jízdních pruhů
- Vyvarování se prudkých změn směru a výšky
- Zajištění dostatečné délky připojovacích a odpojovacích pruhů
- Zapuštění konců ocelových svodidel

2.8 Ekonomické důsledky nehod

Zvyšování bezpečnosti silničního provozu a snižování počtu vážných dopravních nehod je cílem Národní strategie bezpečnosti silničního provozu pro období 2011 – 2020. Během těchto let by mělo být zachráněno více než 3 000 životů. [7]

Snižováním počtu nehod dochází k poklesu výdajů v rezortu dopravy, ale k významným úsporám dochází také ve zdravotnictví a v oblasti sociálního zabezpečení. Vyskočilová [20] vyjádřila v roce 2013 finanční náklady dopravních nehod následovně:

- 18 669 000 Kč za každou usmrcenou osobu
- 5 062 000 Kč za každou těžce raněnou osobu
- 413 000 Kč za každou lehce zraněnou osobu
- 226 000 Kč za každou nehodu pouze s hmotnou škodou

Pokud by byl dodržen stanovený předpoklad záchrany 3 000 životů, došlo by v uvedeném období k úspoře rovnající se přibližně 57 miliard Kč.

Následným problémem vzniklým při dopravních nehodách mohou být částečná či úplná neprůjezdnost (uzavírky) silničních úseků. V případě uzavření úseku v obou směrech je třeba provoz přeměrovat na náhradní objízdné trasy. Takové změny dopravy s sebou nesou další dodatečné finanční náklady pro všechny dotčené účastníky silničního provozu. Objízdné trasy nejsou obvykle na nárazové zvýšení provozu zcela dimenzované, dochází zde ke zvýšení pravděpodobnosti vzniku nehody vlivem nadměrného dopravního proudu. Situaci po nehodě blokující provoz neprospívá ani existence plánovaných či jiných neplánovaných uzavírek, které ještě více protahují objízdnou trasu. [8]

3 Statistické metody

V této kapitole popíší statistické metody, které jsou užívány při určování rizikových lokalit na dopravních komunikacích. Aplikace statistických metod vyžaduje existenci strojově zpracovatelné databáze dopravních nehod či nehodových lokalit s dostatečnou přesností lokalizace.

V textu této a následujících kapitol budu užívat k dopravě vztažené termíny silniční úsek a dopravní nehoda. V obecnějším kontextu budou tato pojmenování občas nahrazována za linie a body (jevy) na liniích.

3.1 Základní pojmy

Shluková analýza (shlukování) je proces rozdělování objektů do předem nedefinovaných tříd – **shluků** – na základě podobnosti objektů podle zvolené funkce podobnosti. Do téhož shluku patří objekty, které jsou si co nejvíce podobné, a zároveň se odlišují od objektů, které do téhož shluku nepatří. [21]

V terminologii dopravních nehod si pod pojmem shluk lze přestavit takovou skupinu dopravních nehod, které jsou si navzájem prostorově blízké a naopak jsou relativně vzdálené od jiných samostatných nehod nebo shluků. Na pojem blízkosti nehod či shluků lze nahlížet různě. Může se jednat o planární vzdálenost dopravních nehod (euklidovská vzdálenost vzdušnou čarou), či o vzdálenost měřenou po linii (viz kapitola 2.5). [7]

Metody hledající shluky mohou být rozděleny na globální či lokální. Globální metody nalézají staticky významné shluky, které vybočují svým náhodným rozdělením napříč celou množinou analyzovaných bodů. Pomocí lokálních metod lze nalézt shlukující se objekty v rámci definovaného okolí.

Dále lze shlukovací metody rozlišit podle komplexnosti vypočtených výsledků (viz kapitola 3.3). Některé metody umožňují pouze určení existence shluku. Je tedy otestováno, zda mají data v rámci daného úseku tendenci ke shlukování. Nejsou však již označeny konkrétní shluky. Další „pokročilejší“ metody dovedou stanovit i přesnou lokalizaci shluků. [16]

Hustota rozdělení pravděpodobnosti (hustota pravděpodobnosti) je spojité rozdělení pravděpodobnosti pro danou náhodnou veličinu, které definuje pravděpodobnost výskytu konkrétní hodnoty dané náhodné veličiny. Plocha pod křivkou znázorňující hustotu pravděpodobnosti je vždy rovna jedné. Tedy platí

$$\int_{x_2}^{x_1} P(x) dx = 1 \quad (3.1)$$

kde hodnoty x_1 a x_2 označují definiční obor veličiny X , tedy rozsah zkoumaného intervalu. [22]

3.2 Metoda KDE+

Metoda KDE+⁴ je statistická metoda založená na jádrovém odhadu funkce hustoty (KDE⁵). Pomocí metod založených na tomto principu lze nalézt shluky bodů dopravních nehod včetně určení jejich přesné lokalizace. Metoda KDE+ vylepšuje standardní metodu KDE o možnost rozlišení nevýznamných a významných shluků bez nutnosti stanovení mezní hodnoty jako vstupního parametru výpočtu. Dalším pozitivním aspektem metody KDE+ je možnost absolutního porovnání významnosti shluku v rámci celé silniční sítě. Následující popis metody KDE+ byl vytvořen s použitím zdrojů [16], [7], [23], ve kterých je metoda podrobně popsána.

Algoritmus metody KDE+ se pro každou zpracovávanou linii skládá z následující sekvence kroků. Od základní metody KDE se odlišuje v krocích 2 a 3, které zajišťují automatické rozlišení významných shluků.

1. Výpočet jádrové funkce hustoty
2. Monte Carlo simulace
3. Určení shluků a stanovení jejich významnosti

První a druhý krok lze vykonávat nezávisle na sobě, jejich vstupem je informace o délce úseku a souřadnice nehodových bodů. Třetí krok bere jako vstupní data výsledky obou předchozích kroků.

3.2.1 Výpočet hustoty pravděpodobnosti

Mějme silniční úsek délky L , na kterém se vyskytují dopravní nehody ve vzdálenostech $x_1, x_2, x_3, \dots, x_n$ měřených od počátku úseku, kde n je počet dopravních nehod. Zřejmě musí platit vztah

$$\forall x_i : 0 \leq x_i \leq L, \quad i \in \langle 1, n \rangle \quad (3.2)$$

⁴ KDE+ – Kernel Density Estimation plus (Jádrový odhad hustoty plus)

⁵ KDE – Kernel Density Estimation (Jádrový odhad hustoty)

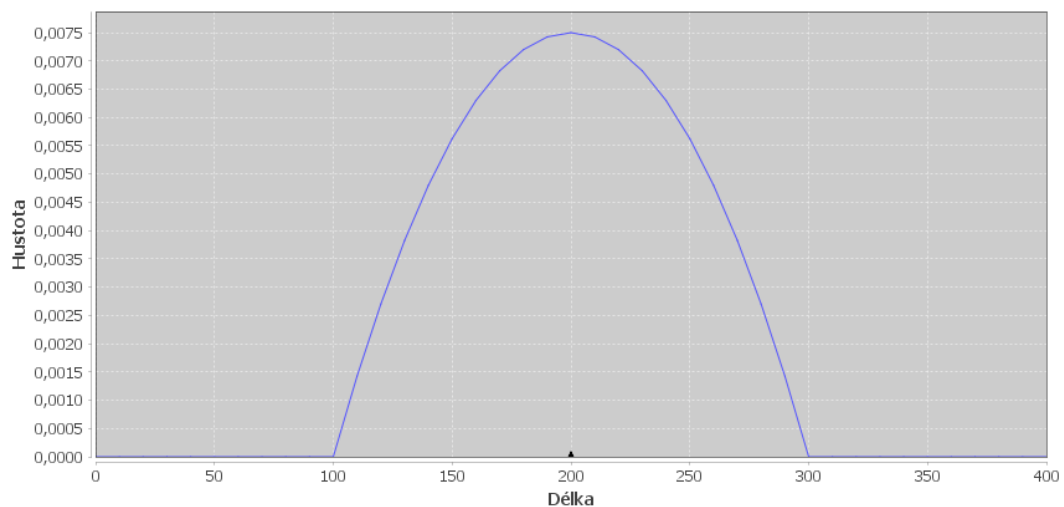
Nyní zvolíme typ jádrové funkce $K(x)$ a velikost okna $d > 0$. Jako jádrová funkce se často volí Gaussova křivka nebo její aproximace. Podstatné je, že její průběh by měl vyjadřovat míru nejistoty v zaznamenání nehody. Vzhledem k tomu, že Gaussova křivka prochází nad celou reálnou osou (je definována na celém intervalu $(-\infty, \infty)$ a nikdy jej neprotne), byl autory metody zvolen jiný typ jádrové funkce, a to Epanechnikova kvadratická funkce (viz Obrázek 3.1), která je nenulová do vzdálenosti d od svého středu. Jinak nabývá nulové hodnoty. Výpočet této funkce nad celým zkoumaným úsekem je dán předpisem

$$K_d(x) = \frac{3}{4d} \left(1 - \left(\frac{x}{d}\right)^2\right) I_{(-d; d)}(x) \quad (3.3)$$

kde funkce $I_{(-d; d)}(x)$ indikuje přítomnost na intervalu $(-d; d)$ tak, že

$$I_{(-d; d)}(x) = \begin{cases} 1, & x \in (-d; d) \\ 0, & x \notin (-d; d) \end{cases} \quad (3.4)$$

Jinými slovy, do každého bodu náhodné veličiny x_i , ve kterém se stala nehoda, se umístí jádrová funkce $K_d(x)$.

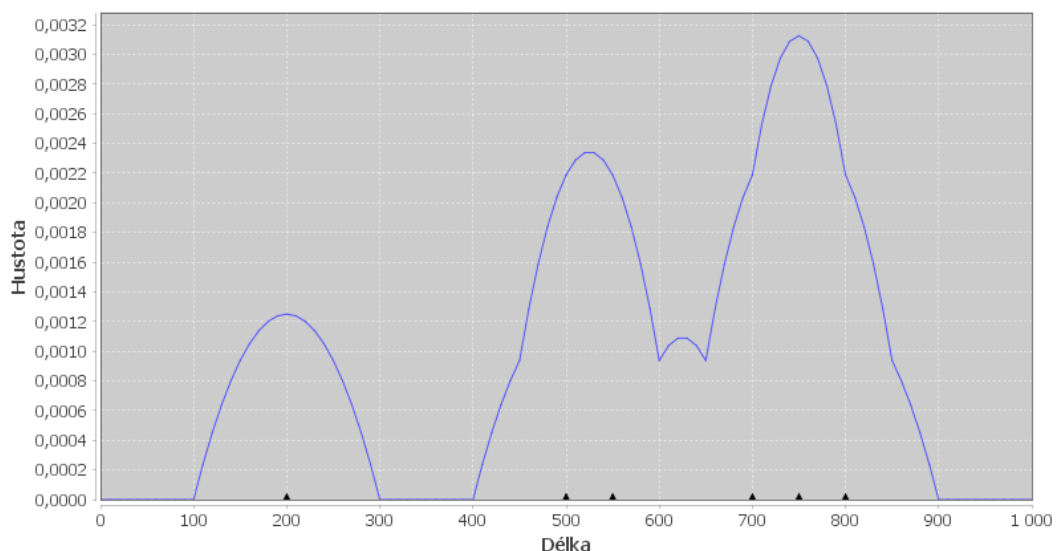


Obrázek 3.1 – Průběh Epanechnikova jádra

Následně jsou jádrové funkce ve všech nehodových bodech sečteny (znázorněno na obrázku 3.2) a je provedeno vydělení celkovým počtem bodů. Jelikož je jádro nenulové pouze ve svém definovaném okolí, součet jader se nárůstem funkce hustoty projeví pouze v průniku několika okolí více bodů. Podoba souhrnné funkce hustoty je tedy dána předpisem

$$f_n(x) = \frac{1}{n} \sum_{i=1}^n K_d(x - X_i) \quad (3.5)$$

Na obrázku 3.2 lze pozorovat výslednou podobu hustoty pravděpodobnosti nad zkoumanou linií o délce 1000 m. Na této linii se nachází šest bodů, nad kterými bylo vyneseno a sečteno Epanechnikovo jádro. Vypočtená funkce hustoty dosahuje čtyř lokálních maxim. Hledané shluky budou ležet v oboustranném okolí některých vrcholů této křivky. Problém nastává s určením minimální hodnoty hustoty (vynesené na svislé ose), od které bude shluk považován za významný. Ke stanovení této mezní hodnoty – hladiny významnosti – je třeba provést další krok metody KDE+, který metodu odlišuje od ostatních statistických metod založených na jádrovém odhadu hustoty, u nichž je třeba tuto hladinu stanovit předem, jako parametr výpočtu, či následně při analýze výsledných dat.



Obrázek 3.2 – Sčítání jádrové funkce

3.2.2 Monte Carlo simulace

Metoda Monte Carlo byla vyvinuta ve 40. letech 20. století v Los Alamos, kde sloužila k provádění výpočtů, které nebylo možno standardní, deterministickou cestou provést. Zpočátku byla držena v tajnosti, postupně se ale vzhledem ke svému vhodnému použití rozšířila a stala se široce využívanou v nejrůznějších matematických, fyzikálních a jiných aplikacích.

Metody Monte Carlo je při výpočtu shluků použito k určení hladiny významnosti h . Při překročení hladiny významnosti hustotou rozdělení pravděpodobnosti lze intervaly na linii pod touto nerovností nazvat signifikantními shluky.

Pro každou zkoumanou linii je provedena simulace Monte Carlo. Simulace se skládá ze dvou kroků. K jejich provedení je třeba znát délku linie L a počet bodů n , které se na ní nachází.

1. Na intervalu $\langle 0; L \rangle$ je vygenerováno n bodů z náhodného rovnoměrného rozložení.
2. Pro takto vygenerované body vypočteme jádrový odhad hustoty pravděpodobnosti.

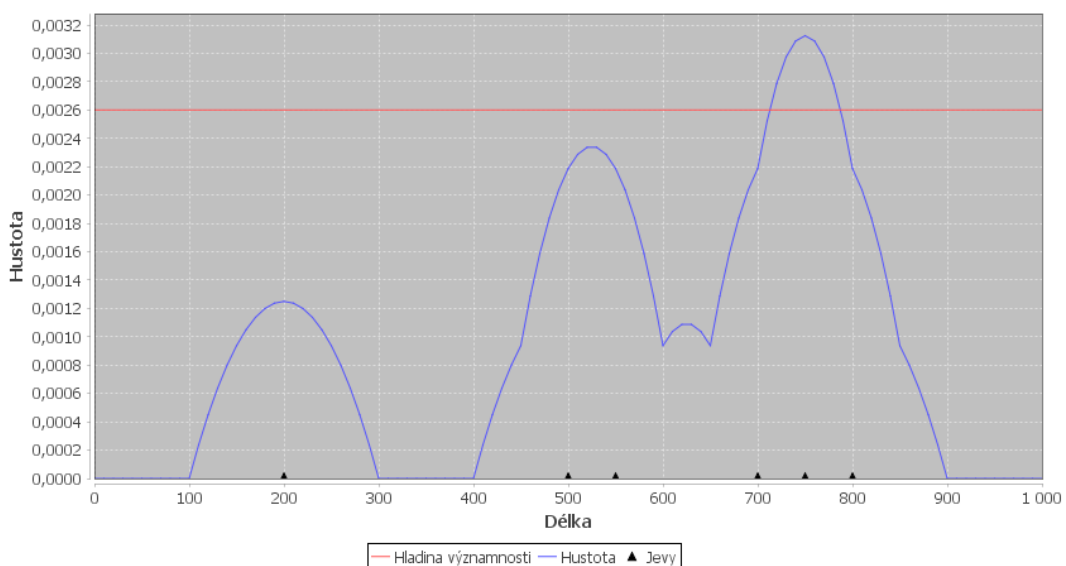
Jelikož se jedná o stochastický postup, je nutné provést dostatečně velký počet opakování předchozích dvou bodů. Pro vyvarování se výsledkům ovlivněných náhodnou chybou je třeba provést alespoň 100, ale raději i více opakování.

Provedené jádrové odhady jsou označeny jako $g_1(x), g_2(x), g_3(x), \dots, g_M(x)$, kde M vyjadřuje počet běhů simulace Monte Carlo.

Nejlépe si lze průběh tohoto kroku představit jako vytvoření velkého počtu stejně dlouhých úseků, délka všech je rovna délce zpracovávaného úseku. Na každém úseku je náhodně rozmístěn daný počet virtuálních nehod, který odpovídá skutečnému počtu nehod na zpracovávaném úseku.

3.2.3 Určení shluků a stanovení hladiny významnosti

V předchozí podkapitole byla připravena data pro testování hypotézy o rovnoměrném rozdělení bodů na linii. Stanovme nulovou hypotézu H_0 předpokládající, že „Body na linii mají náhodné rovnoměrné rozložení“.



Obrázek 3.3 – Stanovená hladina významnosti

Náhodné rovnoměrné rozložení bodů je splněno u jádrových odhadů $g_1(x), g_2(x), g_3(x), \dots, g_M(x)$. Za pomoci těchto odhadů je pro každý bod linie x_i vypočten $1 - \alpha$ kvantil z čísel $g_1(x_i), g_2(x_i), g_3(x_i), \dots, g_M(x_i)$, který je označen jako $v(x_i)$. Hladina významnosti h je stanovena jako střední hodnota ze všech $v(x_i)$. Jako hodnota α se může použít např. obvyklých 5 %. Při M -násobném opakování simulace Monte Carlo bude tato hodnota znamenat, že $95 \% \times M$ náhodných hodnot bude ležet níže, než získaná hladina h .

Jako statisticky významné shluky pak označíme takové intervaly úseků, nad nimiž překročí rozdělení pravděpodobnosti vypočtené nad původním úsekem hladinu významnosti h . Ilustrace této situace je na obrázku 3.3.

Z poměru výšky maxima hustoty pravděpodobnosti nad hladinou významnosti h a její absolutní výšky v bodě maxima lze odvodit sílu shluku. Síla shluku nabývá hodnoty z intervalu od nuly do jedné.

3.3 Další metody

V předchozí podkapitole byla představena metoda založená na jádrovém odhadu hustoty pravděpodobnosti KDE+. Byla zmíněna také metoda KDE, která se od její vylepšené varianty liší nutností expertního stanovení mezní úrovně významnosti (rizikivosti) nalezených shluků. Účel použití dalších metod, které se používají k hledání nehodových míst na silnicích, jsou uvedeny v tabulce 3.1.

Metoda	Detekce shluků	Lokalizace shluků	Určení významnosti shluku	Uspořádání shluku
KDE	•	•		
Index nebezpečí	•	•	•	
K-funkce	•			
Hierarchické shlukování	•	•		
Clumping metoda	•	•	•	
KDE+	•	•	•	•

Tabulka 3.1 – Porovnání metod nalézajících shluky

(Zdroj: The KDE+ software..., 2016 [23])

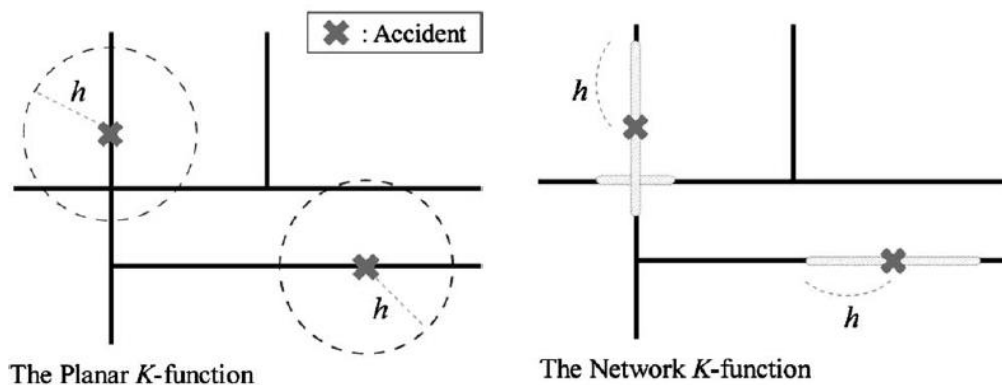
3.3.1 K-funkce

K-funkce je statistická metoda, která umožňuje detekci tendence ke shlukování, neposkytuje však již informace o lokalizaci shluku. Tento postup poskytuje přehled vzájemné závislosti bodů rozptýlených v prostoru. K-funkce bývá vypočtena jako standardizovaný průměr počtu bodů do vzdálenosti h od libovolného bodu. Pro výpočet se použije vzorec

$$K(h) = \frac{1}{\lambda} E(N_h) \quad (3.6)$$

Levá strana $K(h)$ představuje hodnotu K-funkce ve vzdálenosti h , kde $h \geq 0$. Funkce $E(N_h)$ odpovídá střední hodnotě počtu bodů do vzdálenosti h od libovolného jiného bodu. Konstantu λ tvoří intenzita bodů v jednotce plochy. Výše uvedený vzorec (3.6) má několik dalších specifických variant dle konkrétního užití. [24]

Názorně si lze postup představit jako vytvoření řady soustředných kruhů okolo míst nehod a následné sečtení počtu kruhy zasažených nehodových bodů. V základní variantě je vzdálenost měřena standardně po přímé linii v euklidovském prostoru. Pro potřeby analýzy silničních nehod se používá spíše upravená síťová varianta metody, kdy se vzdálenost měří pouze po definovaných hranách silniční sítě. Rozdíl mezi těmito dvěma způsoby měření je patrný z obrázku 3.4. Síťový způsob měření lépe odpovídá dopravní sémantice, jelikož nejsou jako blízké označeny nehody ležící na dvou vedle sebe rovnoběžně ležících úsecích. [25]



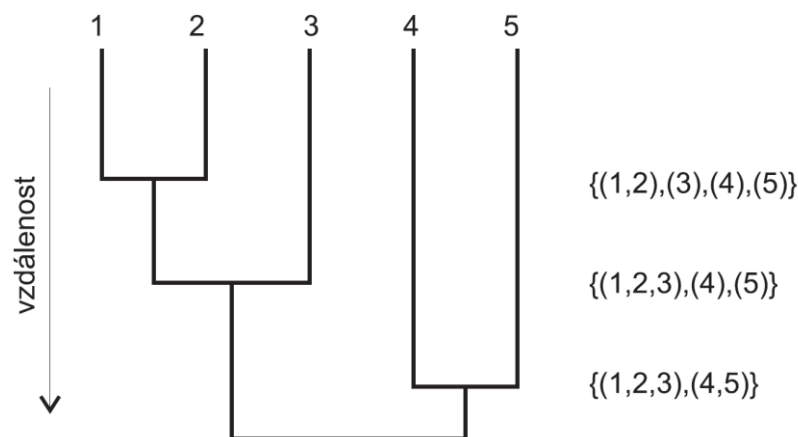
Obrázek 3.4 – Rozdíl mezi planárním a síťovým měřením vzdálenosti
(Zdroj: Comparison of planar and network K-functions..., 2004 [25])

3.3.2 Hierarchické shlukování

Při hledání shluků hierarchickým shlukováním je vytvářen systém podmnožin (vnořených rozkladů) z původní množiny obsahující všechny analyzované objekty. Existují dva ex-

trénní případy těchto rozkladů. Na jedné straně se mohou všechny zkoumané objekty nacházet v témže shluku, na straně druhé leží každý objekt ve svém vlastním shluku. Počet shluků je pak roven počtu objektů. Úkolem výpočtu je nalézt vhodné rozdělení objektů do shluků, které bude ležet mezi těmito dvěma okrajovými stavy.

Vytvářený systém podmnožin lze znázornit matematickou strukturou binárního kořenového stromu. Tento graf je v popisovaném kontextu nazýván jako dendrogram (ilustrace na obrázku 3.5). Každý uzel dendrogramu reprezentuje jeden shluk. Uzel může obsahovat libovolný kladný počet objektů. [26]



Obrázek 3.5 – Dendrogram pěti objektů

(Zdroj: Shluková analýza [26])

Existují dvě hlavní strategie vytváření shluků: [27]

1. Přístup aglomerativní

Stromová struktura je tvořena od listů (uzlů bez potomků). Na počátku obsahuje každý shluk pouze jeden objekt. Postupně jsou slučovány objekty podle funkce vzdálenosti, dokud není splněno ukončovací kritérium nebo se nenachází všechny objekty v jednom shluku.

2. Přístup divizní

Stromová struktura je tvořena od kořene (význačného uzlu bez předků). Na začátku výpočtu se všechny objekty nachází v jediném shluku, který je postupně dělen na menší shluky. Výpočet pokračuje do splnění ukončovacího kritéria nebo do okamžiku, kdy se každý objekt nachází ve svém samostatném shluku.

Definice funkce vzdálenosti (nepodobnosti) objektů, která na svém výstupu vrací koeficient odlišnosti objektů, může být založena např. na minimální (metoda nejbližšího souseda) nebo maximální (metoda nejvzdálenějšího souseda) vzdálenosti objektů. [27]

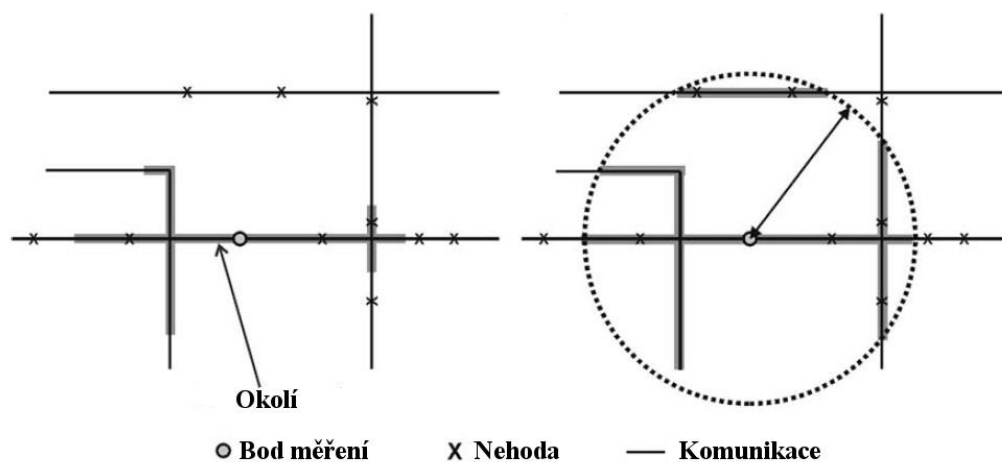
Nevýhodou metody je neschopnost provedení úprav po jednou provedeném rozdělení či spojení shluků. Dále také skutečnost, že pomocí ní není možno zjistit statistickou významnost shluku. [28]

3.3.3 Index nebezpečí

Metoda indexu nebezpečí (DI^6) využívá stejně jako metoda KDE jádrového odhadu hustoty pravděpodobnosti. Na rozdíl od ní ale definuje tzv. body měření, vůči kterým je vypočítána hodnota funkce hustoty. Body měření, které si lze představit jako jednorozměrnou či dvou-
rozměrnou mřížku, je rovnoměrně pokryta zkoumaná oblast. Stanovení bodů měření do velké míry ovlivňuje kvalitu výsledku. [29]

Při procházení všech bodů mřížky jsou ke každému bodu započítány objekty, jejichž vzdálenost k bodu mřížky je menší než definovaná konstanta. Určení objektů blízkých k bodu měření je ilustrováno na obrázku 3.6. Okolí bodu měření je znázorněno pomocí zvýraznění komunikace. V levé části obrázku je příklad síťového měření vzdálenosti, v pravé části je použito euklidovského určení délky.

Pokud není síť bodů měření dostatečně hustá, může dojít k nedetekování shluků nacházejících se mezi body měření. Pokud je struktura sítě jemná, blíží se výpočet touto metodou standardní metodě KDE. [29]



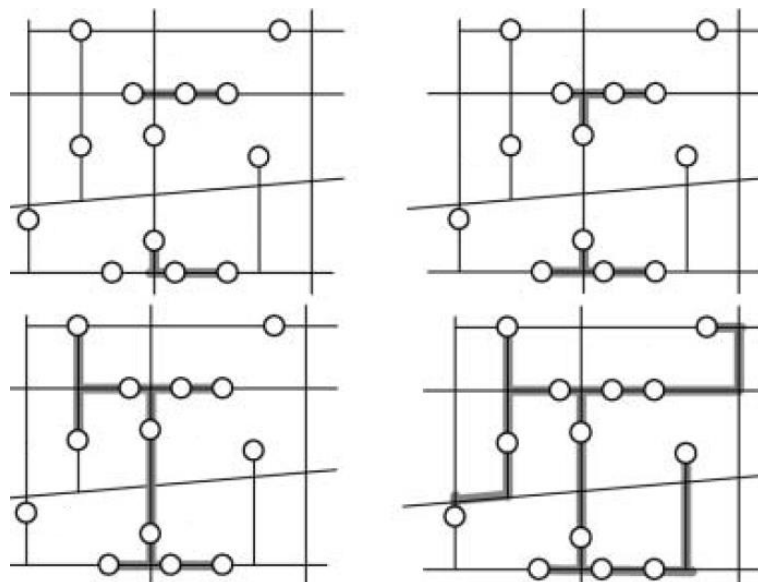
Obrázek 3.6 – Objekty v okolí bodu měření
(Zdroj: Spatial clustering of events on a network, upraveno [29])

⁶ DI – Dangerousness index (Index nebezpečí)

3.3.4 Clumping metoda

Dle definice v článku [30] je jako shluk („clump“) označena množina středů překrývajících se kruhů, vytvořených s daným průměrem okolo analyzovaných bodů. Pro všechny body shluku tedy platí, že jejich vzdálenost od libovolného jiného bodu nesmí překročit stanovenou mez, která se nazývá poloměrem shluku. Stejně jako u předchozích metod, lze také tuto metodu aplikovat na planární i síťová data.

Clumping metoda je schopna určit přesnou lokalizaci shluků. Poloha i počet shluků velmi silně závisí na konkrétním rozložení zkoumaných bodů. I malá změna vstupních údajů (jiná poloha několika málo bodů, jiný poloměr shluku) může vyústit v značně rozdílně nalezené shluky. Jedná se tedy o relativně nestabilní metodu. Na obrázku 3.7 lze vidět průběh vytváření shluků touto metodou. K vymezení hranice zastavení připojování dalších bodů do shluků může být použita metoda Monte Carlo. [28]



Obrázek 3.7 – Clumping metoda

(Zdroj: Detection of multi-scale clusters in network space, upraveno [30])

4 Implementace

V následující kapitole popíši postup tvorby počítačové aplikace. Nejprve byla provedena analýza problému a jeho návrh do programových struktur, na základě kterých jsem mohl posléze přistoupit k samotné implementaci – programování spustitelné aplikace s grafickým uživatelským rozhraním.

4.1 Použité technologie a nástroje

Během vývoje aplikace jsem pracoval s několika existujícími technologiemi a nástroji. V následujících podkapitolách uvedu jejich přehled.

4.1.1 Programovací jazyk Java

Aplikace byla implementována v objektovém programovacím jazyce Java, který se používá od roku 1991. Od této doby bylo vydáno celkem osm hlavních verzí, které postupně do programovacího jazyka vnesly všechny prostředky, které jsou v moderních programovacích jazycích běžné, např. podporu zabezpečení, paralelního zpracování nebo lambda výrazy. [31]

Spuštění aplikací vytvořených v jazyce Java probíhá ve virtuálním prostředí označovaném jako JVM⁷ nebo JRE⁸. Provádění výpočtů ve virtuálním stroji v sobě nese benefit v podobě vyššího bezpečí fyzického stroje oddělením běhového prostředí aplikací od operačního systému uživatele. Nevýhodou může být pomalejší provádění výpočtu oproti použití nativního binárního programu připraveného pro konkrétní operační systém.

Příprava aplikací do spustitelné podoby probíhá ve dvou krocích. Zdrojový kód aplikace připravený programátorem je zkompileován do tzv. „bajtkódu“, což je přenositelný kód určený ke spuštění ve virtuálním prostředí. Tento kód je v případě této aplikace distribuován uživateli v podobě kompresního souborového formátu „jar“⁹, který se používá pro distribuci aplikací napsaných v jazyce Java. Na počítači uživatele následně proběhne překlad bajtkódu do strojové podoby určené pro konkrétní systém a procesor.

Jazyk Java jsem pro implementaci aplikace zvolil především z důvodu snadné přenositelnosti mezi různými operačními systémy (Windows, Linux, ...) a dostupnosti kvalitních knihoven, jejichž použití umožní rychlý vývoj nového software.

⁷ JVM – Java Virtual Machine (Virtuální běhové prostředí pro jazyk Java)

⁸ JRE – Java Runtime Environment (Běhové prostředí pro aplikace v jazyce Java)

⁹ Jar – Java Archive (Kompresní formát pro jazyk Java)

4.1.2 Knihovna jFreeChart

Knihovna jFreeChart umožňuje vykreslování různých druhů grafů (např. základní sloupcové či koláčové) a obdobných dvourozměrných či třírozměrných vizualizací. Kromě zobrazení vygenerovaných grafů je možno tyto grafy uložit do několika grafických formátů a také s nimi z části interaktivně pracovat. [32]

Propojení s touto knihovnou je realizováno zahrnutím jejího kódu do zdrojového kódu aplikace a volání její funkcionality pomocí definovaného API¹⁰. V aplikaci knihovnu používám při zobrazení vypočteného průběhu hustoty rozdělení pravděpodobnosti. Tato funkcionality nebyla explicitně v zadání požadována, považoval jsem však za vhodné poskytnout uživateli názornou představu o vypočtených výsledcích.

4.1.3 Knihovny Apache Commons

Projekt Apache Commons se zaměřuje na tvorbu znovupoužitelných knihoven obsahujících často prováděné operace, které ve svém kódu opakovaně používá velký počet programátorů. Výhodou užití těchto knihoven je vyšší pravděpodobnost korektní funkcionality nabízených operací, jelikož je otestovala širší skupina uživatelů. [33]

V představované aplikaci je použita knihovna Apache Commons IO, která poskytuje funkcionality pro práci se souborovým systémem. Je využita k načítání vstupních souborů ve formátu „csv“¹¹ a k zápisu výstupních souborů rovněž ve stejném formátu. Dále bylo experimentováno s knihovnou Apache Commons Math, která poskytuje množství předprogramovaných matematických funkcí a algoritmů. Po zvážení bylo ale v této oblasti použito pouze standardních knihovnických funkcí jazyka Java.

4.1.4 Vývojové prostředí NetBeans IDE

Jedná se o vývojové prostředí, které původně vzniklo v roce 1996 jako studentský projekt na Matematicko-fyzikální fakultě Univerzity Karlovy v Praze. Pro svoji vzrůstající oblíbenost bylo následně koupeno společností Sun Microsystems, která se později následkem fúze stala součástí společnosti Oracle Inc. I přes vstup do komerčního sektoru byla zachována bezplatná dostupnost a otevřenost kódu vývojového prostředí. [34]

¹⁰ API – Application programming interface (Rozhraní aplikace – funkcionality, kterou lze volat a použít z jiných aplikací.)

¹¹ CSV – Comma-separated values (Soubor s hodnotami oddělenými vybraným oddělovačem, typicky čárkou či středníkem.)

Užití kvalitního vývojového prostředí napomáhá k rychlému a chyb prostému vývoji nových aplikací. Je prováděna průběžná syntaktická kontrola kódu, která upozorní na výslovné i potencionální chyby dříve, než dojde ke spuštění programu. Odstranění sémantických pochybení je pochopitelně stále v režii autora vznikající aplikace.

4.1.5 Launch4j

Protože jednoduchost aplikace měla spočívat i ve snadném postupu jejího spuštění, použil jsem při vytváření spustitelného balíčku s aplikací nástroj Launch4j. Pomocí tohoto nástroje je možno redukovat počet distribuovaných souborů a celou aplikaci přichystat do jednoho souboru obsahujícího vytvořenou aplikaci spolu s použitými knihovnamí. Takto zabalenou aplikaci pak lze poskytnout uživatelům operačního systému Windows v nativním spustitelném formátu souboru s příponou „.exe“¹². [35]

Do stejného souboru je možno také přibalit celé virtuální prostředí pro běh Java aplikací (JRE), uživatel pak může aplikaci spustit i bez jeho trvalé instalace v počítači. Tato alternativa zvýší velikost spustitelného souboru na přibližně 120 MB. Velikost samotné aplikace je cca 3 MB. Tato varianta distribuce zvolena nebyla, a to i z důvodu, že soubor vytvořený tímto způsobem je specifický pro daný operační systém a prostředí. Při problémech se spuštěním aplikace spočívajících v nedostupnosti příslušné verze JRE je zobrazen uživateli postup instalace prostředí včetně webové adresy, odkud je možno jej stáhnout.

4.2 Návrh aplikace

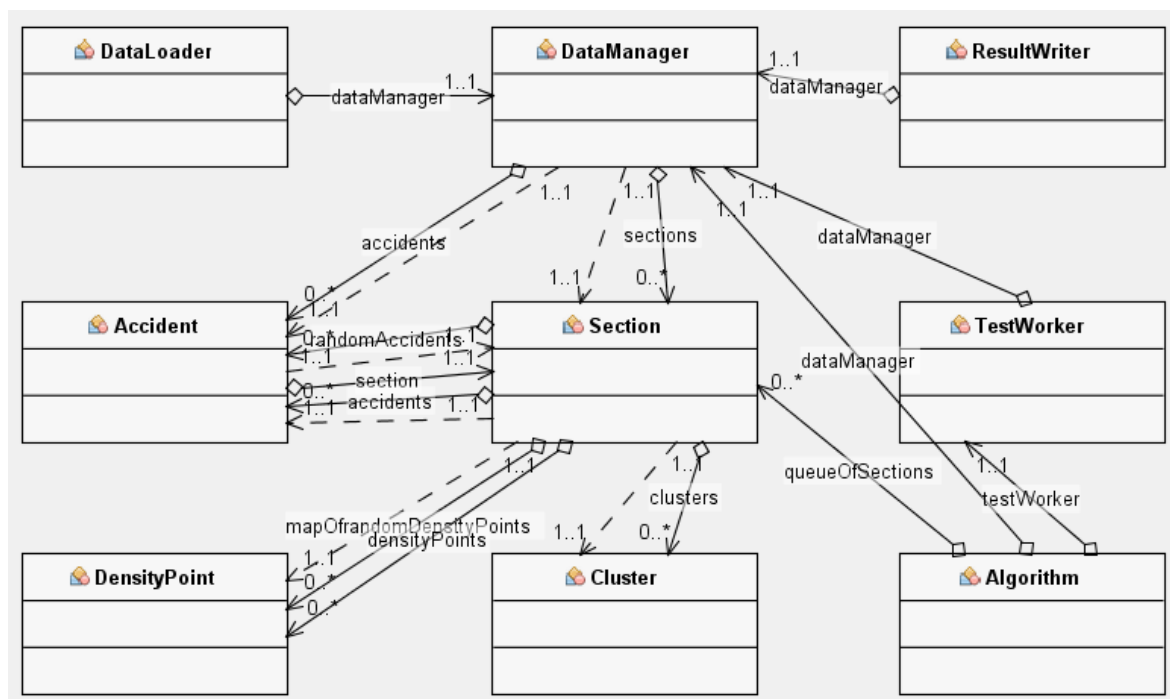
Návrh aplikace je fáze vývoje software, v rámci které probíhá přetvoření požadavků získaných od zákazníka či potřeb vzniklých analýzou problémové oblasti do programových struktur, které budou později programátorem implementovány. K vyjádření návrhu lze využít sadu grafických diagramů UML¹³, pomocí kterých lze snadno získat přehled o navrhovaném software i bez nutnosti detailního zkoumání zdrojového kódu. Soubor diagramů UML obsahuje přibližně 14 typů diagramů, pomocí kterých lze modelovat různé aspekty programového produktu. Užití většiny diagramů je vhodné pro rozsáhlé a kritické aplikace. U menších aplikací obvykle dostačuje použití jen malé podmnožiny z nabízených prostředků. [36]

Jelikož bylo na začátku vývoje aplikace rozhodnuto, že bude implementována objektivním způsobem, jako vhodné se nabízelo užití diagramu tříd, jenž poskytuje základní představu o objektech vyskytujících se v software. Diagram obsahuje obdélníky vyjadřující

¹² Exe – Executable file (Spustitelný soubor)

¹³ UML - Unified Modeling Language (Standardizovaný jazyk pro modelování software)

samostatné programové entity – objekty, které mají v sobě zapouzdřenou jak funkcionalitu, tak také stav (atributy). Obdélníky jsou na diagramu tříd propojeny spojnicemi, jež znázorňují vazby mezi objekty.



Obrázek 4.1 – Diagram tříd dle UML

Na diagramu 4.1 lze vidět zjednodušený diagram tříd popisované aplikace. Reálně se v ní nachází více tříd, jejich kompletní grafické znázornění by však bylo na malém tiskovém prostoru nepřehledné. V horním řádku jsou na diagramu znázorněny třídy, které zajišťují manipulaci s daty. Třída `DataLoader` provádí načítání vstupních dat ze souborů a kontrolu jejich konzistence. Načtená datová struktura je předána třídě `DataManager`, jež zajišťuje uložení informací o silničních nehodách a úsecích během výpočtu a při prezentaci získaných výsledků. Po ukončení výpočtu jsou výsledky předány třídě `ResultWriter`, která zajistí jejich zápis do výstupních souborů.

Datová struktura, kterou v paměti aplikace spravuje třída `DataManager`, a nad kterou je prováděn výpočet, v sobě zahrnuje několik objektů vyskytující se v reálném světě. Třída `Section` odpovídá silničnímu úseku. Mezi atributy, které jsou vlastností každé instance objektu tohoto typu, patří např. název silničního úseku a jeho délka. Ke každému úseku může být přiřazen určitý počet dopravních nehod, které jsou reprezentovány třídou `Accident`. Parametrem této třídy je především lokalizace dané nehody (staničení) vzhledem k odpovídajícímu silničnímu úseku. Nalezené shluky dopravních nehod jsou ukládány

do objektu typu `Cluster`, jenž je definován počátečním a koncovým staničením a významností daného shluku. Třída `DensityPoint` je pomocná třída zajišťující uložení hodnot hustoty pravděpodobnosti v diskretních bodech. Při vizualizaci je uživateli prezentována zdánlivě spojitá funkce hustoty pravděpodobnosti, která vznikla spojením těchto diskretních bodů. Třídy `TestWorker` a `Algorithm` zajišťují paralelní průběh výkonné části výpočtu.

4.3 Definice souborových formátů

Pro potřebu importu vstupních dat z jiných aplikací a exportu spočtených výsledků k dalšímu zpracování bylo třeba navrhnout souborové formáty pro výměnu dat. Jako vstupní i výstupní formát byl zvolen soubor typu „csv“, což je textový formát s lidsky jednoduše čitelnou syntaxí.

4.3.1 Vstupní soubory

Ze vstupních souborů jsou načítány informace o úsecích a na nich se nacházejících nehodách. Při výpočtu je třeba znát název úseku, který je použit k identifikaci úseku ve výsledcích, a jeho délku, jež je potřebná při statistickém zpracování. Dále jsou pak třeba informace o nehodách na úsecích a jejich staničení. Ve vstupních souborech mohou být uvedeny lokalice nehod na více úsecích, jež pak budou zpracovány současně.

Atribut	Typ	Rozsah
<code>název_úseku</code>	Řetězec	1 až 255 znaků
<code>délka_úseku</code>	Desetinné číslo	0.01 až 10^9 s přesností na dvě desetinná místa
<code>staničení_nehody</code>	Desetinné číslo	0.01 až <code>délka_úseku</code> s přesností na dvě desetinná místa

Tabulka 4.1 – Vstupní hodnoty

Během načítání jsou prováděny kontroly vstupních dat. Je ověřována příslušnost hodnot do domén uvedených v tabulce 4.1. Současně je kontrolováno, zda načítaná nehoda leží na některém již importovaném úseku a zda její staničení není vyšší než délka daného úseku. Na obrázku 4.2 lze vidět předpis struktury těchto souborů s příklady specifikace konkrétních úseků a nehod.


```
Soubor s úseky:  
název_úseku;délka_úseku
```

```
Příklad souboru:  
Silnice 413;1830  
52654;502.6
```

```
Soubor s nehodami:  
název_úseku;staničení_nehody
```

```
Příklad souboru:  
Silnice 413;200  
Silnice 413;250.2  
52654;315
```

Obrázek 4.2 – Vstupní formát souboru

4.3.2 Výstupní soubor

Nalezené shluky napříč všemi úseky jsou uloženy do jednoho výstupního souboru, který je možno použít pro další zpracování (např. vizualizaci v jiném software). Pro usnadnění optické kontroly výsledků jsou jednotlivé shluky lexikograficky seřazeny dle názvu odpovídajícího úseku, dále pak od nejnižšího staničení počátku shluku.

```
Výstupní soubor se shluky:  
název_úseku;začátek_shluku;konec_shluku;  
počet_nehod;síla_shluku
```

```
Příklad výstupního souboru:  
Silnice 413;250;292.5;7;0.5234  
Silnice 413;1230;1260;12;0.8500
```

Obrázek 4.3 – Formát výstupního souboru

Vypisované atributy mohou nabývat hodnot popsanych v tabulce 4.2. Při výpočtu je omezen nejnižší počet nehod, které jsou schopny tvořit shluk, proto jejich počet nemůže být menší než tato předem stanovená mez (implicitně je tato mez nastavena na hodnotu tři nehody). Úseky, na kterých nebyl zjištěn žádný shluk, se ve výpisu nenachází.

Atribut	Typ	Rozsah
název_úseku	Řetězec	1 až 255 znaků
začátek_shluku	Desetinné číslo	0 až délka_úseku s přesností na dvě desetinná místa
konec_shluku	Desetinné číslo	0 až délka_úseku s přesností na dvě desetinná místa
počet_nehod	Celé číslo	minimální počet nehod ve shluku až celkový počet nehod na úseku
síla_shluku	Desetinné číslo	0 až 1 s přesností na čtyři desetinná místa

Tabulka 4.2 – Výstupní hodnoty

4.4 Implementace statistického algoritmu

Statistický algoritmus byl implementován odděleně od programování jiných podpůrných částí, jež tvoří např. grafické rozhraní aplikace, se kterým uživatel interaktivně pracuje. Algoritmus není na grafickém rozhraní závislý a lze jej použít i samostatně. Samotný algoritmus statistické metody má rozsah 608 řádků zdrojového kódu, celou aplikaci pak bez doprovodných knihoven tvoří 4 075 řádků kódu.

Implementace statistické metody byla provedena podle postupu popsaného v kapitole 3.2. Ve zmíněné kapitole je pravděpodobnostní funkce hustoty představená jako spojitá křivka, která probíhá od počátku úseku do jeho konce. Pro účely implementace je ale vhodnější pracovat s křivkou definovanou nad diskrétními hodnotami. Pokud je rozstup diskrétních hodnot dostatečně malý, aproximuje diskrétní křivka křivku spojitou velmi věrohodně. V aplikaci jsou implicitně diskrétní body generovány se vzdáleností 10 metrů. Příklad vygenerování diskrétních bodů ve zdrojovém kódu lze vidět na obrázku 4.4.

```

1 // create density points
2 densityPointsList = new LinkedList<>();
3
4 double nextLocation = start;
5 for (int i = 0; i < numberOfDensityPoints; i++) {
6     DensityPoint ds = new DensityPoint(nextLocation);
7     densityPointsList.add(ds);
8     nextLocation += Algorithm.STEP;
9 }

```

Obrázek 4.4 – Vytvoření diskrétních bodů hustoty

Pro každou nehodu je v každém diskretním bodě hustoty vynesena hodnota Epanechnikova jádra. Metodu zajišťující jeho výpočet lze vidět na obrázku 4.5. Hodnota `dpLoc` odpovídá staničení bodu hustoty, `acLoc` představuje staničení nehody. Celočíslnou proměnnou `bw` je vyjádřena nejistota v měření místa nehody, která se projeví v šířce jádra.

```

1     private double jadro(double dpLoc, double acLoc, int bw){
2         double power = Math.pow((dpLoc - acLoc) / bw, 2);
3         return (3.0 / (4.0 * bw)) * (1 - power);
4     }

```

Obrázek 4.5 – Vzorec pro výpočet Epanechnikova jádra

V průběhu metody Monte Carlo je třeba generovat náhodné virtuální body nehod, které jsou využity k ověření hypotézy o rovnoměrném rozdělení reálných nehod na silničním úseku. K tomuto účelu je použito generátoru pseudonáhodných čísel poskytovaném ve standardním balíčku `java.util.Random` jazyka Java. V ideálním případě by bylo lépe použít skutečného generátoru náhodných čísel. Takový generátor však obvykle ke svému provozu potřebuje hardwarový zdroj nahodilosti, jakým může být např. šumová dioda nebo radioaktivní zářič, a nelze předpokládat, že takovým zařízením bude uživatel disponovat.

Výše uvedené ukázky zdrojových kódů jsou pouze ilustrativní, kompletní zakódování algoritmu metody KDE+ je k nalezení v elektronické příloze práce. Seznam souborů elektronické přílohy práce je uveden v příloze A.

4.5 Paralelizace

První experimentální verze aplikace, na které byla ověřována základní funkcionální a správnost vypočtených výsledků, byla implementována sekvenčně. Všechny úseky byly zpracovávány v jednom vlákně jediným procesorovým jádrem, což lze považovat u soudobých víceprocesorových systémů za zbytečné plýtvání zdroji. Vzhledem k tomu, že je zpracování jednotlivých úseků na sobě navzájem nezávislé, vcelku přirozeně se zde nabízela možnost rozvětvení algoritmu do několika paralelních výpočetních větví a z toho plynoucího několikanásobně rychlejšího dokončení výpočtu vstupního souboru úseků. Pokud vstupní data obsahují alespoň dva úseky, je možno provést jejich paralelní zpracování.

Implicitně je v aplikaci nastaven počet použitých výpočetních vláken v průběhu výpočtu na hodnotu rovnou počtu logických jader fyzického procesoru, od kterého je odečtena malá konstanta (1 nebo 2), jako rezerva výpočetního výkonu na správu spuštěných vláken a udržení interaktivity grafického rozhraní bez uživatelsky nepříjemného zamrznutí okna. Maximální počet vláken lze také explicitně specifikovat při spuštění aplikace.

4.6 Internacionalizace

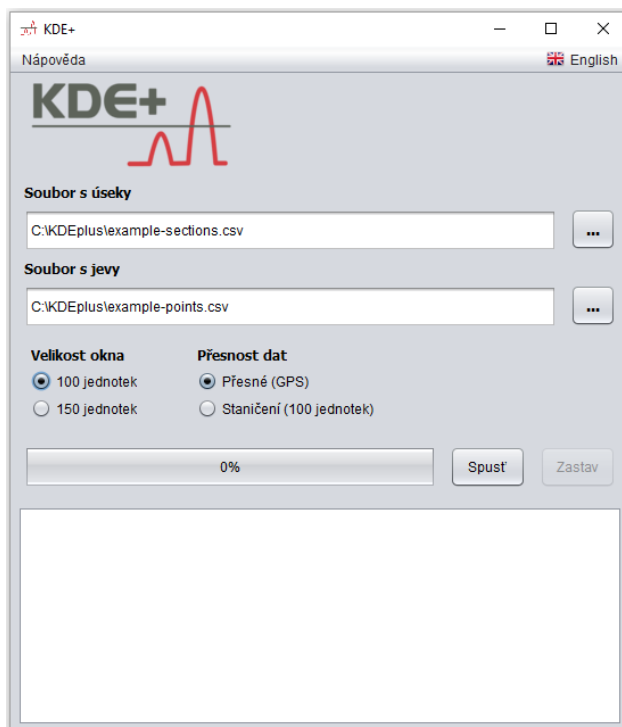
Jedním z požadavků na vznikající aplikaci byla možnost její lokalizace alespoň do dvou jazyků. Zvolené řešení je obecnější, umožňuje přidání libovolného počtu různých jazyků. V předkládané verzi aplikace je možno pracovat se dvěma jazykovými mutacemi, s českou a anglickou. V případě potřeby lze relativně snadno přidat překlady do dalších jazyků vytvořením nové varianty lokalizačního souboru. Všechny použité textové popisky pro jednu jazykovou mutaci jsou totiž uloženy v jediném textovém souboru. Na úrovni zdrojového kódu je lokalizace textových řetězců zajištěna pomocí objektů `ResourceBundle`.

Po spuštění aplikace je zjištěno jazykové nastavení operačního systému uživatele a dle detekovaného jazyka se zvolí odpovídající verze aplikace. V případě použití české či slovenské varianty operačního systému je aplikace implicitně zobrazena v češtině. V jakémkoli jiném případě se aplikace spustí v anglické jazykové mutaci. Mezi jazyky lze přepínat v horním panelu hlavního okna aplikace.

4.7 Grafické rozhraní

Grafické uživatelské rozhraní bylo navrhováno s ohledem na snadnou použitelnost a uživatelskou přívětivost. Při splnění těchto ergonomických aspektů by uživatel neměl příliš váhat nad účelem ovládacích prvků aplikace. Z tohoto důvodu bylo při jejím vývoji prováděno průběžné uživatelské testování grafického rozhraní, které spočívalo v analýze postupu práce s aplikací u předem neproškoleného uživatele.

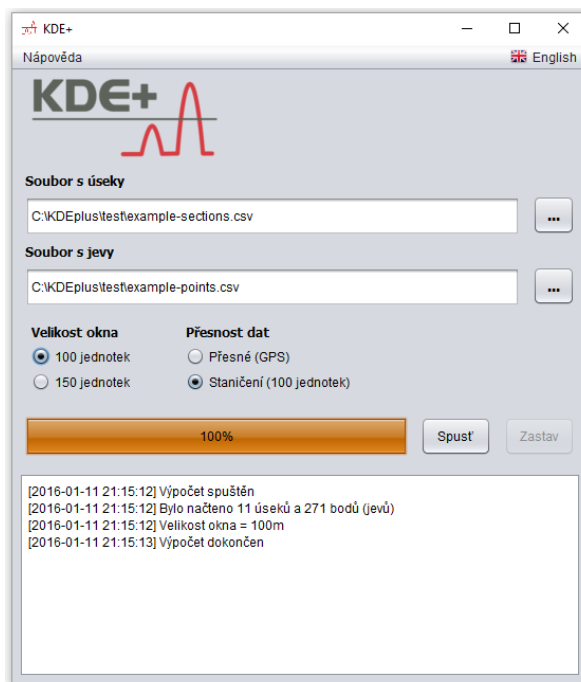
Při návrhu grafického rozhraní aplikace bylo stanoveno, že software bude s uživatelem komunikovat převážně pomocí jedné obrazovky – hlavního okna. Po spuštění aplikace v grafickém režimu je na obrazovce zobrazeno hlavní okno, jehož podoba je znázorněna na obrázku 4.6. V rámci hlavního okna je možno zadat všechny nutné vstupní parametry výpočtu a poté kliknutím na tlačítko „Spust“ zahájit výpočet. Mezi povinné vstupní parametry patří cesty k souborům, ze kterých budou načtena data. Dále je možno zvolit velikost okna (šířku Epanechnikova jádra) a přesnost dat. U popisků jednotlivých voleb není specifikována konkrétní měrná jednotka, ve které jsou vstupní data uložena. Data mohou být v metrech, stopách či jakékoli jiné jednotce. Z toho důvodu je v popisku použito pouze obecného pojmu „jednotka“.



Obrázek 4.6 – Hlavní okno aplikace

Po spuštění výpočtu jsou nejprve načtena data o úsecích a nehodách do vnitřních struktur aplikace. V rámci načítání je provedena série kontrol zajišťující validitu vstupních dat. Pokud se během načítání narazí na chybu, je tato chyba i s vysvětlujícím textem vypsána do textového pole v dolní části aplikace. Na stejné místo jsou vypisována také hlášení o úspěšném začátku, konci a aktuálním stavu výpočtu. Pokud jsou vstupní data v pořádku, dojde ke spuštění samostatných výpočetních vláken a k postupnému zpracovávání všech importovaných úseků.

Postup výpočtu je znázorněn změnou zbarvení obdélníkové plochy (indikátoru průběhu) a číslem vyjadřujícím procento již zpracovaných úseků. Doba výpočtu je závislá na počtu silničních úseků, jejich délce, počtu dopravních nehod a přirozeně také na rychlosti počítače. Zpracování je možno kdykoliv přerušit stisknutím tlačítka „Zastav“. Na obrázku 4.7 je snímek hlavního okna po úspěšném dokončení výpočtu. Z hodnot časů v textovém poli lze odečíst délku trvání výpočtu.



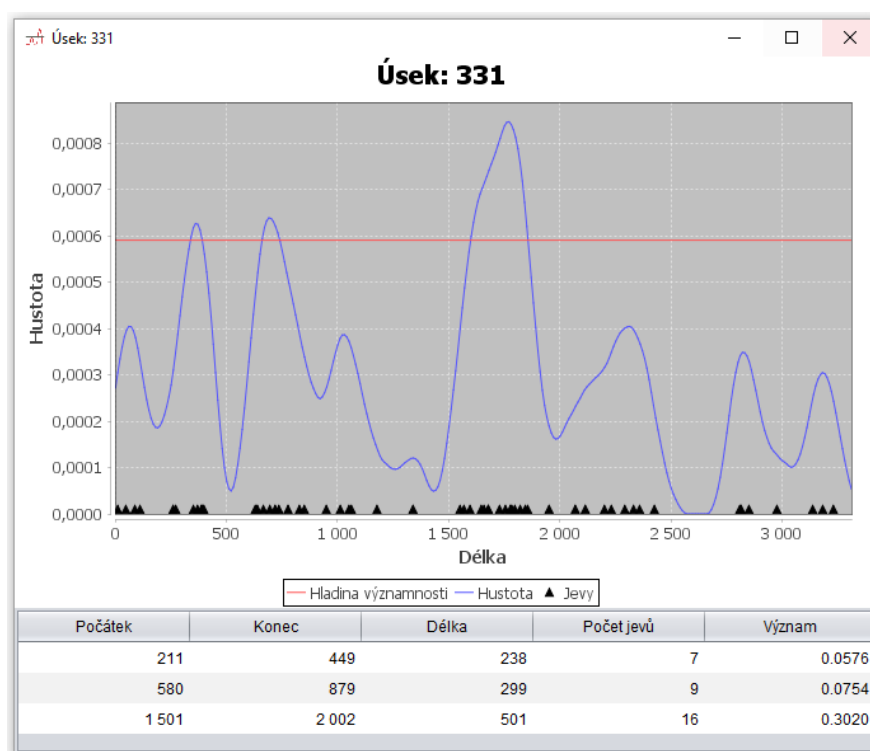
Obrázek 4.7 – Hlavní okno aplikace po dokončení výpočtu

Po ukončení výpočtu je kromě souhrnných informací v hlavním okně zobrazeno nové okno s tabulkou obsahující seznam všech zpracovaných úseků (Obrázek 4.8). Kromě popisných údajů ze zadání (názvu úseku a jeho délky) je u každého úseku zobrazen počet nalezených shluků. U těch úseků, na jejichž řádcích je zobrazen symbol lupy, je možno získat detailní informace kliknutím na symbol lupy. Úseky, u kterých se lupa nenachází, byly při zpracování přeskočeny, a proto k nim není možno zobrazit žádné další podrobnosti. Důvodem přeskočení může být nedostatečná délka úseku, či nulový nebo velmi malý počet dopravních nehod, který nemá smysl dále statisticky zpracovávat. Kromě zobrazení tabulky se zpracovanými úseky jsou výsledky uloženy do výstupního souboru v dříve definovaném formátu.

Úseky	Délka	Bodů (jevů)	Shluky	Detaily
203	5 783	37	2	
310	168	0	0	
311	4 272	21	2	
312	136	0	0	
327	4 746	52	3	
328	3 669	26	3	
329	597	2	1	
330	4 773	54	2	
331	3 312	57	3	
435	2 707	14	1	
road 01	2 000,5	8	2	

Obrázek 4.8 – Výpis seznamu zpracovaných úseků

V okně s detailními informacemi o jednom úseku (Obrázek 4.9) je vidět grafické znázornění aplikace metody KDE+ nad konkrétním úsekem. Modrou křivkou je znázorněn průběh hustoty rozdělení pravděpodobnosti, která se zakládá na rozmístění nehod, jež se na daném úseku (bodě osy x) udály. Čím více nehod bylo zaznamenáno v jednom bodě či jeho okolí, tím je hodnota pravděpodobnostní funkce hustoty vyšší. Naopak, pokud nebude ve stanovené vzdálenosti od konkrétního bodu osy x zaznamenána žádná nehoda, hodnota funkce v tomto bodě bude nulová. Samotné nehody jsou znázorněny černými značkami (trojúhelníky) v dolní části grafu. Z důvodu typicky velkého počtu nehod dochází často k jejich překrývání.



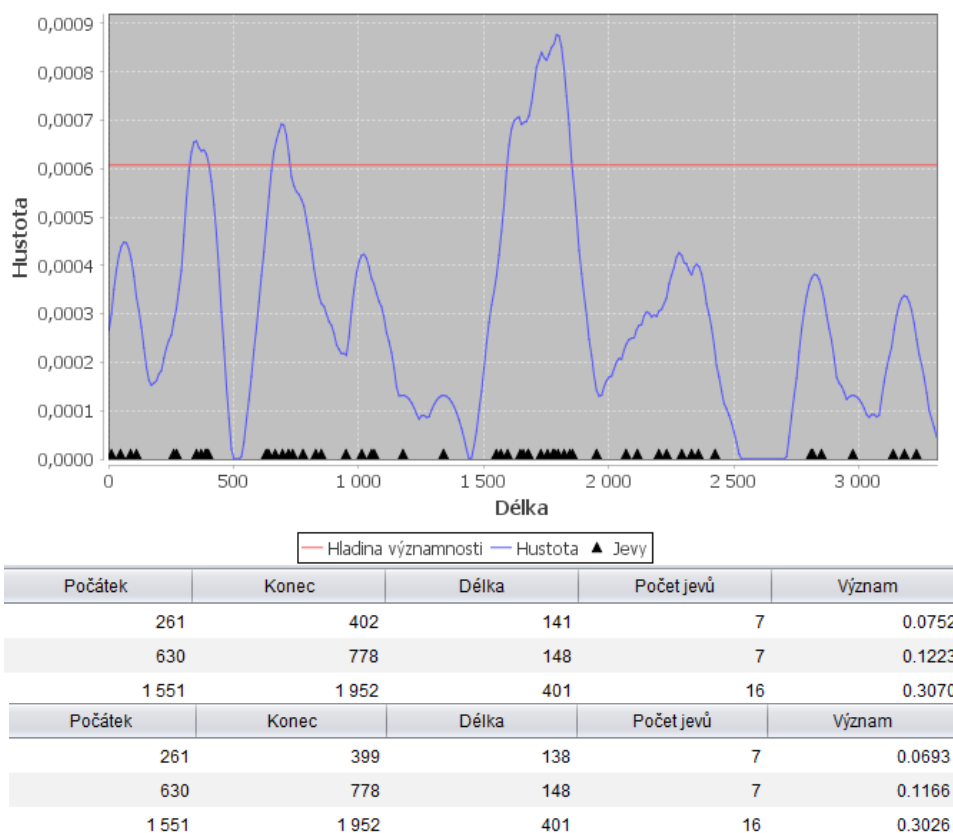
Obrázek 4.9 – Vizualizace zpracování jednoho úseku

Červenou horizontálně vedenou přímkou je znázorněna vypočtená hladina významnosti. Jako shluk nehod je označen takový interval úseku, nad nímž hodnota pravděpodobnostní funkce hustoty vystoupila nad hodnotu hladiny významnosti. Na výše uvedeném snímku (Obrázek 4.9) pozorujeme výskyt tří shluků dopravních nehod. Výčet shluků i s upřesněním jejich lokalizace, počtem obsažených nehod a jejich významností, je uveden v tabulce nacházející se v dolní části okna. Na jednom úseku platí, že s vyšší hodnotou funkce hustoty pravděpodobnosti nad hladinou významnosti má shluk vyšší význam. Aby bylo možno porovnat shluky napříč více úseky, je při výpočtu významu zahrnuta i výše hladiny významnosti.

4.8 Automatizované testování

Testování patří k vývoji každé aplikace. Pro testovací účely jsem vytvořil několik pokusných vstupních souborů s validními i chybnými daty, abych mohl při úpravách aplikace rychle otestovat, zda nebyla touto modifikací narušena její funkcionality, či zda aplikace korektně odhalila nesprávné zadání výpočtu. V aplikacích užívajících pouze deterministické algoritmy můžeme předpokládat, že pro stejné vstupní hodnoty budou vypočteny vždy totožné výsledky. Jelikož ale metoda KDE+ obsahuje také stochastický algoritmus (Monte Carlo), nemusí nutně vést dva výpočty se stejným zadáním k naprosto identickému řešení. Proto bylo potřeba při ověřování testováním vypočtených výsledků oproti očekávaným výsledkům dovolit experimentálně stanovenou odchylku.

Na obrázku 4.10 lze pozorovat mírně rozdílné výsledky ze dvou běhů algoritmu. Na grafu průběhu hustoty žádný rozdíl nebyl patrný, proto jej zde ani duplicitně neuvádím. Z dat v tabulce výsledků lze vyčíst, že se o tři metry liší staničení konce prvního shluku a také mírně význam všech shluků. Tuto nepřesnost jde do určité míry eliminovat zvýšením počtu opakování simulací Monte Carlo.



Obrázek 4.10 – Rozdílné výsledky dvou běhů výpočtu

4.9 Měření rychlosti výpočtu

V rámci testování software jsem analyzoval závislost délky trvání výpočtu na rozsahu a podobě vstupních dat. Pro testování rychlosti aplikace byl použit konvenční osobní počítač s vícejádrovým procesorem¹⁴. Bylo provedeno několik testovacích běhů lišících se počtem silničních úseků určených ke zpracování. Zadání výpočtu a naměřené hodnoty jsou k nalezení v tabulce 4.3.

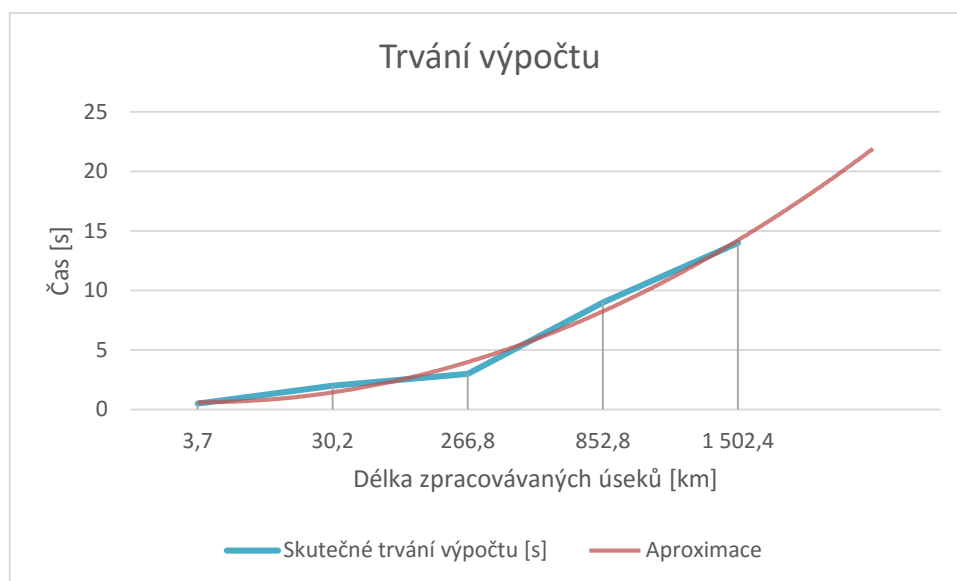
Počet úseků	Celková délka úseků [km]	Celkový počet nehod	Čas výpočtu 1 úseku [s]	Celkový čas výpočtu [s]
1	3,7	26	0,5	0,5
10	30,2	263	0,2	2
100	266,8	2 015	0,03	3
300	852,8	5 917	0,03	9
615	1 502,4	10 820	0,02	14

Tabulka 4.3 – Porovnání rychlosti výpočtu

Před začátkem měření jsem stanovil předpoklad, že trvání výpočtu bude záviset na celkové délce zpracovávaných výsledků přibližně lineárně. Trvání výpočtu závisí i na počtu nehod ležících na úsecích. Vzhledem k tomu, že lze předpokládat jejich statisticky obdobné rozložení napříč větším množstvím úseků, nebyl počet nehod do předpokladu zahrnut. Důvodem předpokládané lineární závislosti je implementace algoritmu, v rámci které je pro každý diskrétní bod funkce hustoty na úseku při inkrementaci hodnoty funkce procházeno jeho okolí.

Závislost trvání výpočtu je vynesena na grafu 4.1. Modrou křivkou jsou vyneseny skutečně naměřené hodnoty, oranžová spojnice trendu aproximuje nárůst délky trvání výpočtu. Závislost celkové délky úseků a trvání výpočtu je tedy horší než lineární, avšak není kvadratická (či obecně polynomiální). Aplikaci lze tedy s rozumnou dobou trvání použít i na data většího rozsahu.

¹⁴ Procesor Intel Core i5 4460S (4 jádra, 2.9 GHz), operační paměť 8 GB DDR3, SSD disk.



Graf 4.1 – Závislost velikosti vstupních dat na rychlosti výpočtu

5 Použití a vyhodnocení

V této kapitole popíši proces přípravy vstupních dat v software typu GIS¹⁵, následně uvedu postup práce s vytvořenou aplikací a poté se budu zabývat vizuální prezentací výsledků zpět v GIS software. Na závěr uvedu výsledky otestování aplikace spolu s několika příklady jejího možného využití.

5.1 Příprava vstupních dat

Jak bylo popsáno v kapitole 4.3.1, jsou k nalezení shluků potřeba dvě datové sady (definice linií – silničních úseků a lokalizace bodů – dopravních nehod), které jsou uloženy ve dvou textových souborech. Pro málo rozsáhlá data lze vstupní soubory vytvořit ručním zadáním informací o úsecích a nehodách v libovolném textovém editoru. Při zpracování většího počtu úseků a nehod se již neobejdeme s alespoň částečně automatizovaným převodem z dvourozměrné prostorové reprezentace.

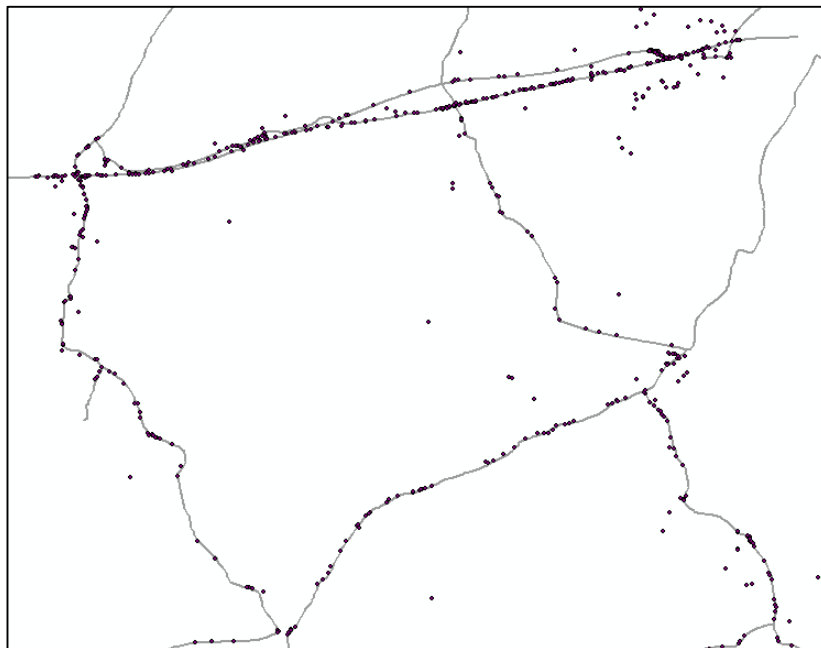
5.1.1 Převod z dvourozměrného formátu

Prostorová data jsou ukládána v některém k tomuto účelu vhodném formátu. Tímto formátem může být například ESRI¹⁶ shapefile. Mapové podklady uložené v tomto formátu se nacházejí v několika souborech. Hlavní soubor s příponou „shp“ lze importovat do vybraného GIS software. Já jsem k tomuto účelu využil komerční aplikace ArcMap, nacházející se v softwarovém balíku pro zpracování geografických dat ArcGIS 10.3. Příklad konkrétních mapových podkladů je znázorněn na obrázku 5.1. Šedou barvou jsou zde vyznačeny linie, reprezentující silniční úseky. Drobnými body jsou pak vyneseny zaznamenané dopravní nehody.

Nejprve je nutno určit linie, které budou dále zpracovávány. Ze zvolených linií lze rovnou odfiltrovat takové, které nesplňují podmínky pro analýzu metodou KDE+, což jsou úseky kratší než 200 či 300 metrů v závislosti na zvolené přesnosti. Takové úseky by byly aplikací i při jejich zanechání v datové sadě přeskočeny. Výsledkem tohoto kroku je první vstupní soubor obsahující seznam názvů zpracovávaných úseků spolu s jejich délkou.

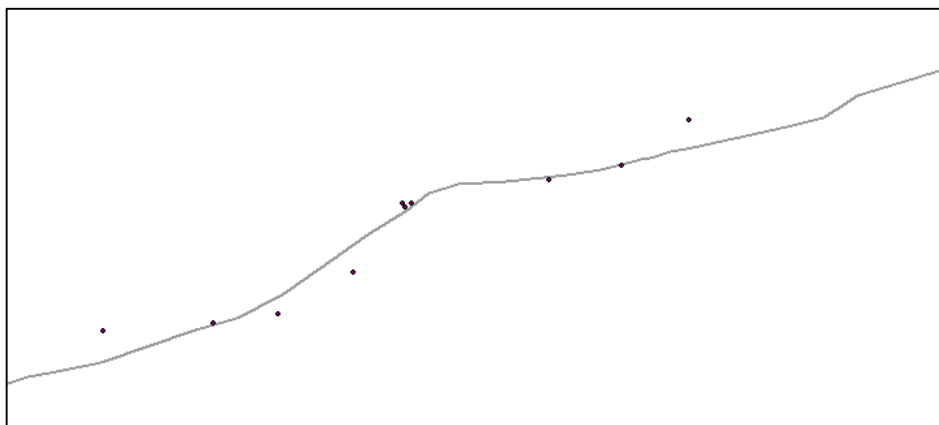
¹⁵ GIS - Geographic information system (Geografický informační systém)

¹⁶ ESRI - Environmental Systems Research Institute (Společnost vyvíjející software pro práci s geografickými informačními systémy)



Obrázek 5.1 – Prostorová reprezentace úseků a nehod

V dalším kroku je třeba vytvořit textový soubor se staničením nehod na vybraných úsecích. Na obrázku 5.2 lze vidět průběh jednoho úseku spolu s nehodami, jejichž souřadnice byly zaznamenány policií. Můžeme pozorovat, že ne všechny nehodové body leží přímo na liniích. Většina bodů je vlivem nepřesnosti systému GPS umístěna v okolí linií. Tyto body je potřeba připojit na linie, což se provádí promítnutím zaměřeného bodu do nejbližšího bodu náležejícího linii. Pokud je nehodový bod vzdálen od linie více než předem zvolená konstanta (typicky 10 m), je ignorován. Pokud blízko sebe probíhá více linií, je možno z vlastností připojených k bodu nehody zjistit, ke které linii má být nehoda připojena. Vzhledem k úrovni nepřesnosti to nemusí nutně být ta nejbližší ležící. Po připojení všech bodů lze vypočítat hodnotu staničení, tedy vzdálenost nehody od počátku linie, který má staničení rovno nule. Nejvýše pak může staničení nabývat hodnoty rovnající se délce linie. Technika měření vzdálenosti bodů podél linií bývá označována jako lineární referencování.



Obrázek 5.2 – Průběh úseku s nehodami

Po exportu přiřazených nehod s vypočteným staničením do textové podoby jsou již připraveny oba potřebné soubory, které aplikace KDE+ při svém spuštění požaduje.

5.2 Postup spuštění a práce s aplikací

Pro vyzkoušení aplikace je nutno provést několik dále popsanych kroků. Jako vstupní data budou v tomto příkladu použity testovací vzorové soubory doprovázející aplikaci. Podobný návod ke spuštění je k nalezení také v souboru `readme.txt`.

Nejprve je nutno rozbalit připravený balíček z elektronického archivu závěrečné práce, nebo z příloženého CD. Nejnovější verzi aplikace je možno stáhnout z webové stránky www.kdeplus.cz. Z důvodu evidence uživatelů aplikace je na zmíněném webu vyžadována bezplatná registrace.

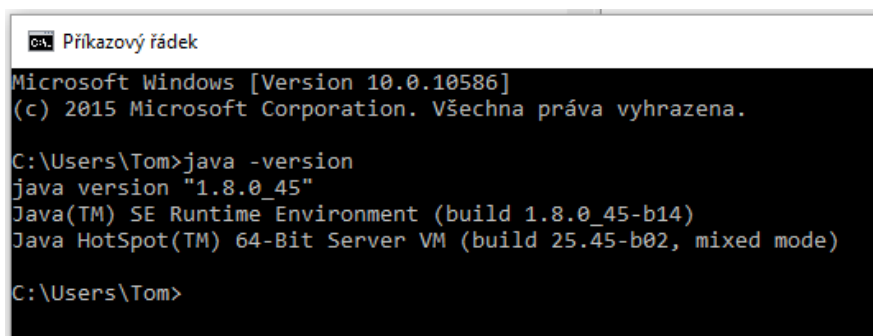
Název	Přípona	Velikost	Datum	Čas
..		ADR	21.5.2016	12:26:16
src		ADR	21.5.2016	12:26:15
config.properties		1,43 KB	19.5.2016	15:05:32
example-points.csv		2,58 KB	20.6.2015	11:31:58
example-sections.csv		113 B	20.6.2015	10:51:19
KDEplus.exe		2,55 MB	19.5.2016	15:41:25
KDEplus.jar		2,29 MB	19.5.2016	15:41:05
manual.pdf		284 KB	19.5.2016	15:16:52
readme.txt		2,32 KB	19.5.2016	14:59:00

Obrázek 5.3 – Obsah balíčku s aplikací

Rozbalením komprimovaného souboru s příponou ZIP získá uživatel ve zvolené složce následující soubory:

test – složka se vzorovými daty
 KDEplus.exe – spustitelný soubor (případně lze použít i verzi s příponou jar)
 config.properties – soubor s implicitním nastavením aplikace
 readme.txt – základní informace o spuštění
 Manual.pdf – kompletní nápověda

Pro běh programu je nutno mít v počítači nainstalován virtuální stroj, který umožňuje spuštění aplikací v jazyce Java alespoň sedmé verze. Může jím být např. Java Virtual Machine (JVM), jenž je volně ke stažení na oficiálních webových stránkách společnosti Oracle na adrese www.java.com/en/download. Přítomnost odpovídajícího JRE lze ověřit zadáním příkazu `java -version` (v systému unixového typu v terminálu, ve Windows je možno provést v příkazovém řádku). Výsledek by měl přibližně odpovídat výpisu znázorněnému na následujícím obrázku (Obrázek 5.4). Především je třeba věnovat pozornost prvnímu řádku, na kterém musí být uvedena alespoň verze 1.7.



```

cs: Příkazový řádek
Microsoft Windows [Version 10.0.10586]
(c) 2015 Microsoft Corporation. Všechna práva vyhrazena.

C:\Users\Tom>java -version
java version "1.8.0_45"
Java(TM) SE Runtime Environment (build 1.8.0_45-b14)
Java HotSpot(TM) 64-Bit Server VM (build 25.45-b02, mixed mode)

C:\Users\Tom>
  
```

Obrázek 5.4 – Ověření přítomnosti a verze JRE

Dále je možno v souboru `config.properties` ověřit implicitní nastavení aplikace použité při inicializaci GUI, či nastavení pro dávkové provádění. Vybrané důležité parametry:

`sections` – cesta k souboru s definicí linií (úseků)
`accidents` – cesta k souboru s definicí jevů (nehod)
`output` – název souboru pro uložení nalezených shluků
`bandwidth` – úroveň nejistoty přesnosti lokalizace jevu
`dataType` – typ vstupních dat

Existují dvě možnosti spuštění aplikace:

1. V grafickém rozhraní

Aplikace se spouští poklepnutím na soubor `KDEplus.exe` či `KDEplus.jar`. V případě, že se vám při poklepnutí na soubor s příponou

jar pouze zobrazí seznam komprimovaných souborů, použijte pravého tlačítka myši a zvolte ze zobrazeného kontextového menu volbu Otevřít, či v příkazovém řádku použijte příkaz

```
java -jar KDEplus.jar
```

2. Dávkově bez grafického rozhraní

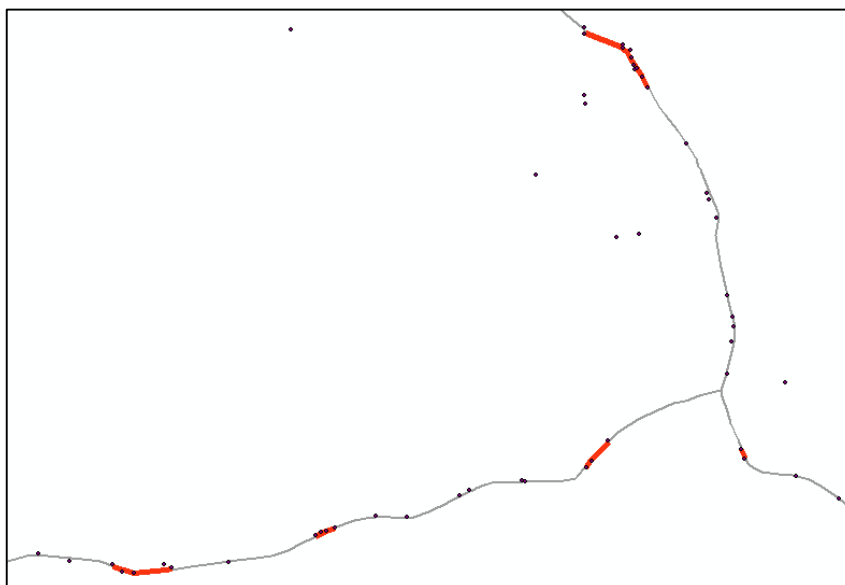
Definujte požadované parametry výpočtu v souboru

```
config.properties a spusťte aplikace provedením příkazu  
java -jar KDEplus.jar -c
```

Během výpočtu se informace o jeho postupu promítají do grafického rozhraní, či do příkazového řádku. Po jeho skončení aplikace uloží výsledky – nalezené shluky – do specifikovaného výstupního souboru a v grafickém rozhraní proběhne jejich vizualizace do podoby grafů s průběhy pravděpodobnostních funkcí hustoty.

5.3 Prezentace vypočtených výsledků

Získaný (jednorozměrný) soubor s nalezenými shluky dopravních nehod je možno prohlížet přímo v textové podobě (ve formátu definovaném v kapitole 4.3.2). Názornější je ale promítnutí získaných dat zpět do dvourozměrné prostorové reprezentace. Tuto projekci lze provést opět za pomoci GIS software. Výřez, na kterém je červenou barvou zvýrazněno pět nalezených shluků na části dopravní sítě uvedené v kapitole 5.1.1, lze pozorovat na obrázku 5.5.



Obrázek 5.5 – Nalezené shluky dopravních nehod

5.4 Aplikace metody KDE+ na nehody v silniční síti ČR

Aplikace byla použita k vyhledání nebezpečných úseků na českých silnicích. Do analýzy byly zahrnuty dopravní nehody vyskytnuvší se na silnicích v extravilánu na úsecích delších než 200 m, jež byly zapříčiněny srážkou se zvěří. Takových nehod se na vybraných komunikacích v souhrnné délce měřících 37 469 km v letech 2009 až 2013 událo dle dostupných záznamů 16 612. [23].

	Délka úseků		Počet shluků		Počet srážek	
	Km	%	počet	%	počet	%
Celá síť	37 469	100	-	-	16 612	100
Nalezené shluky nehod	267	0,71	2060	100	5510	33,2
Nalezené shluky nehod (prvních 100)	20	0,05	100	4,85	402	2,4

Tabulka 5.1 – Aplikace metody KDE+ na silniční síť České republiky
(Tabulka vznikla úpravou tabulky publikované v [23].)

Z tabulky 5.1 lze vyčíst, že z celkové délky úseků v silniční síti bylo za pomoci metody KDE+ nalezeno 267 km potencionálně nebezpečných lokalit, kde vykazují srážky se zvěří významnou tendenci ke shlukování. Pokud je výběr omezen na prvních 100 shluků s nejvyšší silou shluku, jedná se pak o 20 km silnic tvořících 0,05 % z celkové délky sítě, kde je výskyt shluků obzvlášť význačný. Těchto 20 km silnic by bylo vhodné podrobit detailnějším zkoumání a zvážit, zda jsou zde instalovaná nehodová protipatření dostatečná. Takové zkoumání by mohlo zahrnovat provedení místního šetření (bezpečnostní inspekce) na nehodovém úseku.

5.5 Příklady použití

Níže představím dva eventuální příklady aplikace metody KDE+, v rámci kterých může být vhodné použití vytvořeného software. Jelikož se v práci zaměřuji na dopravní problematiku, uvedu příklady z této oblasti. Obecně lze však metodu KDE+ i popisovanou aplikaci užít i pro zjišťování přítomnosti shluků v jiných typech jednorozměrných dat. Např. v temporálních datech, kdy by případným zadáním úlohy mohlo být zjištění tendence ke shlukování

událostí na časové ose (např. výskytu sebevražd v průběhu kalendářního roku). U vícerozměrných dat je třeba nejprve provést jejich transformaci. Příkladem takové transformace je např. i připojení nehodových bodů k úsekům popsané v kapitole 5.1.1.

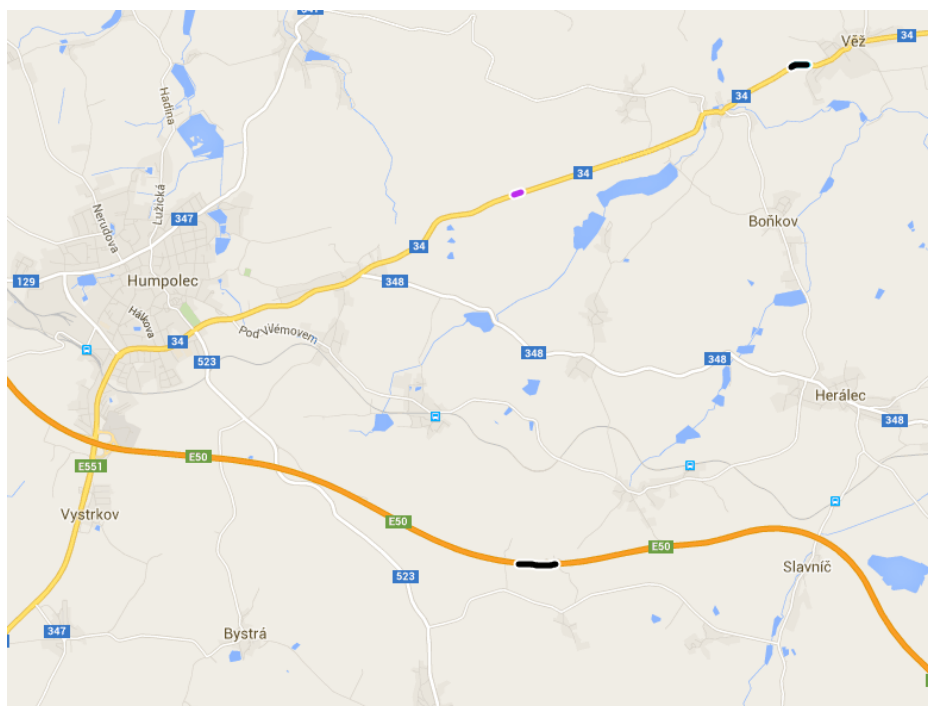
5.5.1 Web „Kde bouráme“

V mapě dostupné na webové prezentaci „Kde bouráme“¹⁷ jsou znázorněny shluky dopravních nehod, které byly zaznamenány v databázi Policie ČR mezi lety 2012 až 2014. Zpracovány jsou nehody na dálnicích, rychlostních silnicích a silnicích první třídy ve správě Ředitelství silnic a dálnic ČR. Aplikací metody KDE+ na tato data bylo označeno 298 významných shluků. Na těchto rizikových úsecích by řidiči měli věnovat jízdě zvýšenou pozornost, či ještě lépe by na nich mohly příslušné úřady provést preventivní bezpečnostní opatření. [37]

Web umožňuje interaktivní procházení mapy. Přiblížením a kliknutím na konkrétní shluk může případný zájemce získat podrobnější informace o tomto shluku. Zvláště informačně přínosná je možnost zobrazení fotografie z místa výskytu shluku ve formě panoramatického pohledu „Street View“, tak jak jej pořídila při fotografování okolí silnic České republiky firma Google Inc.

Na obrázku 5.6 lze nalézt tři shluky dopravních nehod nacházející se poblíž Humpolce, jejichž umístění bylo vypočteno za pomoci metody KDE+. Barva shluku rozlišuje příčinu dopravní nehody (černá – nehoda bez zjištěné příčiny, fialová – srážka se zvířím, modrá – nehoda za mokra). V mapě nejsou zobrazovány jednotlivé dopravní nehody, neboť by z důvodu vysokého počtu výskytů působily rušivě.

¹⁷ Kde bouráme – www.kdebourame.cz



Obrázek 5.6 – Výřez mapy z aplikace www.kdebourame.cz
(Zdroj: Mapový podklad z [37])

5.5.2 Web „Sražená zvěř“

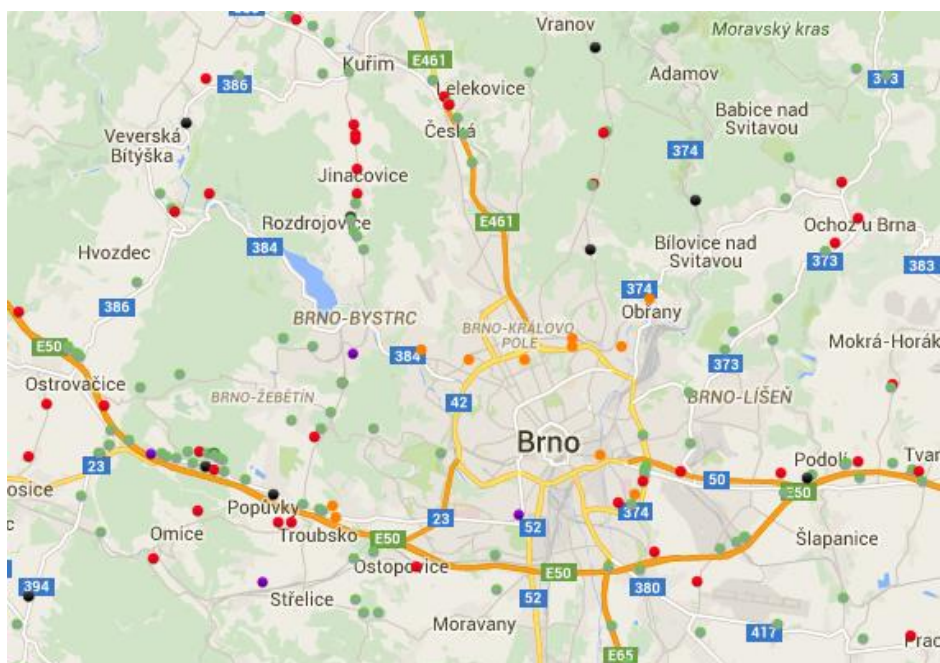
Na webové prezentaci „Sražená zvěř“¹⁸ je možno nalézt evidovaná místa srážek silničního vozidla se zvěří a nově také i se zaběhnutými domácími zvířaty. Data do aplikace mohou dodávat hospodáři z řad myslivců, dobrovolníci skrze webovou prezentaci, či mobilní aplikaci pro operační systém Android. V aplikaci se v současnosti (květen 2016) nachází přes 23 tisíc záznamů o dopravních nehodách tohoto typu. [38]

Dopravní nehody se zvěří patří na silnicích k relativně častým situacím. Účastník provozu má na odezvu zabraňující srážce velmi krátkou reakční dobu. Může docházet k sekundárním srážkám, pokud provede řidič vozidla úhybný manévr a strhne vozidlo do trajektorie jiného vozidla, nebo následkem srážky ztratí nad svým vozidlem kontrolu.

Nad mapovým podkladem je možno prohlížet uložené body dopravních nehod. Barevné rozlišení bodu lze užít k získání přehledu o druhovém zastoupení zvířat na konkrétních úsecích komunikací. Např. celkem pochopitelně je ve městech vyšší počet zaznamenaných srážek se psy. Dále je možno provést přepnutí do analytického pohledu, kdy dochází k agregaci bodů nehod do shluků získaných metodou KDE+. Z umístění těchto shluků lze získat přehled úseků s nejvyšší koncentrací srážek se zvěří, ale je možno také zjistit pravděpodobné

¹⁸ Sražená zvěř – www.srazenazver.cz

migrační trasy volně žijících zvířat. Na obrázku 5.7 lze pozorovat zaznamenané srážky se zvěří v okolí města Brna.



**Obrázek 5.7 – Výřez mapy z aplikace www.srazenazver.cz
(Zdroj: Mapový podklad z [38])**

6 Závěr

V rámci své diplomové práce jsem nejprve provedl základní náhled do problematiky analýzy silničních nehod z hlediska identifikace kritických nehodových lokalit. Ze statistik vývoje nehodovosti uvedených v kapitole 2.3 je zřejmé, nakolik je tato problematika aktuální. Dále jsem se věnoval popisu používaných statistických pojmů. Následně jsem prezentoval několik existujících statistických metod používaných k označování kritických úseků na silnicích. Z porovnání statistických postupů vyplynulo, že metoda KDE+, a tedy i představovaná aplikace, může být k hledání nejkritičtějších silničních úseků velmi vhodná. A to především proto, že není třeba parametricky nastavovat kritickou hodnotu, od které je nalezený shluk při výpočtu považován za významný. Stanovení této mezní hodnoty je integrální součástí metody KDE+. Další výhodou prezentované metody tvoří skutečnost, že jako riziková místa jsou označeny pouze nejnebezpečnější úseky komunikací, jež tvoří na reálných sítích pasáže dosahující v součtu délky pouze malých jednotek (či i zlomků) procent celkové délky sítě. Na reálných datech o srážkách se zvěří v České republice tvoří nejnebezpečnější úseky jen 0,05 % délky ze všech zkoumaných komunikací měřících přes 37 tisíc km. Těmito kritickými úseky, dosahujícími v součtu délky cca 260 km, by bylo tedy vhodné se blíže zabývat.

V kapitole 3.2 jsem se věnoval popisu samotné metody KDE+. Detailněji jsem popsal její dílčí kroky a na příkladech jsem vysvětlil podstatu jejího fungování. Dále jsem popsal kýženu implementaci metody do podoby uživatelsky přívětivé spustitelné počítačové aplikace. V rámci vývoje aplikace bylo třeba projít fázi analýzy a návrhu, kdy jsem provedl rozdělení aplikace do několika z hlediska funkcionality oddělených celků, které z pohledu uživatele tvoří dohromady celistvou aplikaci. Dalším krokem tvorby aplikace bylo samotné programování, v rámci něhož jsem použil několika existujících knihoven. Během této fáze bylo rovněž třeba stanovit formát vstupních a výsledných souborů s daty.

Po dokončení prototypové implementace bylo třeba provést testování korektnosti vypočtených výsledků. S vědomím správné funkčnosti mohly být provedeny optimalizace rychlosti výpočtu, při kterých bylo dosaženo zkrácení doby zpracování (přibližně o 30 %). V souvislosti se zrychlováním výpočtu byl výkonný kód paralelizován, čímž se dosáhlo ještě násobného zlepšení, daného rozdělením práce mezi více procesorových jader.

Následně jsem uvedl několik možností upotřebení vzniklé aplikace. Software byl vyzkoušen na reálných datech popisujících srážky se zvěří na silniční síti České republiky.

6.1 Možnosti dalšího rozvoje aplikace

Ačkoli je vytvořená aplikace plně funkční, je z širší zkoumané oblasti zřejmé, že zde existuje možnost jejího dalšího rozvoje, a to v několika různých směrech. Reálnému nasazení aplikace by také mělo předcházet její nezávislé komplexní otestování. Sám jsem sice provedl řadu testů k zajištění validity výsledků, avšak jako autor aplikace nemohu otestovat aplikaci tak dobře, jako nezávislý odborník. A to především proto, že znám vnitřní principy jejího fungování a nenahlížím na ni, jak by bylo třeba, pouze zvenčí.

6.1.1 Podpora načítání a ukládání dvourozměrných dat

Ve své současné podobě přijímá aplikace jako svůj vstup pouze jednorozměrná data. Tato data obsahují seznam zpracovávaných silničních úseků a seznam staničení, na kterých se v minulosti udály nehody. Pro některé uživatele by mohla být příjemnější práce přímo s dvourozměrnými mapovými podklady. V takových podkladech jsou totiž silniční úseky reprezentovány liniemi a nehody body. Vhodným formátem, který se pro toto využití nabízí, je ESRI shapefile. Jedná se o formát souboru, ve kterém lze ukládat vektorová prostorová data pro zpracování v geografických informačních systémech. Mezi skupinou potenciálních uživatelů aplikace je tento formát relativně hojně rozšířen.

6.1.2 Aplikace jako služba

Pro některé uživatele může být nevýhodou nutnost instalace (nahrání) software na svůj počítač. Obzvláště to pak platí pro aplikaci vyžadované běhové prostředí jazyka Java (JRE). Dále pak uživatel nemusí disponovat dostatečně výkonným počítačem pro zpracování rozsáhlých dat. Příkladem rozsáhlých dat mohou být silniční sítě na úrovni krajů České republiky, které zahrnují historická data za delší časové období. Taková data obsahují velký počet silničních úseků s typicky velkým počtem nehod. K jejich rozumně rychlému zpracování je třeba disponovat větším výpočetním výkonem.

Z těchto důvodů by bylo možno transformovat existující desktopovou aplikaci do webové služby, běžící na dostatečně výkonném serveru mimo pracoviště uživatele. Server by mohl být umístěn ve výpočetním cloudu, bylo by tedy možno měnit jeho výkon v závislosti na aktuálních požadavcích.

Otázkou k zamyšlení zůstává, jakým způsobem by uživatel platil za použití tohoto systému, tvořeného softwarovým i hardwarovým vybavením. Ještě dalším krokem směrem k uživatelské přívětivosti by mohlo být poskytnutí komplexní služby analýzy dat, od počá-

tečního předzpracování po konečnou prezentaci výsledků, včetně uživatelské podpory. Taková služba by již jistě nemohla být plně automatizována, ale v jistých fázích analýzy by bylo třeba netriviální součinnosti experta na danou problematiku.

6.1.3 Implementace více statistických metod

Do aplikace by bylo možno implementovat další existující statistické metody (např. metody zmíněné v kapitole 3.3) a do nastavení vstupních parametrů přidat volbu požadované statistické metody. V souvislosti s tím by se dále vhodně rozšířil modul prezentace výsledků výpočtu tak, aby byl uživatel schopen provést porovnání výsledků získaných z více výpočetních běhů zpracovávaných pomocí různých metod. Pomocí software by pak bylo možno provádět komparaci různých statistických metod. Této funkčnosti by s prospěchem šlo využít pro úpravu existujících metod, či pro vývoj metod zcela nových. Klíčovou vlastností je zde vhodný objektový návrh aplikace, který umožňuje snadné přidání dalších algoritmů.

Použitá literatura

- [1] **POPELKA, Petr.** *Zrod moderní dopravy: modernizace dopravní infrastruktury v Rakouském Slezsku do vypuknutí první světové války.* Vyd. 1. Ostrava : Filozofická fakulta Ostravské univerzity, 2013. 311 s. ISBN 978-80-7464-384-2.
- [2] **Ministerstvo dopravy ČR.** *Statistiky vyplývající z Centrálního registru vozidel vedené na Ministerstvu dopravy* [Online]. 2015. [Citace: 29. 12 2015].
<http://www.mdcr.cz/cs/Silnicni_doprava/Dovoz_registrace_a_schvalovani_vozidel/Registrace+vozidel/Statistiky+vyplývající+z+Centr%C3%A1ln%C3%ADho+registru+vozidel>.
- [3] **European Commission.** *Road Safety: new statistics call for fresh efforts to save lives on EU roads* [Online]. 31. 3. 2016. [Citace: 29. 4 2016].
<http://europa.eu/rapid/press-release_IP-16-863_en.htm>.
- [4] **JENELIUS, Erik a MATTSSON, Lars-Göran.** *Road network vulnerability analysis: Conceptualization, implementation and application* [Online]. 2014. [Citace: 29. 4 2016]. <<http://dx.doi.org/10.1016/j.compenvurbsys.2014.02.003>>.
- [5] **JANÍČEK, Přemysl.** *Systémová metodologie: brána do řešení problémů.* Vyd. 1. Brno : CERM, 2014. 365 s. ISBN 978-80-7204-887-8.
- [6] **FRIČ, Jindřich; HAVRÁNEK, Pavel; LIPL, Martin; NOVÁK, Jan; POKORNÝ, Petr; SIMONOVÁ, Eva; Veronika, VALENTOVÁ; Lucie, VYSKOČILOVÁ.** *Řešení kritických míst na pozemních komunikacích v extravilánu: metodika provádění.* Vyd. 2. Brno : Centrum dopravního výzkumu, v. v. i., 2015. ISBN 978-80-88074-22-9.
- [7] **BÍL, Michal; ANDRÁŠIK, Richard; JANOŠKA, Zbyněk; SEDONÍK, Jiří; Valentová, Veronika.** *Identifikace kritických nehodových lokalit pomocí GIS analýzy polohy dopravních nehod.* Vyd. 1. Brno : Centrum dopravního výzkumu, 2014. ISBN 978-80-86502-76-2.
- [8] **BÍL, Michal; VODÁK, Rostislav; KUBEČEK, Jan; SVOBODA, Tomáš; ANDRÁŠIK, Richard; SEDONÍK, Jiří; KŘIVÁNKOVÁ, Zuzana, BÍLOVÁ, Martina.** *Online identifikace kritických úseků pozemních komunikací pomocí síťových analýz.* Vyd. 1. Brno : Centrum Dopravního výzkumu, v. v. i., 2015. ISBN: 978-80-88074-18-2.

- [9] **MAFRA, a. s.** *Inspekce práce vyměřila první pokuty za tragický pád mostu ve Vilemově* [Online]. 2016. [Citace: 29. 4 2016].
<http://jihlava.idnes.cz/inspektorat-prace-vymeril-prvni-pokuty-za-pad-mostu-ve-vilemove-phk-/jihlava-zpravy.aspx?c=A160119_162058_jihlava-zpravy_mv>.
- [10] **BÍL, Michal; KUBEČEK, Jan; ANDRÁŠIK, Richard.** Total water content thresholds for shallow landslides. *Landslides*. 2016, Volume 13, Issue 2. s. 337–347. ISSN 1312-510X.
- [11] *Rozhodnutí rady ze dne 30. listopadu 1993 o vytvoření databáze Společenství týkající se dopravních nehod v provozu na pozemních komunikacích : 93/704/ES* [Online]. 30. 11. 1993. [Citace: 29. 4 2016].
<http://publications.europa.eu/resource/cellar/5701bed1-5e44-45b5-aa9e-c28ee39d9215.0001.01/DOC_23>.
- [12] *Resolution of the Council and of the representatives of the Governments of the Member States, meeting within the Council of 21 June 1991 on a Community programme of action on road safety* [Online]. 1991. [Citace: 29. 4. 2016].
<<http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:41991X0709&from=EN>>
- [13] *Statistics - accidents data* [Online]. 2016. [Citace: 29. 4. 2016].
<http://ec.europa.eu/transport/road_safety/specialist/statistics/index_en.htm>.
- [14] **PŘIBYL, Pavel; JANOTA, Aleš; SPALEK, Juraj.** *Analýza a řízení rizik v dopravě: tunely na pozemních komunikacích a železnicích*. Vyd. 1. Praha : BEN - technická literatura, 2008. 527 s. ISBN 978-80-7300-214-5.
- [15] **CHMELÍK, Jan.** *Dopravní nehody*. Vyd. 1. Plzeň : Vydavatelství a nakladatelství Aleš Čeněk, 2009. 540 s. ISBN 978-80-7380-211-0.
- [16] **BÍL, Michal, ANDRÁŠIK, Richard a JANOŠKA, Zbyněk.** Identification of hazardous road locations of traffic accidents by means of kernel density estimation and cluster significance evaluation. *Accident analysis & Prevention*. June 2013, Volume 55. 265-273 s. ISSN 0001-4575.
- [17] **TUČKA, Pavel.** *Lokalizace dopravních nehod s využitím GPS* [Online]. 2015. [Citace: 2. 5. 2016].
<<http://docplayer.cz/1669662-Lokalizace-dopravnich-nehod-s-vyuzitim-gps-ing-pavel-tucka-pavel-tucka-cdv-cz.html>>.

- [18] **ANDRES, Josef; MIKULÍK, Josef; ROKYTOVÁ, Jitka; HRUBÝ, Zdeněk; SKLÁDANÝ, Pavel.** *Metodika identifikace a řešení míst častých dopravních nehod.* Vyd. 1. Brno : Centrum dopravního výzkumu, v. v. i., 2001. 38 s. ISBN 80-902141-9-3.
- [19] **STRIEGLER, Radim; VALENTOVÁ, Veronika; POKORNÝ, Petr; AMBROS, Jiří; ŠENK, Petr; JANOŠKA, Zbyněk.** *Identifikace kritických míst na pozemních komunikacích v extravilánu – metodika provádění.* Vyd. 1. Brno : Centrum dopravního výzkumu, v. v. i. 2013., 194 s. ISBN 978-80-86502-70-0.
- [20] **VYSKOČILOVÁ, Alena.** *Metodika výpočtu ztrát z dopravní nehodovosti na pozemních komunikacích: aktualizovaná verze 2013.* Vyd. 1. Brno : Centrum dopravního výzkumu, v. v. i., 2013. 35 s. ISBN 978-80-86502-55-7.
- [21] **ZENDULKA, Jaroslav; BARTÍK, Vladimír; LUKÁŠ, Roman; RUDOLFOVÁ, Ivana.** *Získávání znalostí z databází - Studijní opora.* Vyd. 1. Brno : Vysoké učení technické v Brně, Fakulta informačních technologií, 2006. 160 s.
- [22] **KARPÍŠEK, Zdeněk.** *Statistická analýza : přehledový učební text pro doktorské studium* [Online]. 2008. [Citace: 2. 5. 2016].
<http://www.mat.fme.vutbr.cz/download.aspx?id_file=2033>.
- [23] **BÍL, Michal; ANDRÁŠIK, Richard; SVOBODA, Tomáš; SEDONÍK, Jiří.** The KDE+ software: a tool for effective identification and ranking of animal-vehicle collision hotspots along networks. *Landscape Ecology*. February 2016, Volume 31, Issue 2, s. 231–237. ISSN 0921-2973.
- [24] **HORÁK, Jiří.** *K funkce. Prostorová analýza dat* [Online]. 2002. [Citace: 4. 5. 2016]. <http://gis.vsb.cz/pad/Kap_3/kap_3_2_3.htm>.
- [25] **YAMADA, Ikuho; THILL, Jean-Claude.** Comparison of planar and network K-functions in traffic accident analysis. *Journal of Transport Geography*. June 2004, Volume 12, Issue 2. s. 149–158. ISSN 0966-6923.
- [26] **KELBEL, Jan; ŠILHÁN, David.** *Shluková analýza* [Online]. 2004. [Citace: 25. 4. 2016].
<http://cmp.felk.cvut.cz/cmp/courses/recognition/zapis_prednasky/zapis_02/13/shlukovani.pdf>.
- [27] **ZEMAN, David.** *Shlukování* [Online]. 2005. [Citace: 4. 5. 2016].
<www.fit.vutbr.cz/study/courses/ZZD/public/seminar0405/zemand>.
- [28] **OKABE, Atsuyuki; SUGIHARA, Kokichi.** *Spatial analysis along networks: statistical and computational methods.* Chichester : John Wiley & Sons, 2012. 306 s. ISBN 978-0-470-77081-8.

- [29] **STEENBERGHEN, Thérèse; AERTS, Koen; THOMAS, Isabelle.** Spatial clustering of events on a network. *Journal of Transport Geography*. 2010, Volume 18, Issue 3. s. 411–418. ISSN 0966-6923.
- [30] **SHIODE, Shino a SHIODE, Narushige.** Detection of multi-scale clusters in network space. *International Journal of Geographical Information Science*. 2009, Volume 23, Issue 1. s. 75–92. ISSN 1365-8816.
- [31] **SCHILDT, Herbert.** *Mistrovství - Java*. 1. vyd. Brno : Computer Press, 2014. 1224 s. ISBN 978-80-251-4145-8.
- [32] **GILBERT, David.** *JFreeChart* [Online]. 2014. [Citace: 10. 10. 2015]. <<http://www.jfree.org/jfreechart/>>.
- [33] **The Apache Software Foundation.** *Commons IO* [Online]. 2015. [Citace: 5. 11. 2015]. <<https://commons.apache.org/>>.
- [34] **Oracle Corporation and/or its affiliates.** *A Brief History of NetBeans* [Online]. NetBeans. 2011. [Citace: 5. 5. 2016]. <<https://netbeans.org/about/history.html>>.
- [35] **KOWAL, Grzegorz.** *Cross-platform Java executable wrapper* [Online]. 2015. [Citace: 10. 10 2015]. <<http://launch4j.sourceforge.net/>>.
- [36] **Object Management Group, Inc.** *OMG Unified Modeling Language* [Online]. 2015. [Citace: 15. 3. 2016]. <<http://www.omg.org/spec/UML>>.
- [37] **Centrum dopravního výzkumu, v. v. i.** *Kde bouráme* [Online]. 2015. [Citace: 20. 4. 2016]. <<http://kdebourame.cz>>.
- [38] **Centrum dopravního výzkumu, v. v. i.** *Srážky se zvěří*. [Online]. 2016. [Citace: 15. 4. 2016]. <<http://www.srazenazver.cz>>.

Použité zkratky

API – Application programming interface (Rozhraní aplikace – funkcionalita, kterou lze volat a použít z jiných aplikací.)

CDV – Centrum dopravního výzkumu, v. v. i.

CSV – Comma-separated values (Soubor s hodnotami oddělenými vybraným oddělovačem, typicky čárkou či středníkem.)

DI – Dangerousness index (Index nebezpečí)

ESRI - Environmental Systems Research Institute (Společnost vyvíjející software pro práci s geografickými informačními systémy)

Exe – Executable file (Spustitelný soubor)

GIS - Geographic information system (Geografický informační systém)

GPS – Global Positioning System (Globální polohový systém)

Jar – Java Archive (Kompresní formát pro jazyk Java)

JRE – Java Runtime Environment (Běžové prostředí pro aplikace v jazyce Java)

JVM – Java Virtual Machine (Virtuální běžové prostředí pro jazyk Java)

KDE – Kernel Density Estimation (Jádrový odhad hustoty)

KDE+ – Kernel Density Estimation plus (Jádrový odhad hustoty plus)

UML – Unified Modeling Language (Standardizovaný jazyk pro modelování software)

Seznam obrázků

Obrázek 2.1 – Dopravní nehoda – překlopení tahače s návěsem	16
Obrázek 2.2 – Kolizní diagram	24
Obrázek 3.1 – Průběh Epanechnikova jádra	28
Obrázek 3.2 – Sčítání jádrové funkce	29
Obrázek 3.3 – Stanovená hladina významnosti	30
Obrázek 3.4 – Rozdíl mezi planárním a síťovým měřením vzdálenosti.....	32
Obrázek 3.5 – Dendrogram pěti objektů.....	33
Obrázek 3.6 – Objekty v okolí bodu měření.....	34
Obrázek 3.7 – Clumping metoda	35
Obrázek 4.1 – Diagram tříd dle UML.....	39
Obrázek 4.2 – Vstupní formát souboru.....	41
Obrázek 4.3 – Formát výstupního souboru.....	41
Obrázek 4.4 – Vytvoření diskrétních bodů hustoty	42
Obrázek 4.5 – Vzorec pro výpočet Epanechnikova jádra.....	43
Obrázek 4.6 – Hlavní okno aplikace.....	45
Obrázek 4.7 – Hlavní okno aplikace po dokončení výpočtu	46
Obrázek 4.8 – Výpis seznamu zpracovaných úseků.....	46
Obrázek 4.9 – Vizualizace zpracování jednoho úseku	47
Obrázek 4.10 – Rozdílné výsledky dvou běhů výpočtu	48
Obrázek 5.1 – Prostorová reprezentace úseků a nehod	52
Obrázek 5.2 – Průběh úseku s nehodami	53
Obrázek 5.3 – Obsah balíčku s aplikací.....	53
Obrázek 5.4 – Ověření přítomnosti a verze JRE	54
Obrázek 5.5 – Nalezené shluky dopravních nehod.....	55
Obrázek 5.6 – Výřez mapy z aplikace www.kdebourame.cz	58
Obrázek 5.7 – Výřez mapy z aplikace www.srazenazver.cz	59

Seznam tabulek

Tabulka 3.1 – Porovnání metod nalézající shluky	31
Tabulka 4.1 – Vstupní hodnoty.....	40
Tabulka 4.2 – Výstupní hodnoty.....	42
Tabulka 4.3 – Porovnání rychlosti výpočtu	49
Tabulka 5.1 – Aplikace metody KDE+ na silniční síť České republiky	56

Seznam grafů

Graf 2.1 – Vývoj počtu nehod v ČR a EU	18
Graf 2.2 – Vývoj počtu obětí smrtelných nehod v ČR a EU	18
Graf 2.3 – Úmrtnost na silnicích EU v roce 2014	19
Graf 4.1 – Závislost velikosti vstupních dat na rychlosti výpočtu.....	50

Přílohy

Příloha A – Seznam elektronických příloh

V následujícím seznamu uvádím přehled důležitých souborů obsažených v elektronické příloze práce se stručným shrnutím jimi poskytované funkcionality.

Zdrojové kódy:

- `src/ClusteringApp.java` – třída obsahující metodu `main()`, která tvoří vstupní bod aplikace
- `src/TestWorker.java` – třída odpovědná za spuštění výpočtu a monitorování jeho stavu
- `src/DataLoader.java` – načítání a kontrola vstupních dat o úsecích a nehodách
- `src/Algorithm.java` – inicializace paralelního výpočtu
- `src/AlgRunnable.java` – implementace metody `KDE+`
- `src/ResultWriter.java` – zápis výsledků do výstupních souborů

Další soubory:

- `config.properties` – soubor s konfigurací výpočtu, je použit pro stanovení implicitních hodnot pro zobrazení v GUI, nebo jsou z něho načteny vstupní parametry v případě dávkového spuštění
- `KDEplus.exe` – spustitelný soubor
- `KDEplus.jar` – spustitelný soubor
- `*.csv` – vstupní soubory s testovacími daty
- `readme.txt` – soubor s pokyny pro spuštění a základní nápovědou
- `manual.pdf` – soubor s kompletní nápovědou

Příloha B – Příručka pro práci s aplikací

Následující text uživatelské příručky, která doprovází aplikaci, byl sepsán týmem zaměstnanců Centra dopravního výzkumu, v. v. i., jehož je autor práce členem, zabývajícím se vývojem metody KDE+ a s ní souvisejícím software. Text je v příloze uveden pro případného zájemce, který by chtěl získat rychlý přehled o použití v práci popisované aplikace.

Účel aplikace

Aplikace KDE+ vznikla za účelem prezentace nové metody KDE+, která je rozšířením standardní KDE (Kernel Density Estimation) o metody objektivního testování významu shluků. Metoda je popsána na www.kdeplus.cz a v dalších odkazech tamtéž.

Vstupní soubory - Údaje o úseku

Tento soubor obsahuje pouze identifikaci daného úseku a jeho délku v metrech. Formát (bez záhlaví, záznamy oddělené středníkem a bez mezer) např.:

125;258

126;2789

127;698

Vstupní soubory - Údaje o poloze bodů na úsecích

Soubor obsahuje identifikaci úseku a staničení bodů v metrech. Hodnota staničení nesmí být vyšší, než je délka úseku v doprovodném souboru, např.:

125;1

125;15.5

125;159

126;10.5

126;11

126;11

Velikost okna

K dispozici jsou dvě hodnoty, které se liší mírou nejistoty při určení polohy bodu na linii, a ostatních jevů, které mohou bod ovlivnit, nebo naopak.

Přesnost dat

Pokud jsou polohy bodů na úsecích zaměřeny přesně, zvolíme možnost GPS. Jestliže se body na úsecích lokalizovaly v systému staničení s přesně daným krokem, obvykle 100 j, potom zvolíme možnost staničení. Aplikace podle tohoto výběru mění tvar jádra funkce jádrového odhadu hustoty.

Spust'

Spustí se program. V okně ve spodní části lze sledovat průběh výpočtu, stejně jako celkový čas, který zbývá na progress bar vedle tohoto tlačítka.

Zastav

Pokud výpočet probíhá pomalu, např. z důvodů zadání příliš rozsáhlých vstupních souborů, lze tímto výpočet přerušit.

Tabulka výsledků

Jedná se o shrnutí výsledků výpočtu. Každý řádek odpovídá jednomu úseku. V prvním sloupci jsou ID úseků, v dalším délka, ve třetím počet záznamů (nehod) a v posledním počet identifikovaných shluků. Po poklepání (dvojklik) na řádek se otevře okno s detaily pro daný úsek. Je vykreslena funkce jádrového odhadu hustoty a hladina významnosti. Shluky jsou definovány tam, kde funkce JOH převyší tuto hladinu. Grafika v horní části je pro názornost, níže jsou detaily. Výsledky, pokud existují, tj. pokud se vypočítal alespoň jeden shluk na jednom úseku, se uloží do souboru results.csv ve formátu:

ID úseku; počátek shluku v j; konec shluku v j; počet bodů ve shluku; síla shluku

Hlášení chyb

Pokud jste našli v aplikaci chybu, kontaktujte nás prosím na gis@cdv.cz

Jak citovat

Pokud program využijete při své práci, potom ho prosím citujte:

Bíl, M., Andrášik, R., Svoboda, T., Sedoník, J., 2016. *The KDE+ Software: A Tool for Effective Identification of Animal-Vehicle Collision Hotspots Along Networks*. *Landscape Ecology* 31, 231–237.

Pro výzkumné práce můžete také použít původní publikace, kde byla metoda KDE+ poprvé představena:

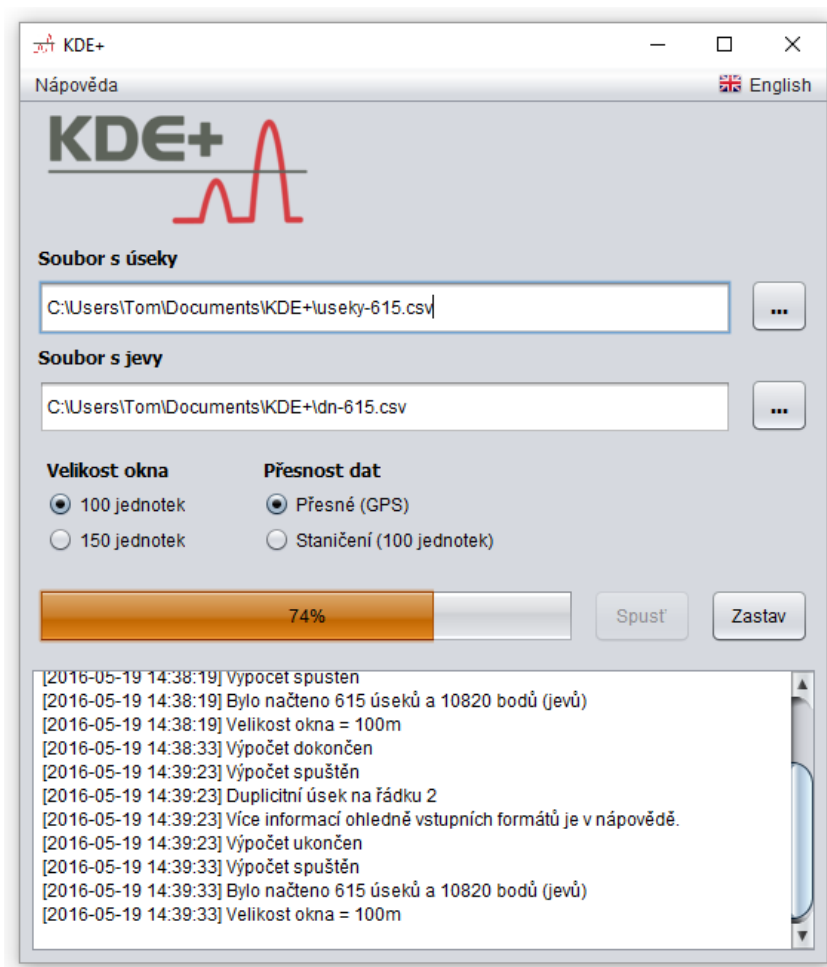
Bíl, M., Andrášik, R., Janoška, Z., 2013. Identification of Hazardous Road Locations of Traffic Accidents by means of Kernel Density Estimation and Cluster Significance Evaluation. *Accident Analysis and Prevention* 55, 265–273.

Autoři

Aplikace KDE+ vytvořilo Centrum dopravního výzkumu, v. v. i., oblast Geoinformatiky. Pokud shledáte aplikaci užitečnou, budeme rádi, když nám dáte vědět.

Příloha C – Snímky obrazovek vytvořené aplikace

Hlavní obrazovka:



Tabulka s výsledky – zpracovanými úseky:

Úseky	Délka	Bodů (jevů)	Shluky	Details
3789	2 935	20	1	
3790	131	0	0	
3791	96	0	0	
3792	2 860	7	1	
3951	154	0	0	
4129	785	29	1	
4195	5 057	23	4	
4196	129	0	0	
4212	123	0	0	
4263	93	0	0	
4264	1 257	18	1	
4265	1 542	19	0	
4266	5 602	78	2	
4267	5 271	76	4	
4338	4 666	16	3	
4339	3 932	17	4	
5347	373	15	0	
5359	2 152	19	1	
5749	7 562	57	6	
6073	187	0	0	
6095	4 685	1	0	
6102	437	0	0	
6242	857	33	1	
6558	1 066	46	1	
6630	173	0	0	
6631	8 216	29	7	
6755	80	0	0	
6769	7 513	88	6	
6770	87	0	0	

Detailní pohled na výsledek analýzy jednoho úseku:

