

SVEUČILIŠTE U ZAGREBU  
GEODETSKI FAKULTET

Tomislav Šimunović i Filip Todi

Geostatistička analiza prostorne distribucije prometnih nesreća na području Grada Zagreba u razdoblju od 2010. do 2013. godine

Zagreb, 2014.

Ovaj rad izrađen je na Geodetskom fakultetu, Zavodu za geomatiku, Katedri za geoinformatiku pod vodstvom prof. dr. sc. Damira Medaka, dipl. ing. geod. i predan je na natječaj za dodjelu Rektorove nagrade u akademskoj godini 2013./2014.

## Sadržaj rada:

|        |                                                                                 |    |
|--------|---------------------------------------------------------------------------------|----|
| 1.     | Uvod .....                                                                      | 1  |
| 1.1.   | Prethodna istraživanja.....                                                     | 4  |
| 2.     | Op i i specifi ni ciljevi rada.....                                             | 9  |
| 3.     | Materijal i metode .....                                                        | 10 |
| 3.1.   | Preuzimanje prometne mreže Grada Zagreba s karte svijeta OpenStreetMap .....    | 10 |
| 3.2.   | Obrada podataka iz policijskih zapisnika pomo u programskog jezika Python ..... | 11 |
| 3.3.   | Unos i obrada podataka pomo u PostgreSQL-a i PostGIS-a .....                    | 16 |
| 3.4.   | Analiza podataka pomo u hi-kvadrat testa.....                                   | 18 |
| 3.5.   | Analiza podataka metodom normalizacije .....                                    | 20 |
| 3.6.   | Analiza podataka statisti kom metodom Kernel Density Estimation .....           | 24 |
| 3.6.1. | Kernel Density Estimation (KDE) .....                                           | 26 |
| 3.6.2. | Odre ivanje širine pojasa $h$ .....                                             | 36 |
| 3.6.3. | KDE u 2D i 3D.....                                                              | 37 |
| 3.6.4. | Primjene KDE-a .....                                                            | 42 |
| 3.7.   | Vizualizacija rezultata .....                                                   | 45 |
| 4.     | Rezultati .....                                                                 | 46 |
| 4.1.   | Osnovni podaci za obradu .....                                                  | 46 |
| 4.2.   | Rezultati obrade hi-kvadrat testovima.....                                      | 51 |
| 4.3.   | Rezultati obrade statisti kom metodom normalizacije.....                        | 56 |
| 4.4.   | Rezultati obrade statisti kom metodom KDE .....                                 | 61 |
| 4.4.1. | Prometne nesre e s poginulim sudionicima .....                                  | 61 |
| 4.4.2. | Prometne nesre e s poginulim i ozlije enim sudionicima .....                    | 64 |
| 4.4.3. | Prometne nesre e s poginulim i ozlije enim pješacima.....                       | 72 |
| 4.4.4. | Prometne nesre e s poginulim i ozlije enim biciklistima.....                    | 76 |
| 4.5.   | Dodatne mogu nosti vizualizacije rezultata .....                                | 80 |

|                                                              |    |
|--------------------------------------------------------------|----|
| 4.6. Problemi uo eni tijekom obrade i analize podataka ..... | 81 |
| 5. Rasprava .....                                            | 84 |
| 6. Zaklju ci.....                                            | 88 |
| Zahvale .....                                                | 89 |
| Popis literature.....                                        | 90 |
| Sažetak .....                                                | 94 |
| Summary .....                                                | 95 |

## 1. Uvod

Prema Svjetskoj zdravstvenoj organizaciji (*eng. World Health Organization - WHO*), godišnje u svijetu 1.24 milijuna ljudi pogine na cestama, od čega su 77% muškarci. U 2010. godini, zabilježeno je da visoko razvijene zemlje imaju najmanju stopu smrtnosti na cestama, slabo razvijene imaju višu stopu, a srednje razvijene zemlje imaju najveću u stopu smrtnosti. Na svjetskim cestama 50% poginulih su posebno ranjivi sudionici u prometu (23% motociklisti, 22% pješaci, 5% biciklisti). Ozljeđe dobivene u prometnim nesrećama su osmi vodeći uzrok smrti i prvi uzrok smrti među osobama od 15. do 29. godine, što predstavlja važan globalni zdravstveni problem (URL 1).

Od 194 zemlje koje su sadržane u statistici, 64% ih ima ograničenu brzinu u urbanim područjima od 50 km/h, ali samo 59 država dopušta lokalnim vlastima smanjenje dopuštene granice ovisno o potrebama; 89 država ima zakone o vožnji u alkoholiziranom stanju; 111 država ima zakone koji propisuju korištenje zaštitnih pojaseva; 90 država ima zakone koji propisuju korištenje kaciga i 96 država ima zakone koji propisuju korištenje dodatne zaštite za djecu (URL 1).

Od 2007. do 2010. godine zabilježen je porast u broju registriranih vozila u svijetu u iznosu od 15%. U srednje razvijenim zemljama, dogodio se nagli porast broja motornih vozila. Loša strana tog procesa je što se pri razvoju prometne infrastrukture ne uzima u obzir sigurnost ostalih sudionika u prometu, odnosno pješaka i biciklista (URL 1).

Broj prometnih nesreća je u porastu na globalnoj razini iz razloga što prometna infrastruktura ne prati razvoj ostalih sektora poput industrije i razvoja stambenih naselja. Prometne nesreće su među glavnim uzrocima smrti na globalnoj razini, a većina nesreća je rezultat ljudskih pogrešaka, najčešće nemara vozača ili pješaka. Vjerojatnost prometnih nesreća i njihova težina mogu biti smanjeni sistematskom analizom scenarija nesreća te korištenjem odgovarajućih rješenja koja uključuju primjenu potrebnih uređaja za kontrolu prometa, prikladnim dizajnom prometnica i u inkovitim intervencijama prometne policije. Donošenje u inkovitim rješenjima iziskuje analizu prostornih i vremenskih obrazaca na mjestima prometnih nesreća što se postiže primjenom geoprostornih tehnologija (Prasannakumar et al., 2011).

Heterogenost vozačke populacije, obilježja prometnica, vremenskih i prometnih prilika rezultira heterogenošću u prometnim nesrećama. Zbog želje i potrebe za sigurnijim prometnim okruženjem, stručnjaci su zaduženi za identifikaciju visokoriziknih lokacija s ciljem

popravljanja istih. Problem nastaje kada nije moguće pouzdano razlikovati mjesta koja su zaista visokorizična od mjesta koja slučajno imaju zabilježen veći broj prometnih nesreća u promatranom razdoblju. Zbog oscilacije broja prometnih nesreća iz godine u godinu, metode koje se oslanjaju na jednostavno praćenje uestalosti nesreća povećavaju rizik od velikog broja tzv. 'lažnih pozitivnih nesreća'. To vodi do mogućeg poboljšanja sigurnosnih uvjeta na relativno sigurnim lokacijama, a otvara mogućnost previda zaista opasnih lokacija što u konačnici rezultira lošim ulaganjima u unaprjeđenje prometne infrastrukture (Cheng i Washington, 2005).

Iz tog je razloga ključna identifikacija žarišta prometnih nesreća. Cilj je razvoj i provođenje učinkovitih sigurnosnih strategija na područjima s velikom gustoćom prometnih nesreća. U tom su procesu važni atributni podaci pridruženi prometnim nesrećama. Oni omogućuju bolje razumijevanje procesa koji se odvijaju na žarištima što vodi kvalitetnijoj raspodjeli resursa namijenjenih poboljšanju sigurnosti u prometu. Literatura o prometnim nesrećama ne pruža univerzalno prihvaćenu definiciju žarišta prometnih nesreća, što prisiljava stručnjake da ih rangiraju na osnovu težine ili uestalosti nesreća. Analiza žarišta tradicionalno se koncentrirala na cestovne segmente ili na specifična križanja te zahtijeva opsežno razumijevanje stupnja uključenosti vozila u proces, težinu ozljeda i okolinu. GIS (*eng. Geographic information system*) aplikacije su praktične za ovu vrstu problema jer omogućuju učinkovitu manipulaciju, analize i vizualizaciju prostornih podataka. Identifikacijom žarišta prometnih nesreća u GIS-u može se ostići potpunije razumijevanje uzroka nesreća. Žarišta prometnih nesreća otkrivaju koncentracije istovrsnih nesreća i ukazuju na prostornu ovisnost između individualnih slučajeva koji mogu imati zajedničke uzroke (Anderson, 2009).

GIS tehnologija je popularan alat za vizualizaciju podataka prometnih nesreća i analizu žarišta te važan upravljački alat za sigurnost u prometu. Omogućava manipulaciju velikim količinama podataka koja se može pohraniti, dijeliti i obrađivati. Pruža platformu za analizu i vizualizaciju podataka s ciljem istraživanja odnosa između istih te može pružiti grafičke i ne-grafičke rezultate. Važnost analize prostornih podataka leži u atributima i njihovim vezama, uzimajući pritom u obzir prostornu lokaciju promatrane pojave. Informacijski sustavi prometnih nesreća zasnovani na GIS-u mogu otkriti veze između prostornih pojava koje su skoro neodredive s ne-prostornom bazom podataka. Prostorni podaci dobiveni GIS-om i prostornim analizama su jedni od najvažnijih izvora informacija pri analizi prometnih nesreća jer pružaju stručnjacima mnoštvo informacija o opasnim lokacijama, žarištima, itd. Stručnjaci pomoću GIS-a mogu spojiti i geokodirati podatke nesreća i prometnica, izraziti u ustalost

nesre a, i sl. S pravom organizacijom GIS-a, prometne institucije mogu analizirati podatke prometnih nesre a i prostorne karakteristike tih podataka (Erdogan et al., 2008).

Identifikacija opasnih lokacija prostornim analizama pomo u GIS-a doprinosi smanjenju broja prometnih nesre a, ali uspjeh ovih analiza uvelike ovisi o to nosti, pouzdanosti i opsežnosti zapisnika prometne nesre e, posebno pri unosu i obradi podataka. Analiza prometnih nesre a je jako složena tema zbog brojnih faktora koji prostorno utje u na nesre e i stoga zapisnici moraju biti detaljni i prikladno strukturirani za prostornu i statisti ku analizu. Umjesto klasi nog analognog zapisnika, GNSS (*eng. Global Navigation Satellite System*) i GIS moraju se integrirati u novi sustav zapisnika ime se uklanjaju mnoge sistematske i slu ajne pogreške te u isto vrijeme pove ava prostorna to nost. Uklju ivanje multimedijских mogu nosti (digitalne slike i video zapisi) dodatno poboljšava u inkovitost sustava (Erdogan et al., 2008).

S prostornog gledišta, u analizi prometne sigurnosti cestovna se mreža esto dijeli na individualna mjesta poput raskrižja, homogenih segmenata ili koridora. Statisti ki modeli se razvijaju kako bi se odredila frekvencija prometnih nesre a na odre enim mjestima. Rezultati modela omogu uju prometnim institucijama lociranje mjesta s u ustalim prometnim nesre ama u cestovnoj mreži i uo avanje zna ajnih faktora koji doprinose nesre ama. Cilj ovog procesa je poboljšanje prometne sigurnosti pomo u inženjerskih i zakonskih mjera te dodatnog obrazovanja. Razvoj modela zahtijeva više tipova podataka o prometnoj nesre i. Naj eš e je rije o vrsti nesre e i lokaciji na kojoj se dogodila. Razli ite vrste sirovih podataka moraju se vremenski i prostorno povezati sa svakim mjestom kako bi se dobili ulazni podaci za ovisne varijable, primjerice broj prometnih nesre a odre enog tipa na nekoj lokaciji u odre enom vremenskom razdoblju te opisne varijable, primjerice dužina segmenta, gusto a prometa ili vremenske prilike. To nost prostornih informacija o nesre ama i to nost prometne mreže je klju na za proces povezivanja podataka, stoga pogreške i nekonzistentnosti u zapisnicima prometnih nesre a predstavljaju realan problem koji esto umanjuje to nost rezultata u obliku sigurnosnog modela (Tegge i Ouyang, 2009).

Podatke vezane za prometne nesre e rijetko kad prikuplja jedna ustanova ve su razli iti tipovi podataka prikupljeni i održavani od strane razli itih agencija i ustanova. Primjerice, informacije o prometnim nesre ama prikuplja Ministarstvo unutarnjih poslova (MUP), podatke o prometnoj mreži Državna geodetska uprava (DGU) ili privatne tvrtke, a informacije o vremenskim uvjetima Državni hidrometeorološki zavod (DHMZ) i sli ne ustanove.

Pogreške i nekonzistentnosti u prostoru, odnosno lokacijama, mogu se dogoditi iz više razloga. S jedne strane, podaci o lokaciji prometne nesre e se nerijetko izvla e iz analognih policijskih zapisnika. To nost ove metode uvelike ovisi o iskustvu policijskog službenika i osoblja za pripremu podataka. Otežavaju a okolnost je što podaci o prometnim nesre ama i o prometnoj mreži mogu biti razvijani na razli itim softverskim platformama koje koriste razli ite prostorne koordinatne sustave. Tako er može do i do pogrešaka prilikom spajanja podataka ukoliko njihove prostorne koordinate nisu usporedive, a stanje se dodatno pogoršava ako su prostorni koordinatni sustavi s vremenom doživjeli nezavisne promjene (ažuriranja i redefinicije). U praksi su mnoge nesre e locirane na granici dvaju ili više segmenata, ne iz razloga što su se dogodile to no na toj lokaciji, ve zato što te lokacije policijski službenici lakše identificiraju. Ukoliko su položaj prometne nesre e ili prometnice podložni pogreškama, potreban je postupak ispravljanja prostornih koordinata. Pritom treba imati na umu da ne postoji procedura za sistematsko i u inkovito ispravljanje mogu ih prostornih pogrešaka u podacima prometnih nesre a i prometnica (Tegge i Ouyang, 2009).

### 1.1. Prethodna istraživanja

Istraživanje provedeno u Hong Kongu (Loo, 2006) predlaže metodologiju potvr ivanja i identifikacije preciznih lokacija prometnih nesre a s cestovnom mrežom bez korištenja zone obuhvata (eng. *buffer*). Autorica tvrdi da vrijednosti rezultata modela ovise o dostupnosti i kvaliteti podataka prometnih nesre a. Prije analize podataka, potrebno je ispraviti krivo georeferencirane varijable. U istraživanju su analizirane prometne nesre e Hong Konga iz razdoblja od 1993. do 2004. godine. Podaci prometnih nesre a su preuzeti iz policijske baze podataka prometnih nesre a (eng. *Traffic Accident Database System - TRADS*), a digitalna cestovna mreža je preuzeta od Odjela za katastarske poslove Hong Konga. Budu i da je jedan od glavnih ciljeva studije priprema podataka za prostorne analize, zona obuhvata oko ceste nije korištena jer dovodi do lažnih podataka. To je posebno izraženo na podru jima s velikom gusto om prometnica. Sa sustavom validacije prostornih podataka zasnovanom na GIS-u 65-80% policijskih zapisnika prometnih nesre a su imali to ne attribute ulica i okruga.

U Riyadhu, glavnom gradu Saudijske Arabije, provedeno je istraživanje (Al-Ghamdi, 2003) karakteristika prometnih nesre a na križanjima i ravnim cestovnim potezima. Korišteni su policijski zapisnici prometnih nesre a kako bi se odredile generalne karakteristike prometnih nesre a na odre enim lokacijama. Skup podataka je podijeljen na teške nesre e i nesre e s materijalnom štetom. Autor zaklju uje da je naj eš i uzrok teških nesre a na ravnim



cestovnim segmentima udarac u fiksni objekt, a 50% teških nesre a uklju uje pješake. Neprimjereni ponašanje voza a je uzrok ve ine prometnih nesre a. To uklju uje preveliku brzinu, oduzimanje prednosti i prolazak kroz crveno svijetlo.

U Ujedinjenom Kraljevstvu provedena su tri istraživanja [(Noland i Quddus, 2004), (Wang et al., 2009), (Quddus, 2008)] koja analiziraju utjecaj prometne infrastrukture i demografskih karakteristika prostornih jedinica na prometne nesre e. Podaci su preuzeti iz razli itih institucija. Prostorni podaci su preuzeti iz EDINA UKBORDERS baze podatka i Ordnance Surveya, demografski podaci iz popisa stanovništva Ujedinjenog Kraljevstva, a podaci o prometnim nesre ama iz STATS19 baze podataka prometnih nesre a. Jedan od klju nih elemenata istraživanja je integracija više baza podataka u opsežan GIS što je omogu eno geokodiranjem lokacija prometnih nesre a. Rezultati prvog istraživanja koje su proveli Noland i Quddus (Noland i Quddus, 2004) su pokazali da urbanizirana i gusto naseljena podru ja imaju manji broj prometnih nesre a. Autori pretpostavljaju da velika gusto a prometa rezultira manjim brzinama. Podru ja s ve om gusto om zaposlenosti imaju ve i broj prometnih nesre a, vjerojatno zbog visoke aktivnosti na ulicama. Podru ja s višom stopom siromaštva imaju više nesre a, a karakteristike prometnica (broj obilaznica i križanja) ne utje u na nesre e. Rezultati drugog istraživanja koje su proveli Wang, Quddus i Ison (Wang et al., 2009) su potvrdili pretpostavku da je prosje na brzina statisti ki zna ajna i povezana s težinom nesre e zbog gusto e prometa. Zakrivljenost ceste je statisti ki zna ajna s negativnim koeficijentom. To ukazuje da su voza i svjesniji stanja na cesti na zakrivljenim cestama i da opreznije voze kroz naseljena i komercijalna podru ja gdje ima više zavoja. Zakrivljene ceste u prostornoj jedinici mogu se smatrati zaštitnim faktorom u pogledu sigurnosti na cestama. Rezultati tre eg istraživanja kojeg je proveo Quddus (Quddus, 2008) su pokazali da je gusto a prometa povezana sa žrtvama u prometu. Varijable poput ku anstva bez vozila i zaposlenost su statisti ki zna ajni u svim modelima, a prosje na brzina prometa u popisnim krugovima nema utjecaja na težinu ozljeda.

U Sjedinjenim Ameri kim Državama su provedena dva istraživanja na podru ju San Francisca [(LaScala et al., 2000), (Wier et al., 2009)]. Analizirani su utjecaji prometne infrastrukture i demografskih karakteristika prostornih jedinica na stopu prometnih nesre a motornih vozila i pješaka. Podaci su preuzeti iz razli itih institucija. Podaci o prometnim nesre ama su preuzeti iz Državnog integriranog prometnog sustava (*eng. Statewide Integrated Traffic Records System – SWITRS*), demografski podaci iz popisa stanovništva, a podaci o prometnoj infrastrukturi od Odjela za javne radove San Francisca. Rezultati prvog istraživanja

koje su proveli LaScala, Gerber i Gruenewald (LaScala et al., 2000) ukazuju na značajnu geografsku korelaciju između u okolišnih i demografskih karakteristika u nesrećama vozila i pješaka. Najveća vjerojatnost nesreće vozila i pješaka postoji u područjima s većom gustoćom stanovništva i većom nezaposlenošću. Detaljniji uvid pokazuje da je riječ o područjima s većim udjelom muškaraca, manjim udjelom djece od 0 do 15 godina i manjim udjelom visokoobrazovanih stanovnika. Autori pretpostavljaju da niže obrazovanje može biti povezano s fizičkim poslovima koji izlažu radnike opasnostima u prometu, a nezaposlenost može biti povezana s većim brojem aktivnosti na otvorenom i izlaganjem rizicima u prometu. Nesreće s pješacima su češće u dijelovima grada s većom gustoćom prometa i većom koncentracijom barova. Rezultati drugog istraživanja koje su proveli Wier, Weintraub, Humphreys, Seto i Bhatia (Wier et al., 2009) su pokazali da je gustoća prometa primarni okolišni uzrok nesreća između u vozila i pješaka na promatranim područjima.

Cilj istraživanja provedenog u središnjoj Floridi (Abdel-Aty i Radwan, 2000) je uspostava matematičkog modela koji objašnjava razliku između u frekvencije prometnih nesreća i karakteristika autocesta. Podaci o prometnim nesrećama su preuzeti iz baze podataka Odjela za sigurnost na autocestama i motornih vozila za razdoblje od 1992. do 1994., a podaci o geometrijskim i prometnim karakteristikama cesta iz baze podataka Odjela za promet Floride. Razmatra se glavna autocesta dugačka 227 km koja povezuje istočnu i zapadnu obalu središnje Floride. Autocesta prolazi kroz centar grada Orlando. Rezultati su pokazali da se povećanjem prosječnog godišnjeg dnevnog prometa (*eng. Annual Average Daily Traffic - AADT*) značajno povećava frekvencija prometnih nesreća. Širina i broj prometnih traka su među značajnijim geometrijskim parametrima. Uske trake i velik broj traka povećava frekvenciju prometnih nesreća, a urbani cestovni segmenti su potencijalno opasniji od ruralnih. Žene, mladi i stariji vozači su podložniji prometnim nesrećama u većim gustoćama prometa, na užim trakama i većem broju traka. Muškarci i mladi vozači su češće uključeni u prometne nesreće povezane s visokim brzinama.

Istraživanje provedeno u Ujedinjenom Kraljevstvu na području Norfolka (Jones et al., 1996) nastoji utvrditi važnost vremena odaziva hitne pomoći na vjerojatnost preživljavanja žrtve prometne nesreće. Analizirani podaci su rezultat istrage smrtnih slučajeva u prometnim nesrećama u Norfolku od 1987. do 1991. godine. Glavni izvor podataka su policijski zapisnici, a korištena je i digitalna cestovna mreža Norfolka pri simulaciji ruta hitne pomoći. Procjene vremena odaziva hitne pomoći napravljene su pomoću GIS-a. Rezultati su pokazali da postoji veća vjerojatnost da će osobe starije od 60 godina umrijeti u nesrećama koje su se

dogodile na cestama s ograničenjem brzine od 100 km/h i u prometnim nesrećama s tri ili više žrtava. Pješaci su najugroženiji sudionici u prometu i vrijeme intervencije hitne pomoći nije presudno u preživljavanju. Model se koristio u predviđanju ishoda preživljavanja svake žrtve s obzirom na težinu i karakteristike nesreće. To nije predvidio većinu ishoda.

U istraživanju provedenom u Belgiji (Depaire et al., 2008) provedena je analiza prometnih nesreća na području Bruxellesa s ciljem proučavanja učinkovitosti tehnike grupiranja (*eng. clustering*). Podaci prometnih nesreća su preuzeti od Državnog zavoda za statistiku Belgije (*eng. Directorate-general Statistics Belgium*) za razdoblje od 1997. do 1999. godine. Autori su došli do zaključka da primijenjena metoda grupiranja po klasama daje zadovoljavajuće rezultate nad heterogenim setom podataka prometnih nesreća. Grupiranje po klasama kao primarna analiza može otkriti skrivene odnose i pomoći stručnjacima u prometnoj sigurnosti pri segmentiranju prometnih nesreća.

Istraživanje provedeno u južnoindijskom gradu Thiruvananthapuramu (Prasannakumar et al., 2011) je provedeno s ciljem identifikacije i analize žarišta prometnih nesreća uzrokovanih neodgovarajućim razvojem prometne mreže. Podaci o prometnim nesrećama su preuzeti iz sjedišta prometne policije grada Thiruvananthapurama za godinu 2008. Granice popisnih krugova su digitalizirane i korištene za izdvajanje cestovne mreže iz topografskih karata mjerila 1:25000. Nakon digitalizacije cestovna mreža je nadopunjena podacima GPS izmjere. Rezultati analize prostornih grupiranja prikazuju prostorne i vremenske varijacije žarišta što potvrđuje da prometne nesreće ovisе o vremenskim i prostornim varijablama.

U istraživanju provedenom u Afyonkarahisar, Turskoj (Erdogan et al., 2008), identificirana su žarišta na autocestama unutar spomenutog administrativnog područja. Podaci o prometnim nesrećama su preuzeti od područnog ureda prometne policije (*eng. District Traffic Agency Officers*) za razdoblje od 1996. do 2006. godine. Autoceste spomenutog područja su digitalizirane s karata mjerila 1:1,000,000. Rezultati pokazuju da je gustoća prometnih nesreća najveća u kolovozu i prosincu. Frekvencija prometnih nesreća na dnevnoj razini je najviša petkom i vikendom, što ukazuje na utjecaj turističkih putovanja. Većina prometnih nesreća se dogodila oko podneva, ali najteže prometne nesreće su najčešće oko ponoći. Najčešći uzrok istih je neprimjereni ponašanje vozača.

U Londonu je provedeno istraživanje (Anderson, 2009) s ciljem razvoja metodologije identifikacije žarišta prometnih nesreća pomoću GIS-a i procjene vjerojatnosti raspodjele slučajne varijable na temelju funkcije gustoće (*eng. Kernel Density Estimation – KDE*).

Također je obrađeno pridruživanje atributnih podataka spomenutim zonama i identifikacija zona na temelju grupiranja atributa oko najbliže rednje vrijednosti (*eng. K-means clustering*). Podaci prometnih nesreća preuzeti su iz STATS19 baze podataka Odjela za promet i prometne policije. Podaci se odnose na razdoblje od 1999. do 2003. godine. Žarištima su pridruženi atributni podaci o okolini i pomoću K-means grupiranja određena su područja s velikom gustoćom prometnih nesreća. Na taj je način uspostavljen model koji omogućuje usporedbu i grupiranje žarišta sa sličnim karakteristikama.

## 2. Opis i specifični ciljevi rada

Na raspolaganju su nam podaci o prometnim nesrećama koje su se dogodile na području Grada Zagreba od 2010. do 2013. godine. Prema našim saznanjima, ovo je prva prostorna analiza i obrada ovog skupa podataka u Republici Hrvatskoj. Prvi korak je utvrđivanje kvalitete preuzetih podataka. To podrazumijeva provjeru pouzdanosti prostornih koordinata i atributa pojedine prometne nesreće. Budući da je riječ o više od 35 000 nesreća, potrebno je razviti metodologiju automatizirane obrade podataka.

S ispravljenim podacima prelazi se u izgradnju baze podataka koja će biti temelj budućeg GIS-a prometnih nesreća. Samo prikupljanje i pohrana podataka nisu isplativi bez primjene statističkih analiza. Iz tog razloga nije stavljen naglasak na pohranu podataka već na primjenu metoda prostorne statistike. Uzevši u obzir prethodno navedena inozemna istraživanja, obrađene su tri metode. Riječ je o hi-kvadrat testu, normalizaciji i Kernel Density Estimationu (KDE). Jedan od ciljeva našeg istraživanja je utvrditi prednosti i mane pojedine metode u obradi danog skupa podataka. Hi-kvadrat test ne omogućuje prostornu vizualizaciju podataka, ali je dobra polazna osnova jer omogućuje generalni pregled stanja podataka. Normalizacija je obrađena kao primjer metode koja uključuje podjelu na prostorne zone. KDE je relativno nova statistička metoda koja omogućuje identifikaciju žarišta prometnih nesreća.

Sigurnost u prometu dobiva sve veći značaj, a statističke analize su neophodne u istraživanju uzroka i provođenju uinkovitih protumjera. Cilj rada je uspostaviti metodologiju obrade i analize prometnih nesreća metodama prostorne statistike. Rezultati bi omogućili gradskim vlastima donošenje kvalitetnih odluka po pitanju sigurnosti u prometu i prometnim službenicima njihovo provođenje. Budući da su rezultati namijenjeni strukama koje nisu u doticaju s prostornom statistikom, posebna je pozornost posvećena vizualizaciji rezultata.

### 3. Materijal i metode

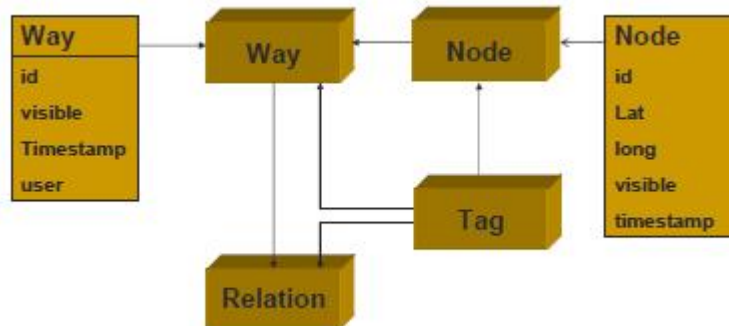
Podaci o prometnim nesrećama su preuzeti od Ministarstva unutarnjih poslova Republike Hrvatske (u daljnjem tekstu: MUP) preko Fakulteta prometnih znanosti Sveučilišta u Zagrebu. Riječ je o podacima za 2010., 2011., 2012. i 2013. godinu. Podaci o prometnoj mreži preuzeti su sa slobodne karte svijeta OpenStreetMap. Kao podloga za vizualizaciju rezultata korišteni su slojevi OpenStreetMap i OpenCycleMap, dostupni preko platforme za razvoj web karata OpenLayersa. Podaci su obrađeni pomoću programskog jezika Python i uneseni u bazu podataka PostgreSQL. Za potrebe obrade prostornih podataka primijenjeno je prostorno proširenje baze podataka PostgreSQL poznato pod nazivom PostGIS. Kao što je ranije navedeno, primijenjene su tri statističke metode za obradu danog skupa podataka: hi-kvadrat test, normalizacija i Kernel Density Estimation. Rezultati sve tri metode su analizirani i istražene su dodatne mogućnosti vizualizacije rezultata posljednje dvije metode.

#### 3.1. Preuzimanje prometne mreže Grada Zagreba s karte svijeta OpenStreetMap

OpenStreetMap (u daljnjem tekstu: OSM), projekt slobodne karte svijeta, omogućuje korisnicima suradnju u prikupljanju i dijeljenju geografskih podataka, prvenstveno prometnica i izmeću naseljenih mjesta. Projekt je nastao iz razloga što većina karata koje se smatraju slobodnima za korištenje imaju pravna ili tehnička ograničenja za njihovu uporabu. To sprječava ljude da ih koriste na kreativne i produktivne načine (URL 2). Projekt je započeo 2004. godine na University College London, s ciljem stvaranja besplatne baze podataka s prostornim informacijama iz cijelog svijeta. Baza podataka projekta sadrži prostorne podatke poput cesta, zgrada, i dr. Bilo koji korisnik može sudjelovati u izradi projekta i uređivati podatke nakon kratke online registracije (Neis i Zipf, 2012).

Geografske informacije u OSM bazi podataka pohranjene su koristeći tri tipa objekata: čvor, put i relacija (*eng. node, way, relation*) (Slika 1). Čvorovi u bazi podataka sadrže informacije o lokaciji. Riječ je o točkama s pridruženom geografskom širinom i dužinom. Linije poput cesta i poligoni se spremaju kao putevi i relacije, koji definiraju logičke i geografske odnose između objekata. Otvoreni pristup i slobodno prikupljanje prostornih podataka podižu pitanja o kvaliteti prostornih podataka. Studije koje su se bavile ovim pitanjem zaključile su da skup podataka OSM-a ima vrlo heterogenu kvalitetu. Drugim riječima, kvaliteta i potpunost baze

podataka varira od države do države. U urbanim područjima Europe, posebno u Ujedinjenom Kraljevstvu, Njemačkoj, Austriji i Švicarskoj, OSM podaci pokazuju potpunost na razini komercijalnih i državnih institucija. U ruralnim područjima (s iznimkom Sjedinjenih Američkih Država) postoji znatno slabija pokrivenost OSM podacima (Neis i Zipf, 2012).



Slika 1. Pojednostavljeni model podataka za OpenStreetMap (URL 3)

Provedena je analiza prometnih nesre a Grada Zagreba iz razloga što Zagreb ima najbolju pokrivenost OSM podacima od svih gradova u Republici Hrvatskoj. S OSM servera preuzeta je mreža ulica Grada Zagreba u SHP (*eng. Shapefile*) formatu koja je korištena kao osnova u obradi podataka prometnih nesre a.

### 3.2. Obrada podataka iz policijskih zapisnika pomo u programskog jezika Python

Podaci o prometnim nesre ama su preuzeti u CSV (*eng. Comma-Separated Values*) formatu. Svako vozilo prometne policije ima GPS prijemnik, što zna i da svaka prometna nesre a ima prostorne koordinate (geografska širina i dužina). Otežavaju a je okolnost nepoznat kutni format zapisa. Primjerice, geografska širina ima vrijednost 4512345, a geografska dužina 1812345. Ovakav na in zapisa geografskih koordinata (bez interpunkcijskih znakova) nije prikladan za daljnju obradu jer postoje tri osnovna zapisa kutnih koordinata:

- *DMS* - Stupnjevi (cijeli broj), minute (cijeli broj), sekunde (cijeli ili realni broj);
- *Decimalne minute* - Stupnjevi (cijeli broj), minute (realni broj);
- *Decimalni stupnjevi* - Stupnjevi (realni broj).

Potrebno je provjeriti u kojem su formatu zapisane geografske koordinate iz policijskog zapisnika i po potrebi ispraviti zapis. To podrazumijeva konverziju koordinata u jedan od

navedenih formata. U tu su svrhu korišteni nazivi ulica iz zapisnika. U zapisniku svaka nesre a može imati najviše dva atributa ulica, *ulica1* i *ulica2*. Koriste i mrežu ulica preuzetu s OSM-a, mogu e je napraviti kontrolu konverzije prostornih koordinata. Problem nastaje kad u zapisniku ne piše naziv ulice na kojoj se dogodila nesre a ili kad naziv ulice ne odgovara nazivu ulice iz OSM-a.

U prvom slu aju ne postoji mogu nost kontrole konverzije prostornih koordinata. Tada se koordinate konvertiraju u decimalne minute. Taj je format odabran iz razloga što GNSS ure aji policijskih službenika ispisuju koordinate u decimalnim minutama (Slika 2, Slika 3).



*Slika 2. Prikaz jedne vrste GNSS ure aji prometne policije Grada Zagreba (izvor: MUP)*



*Slika 3. Prikaz druge vrste GNSS ure aji prometne policije Grada Zagreba (izvor: MUP)*



U drugom se slučaju to se ocjenjuje kao nepouzdana na osnovu Damerau-Levenshteinovog algoritma metrike udaljenosti, bez obzira je li riječ o pravim ulicama ili ne. Damerau-Levenshteinov algoritam udaljenosti (*eng. Damerau-Levenshtein Distance Metric*) je rezultat istraživanja pravopisnih pogrešaka koje su proveli Damerau i Levenshtein. Riječ je o funkciji koja od konačnog skupa znakova abecede dolazi do brojeva. To je udaljenost u smislu da nadanim stringovima (niz znakova)  $s_1$ ,  $s_2$  i  $s_3$  vrijede sljedeći uvjeti (Bard, 2007):

- *Nenegativnost*:  $d(s_1, s_2) \geq 0$
- *Nedegeneracija*:  $d(s_1, s_2) = 0$  ako i samo ako  $s_1 = s_2$
- *Simetrija*:  $d(s_1, s_2) = d(s_2, s_1)$
- *Nejednakost trokuta*:  $d(s_1, s_2) + d(s_2, s_3) \geq d(s_1, s_3)$

Udaljenost  $d(s_1, s_2)$  je definirana na sljedeći način. Operacija može biti ubacivanje, brisanje ili zamjena znaka s drugim na jednoj lokaciji ili zamjena dva susjedna znaka. Moguće je velik broj kombinacija ove četiri operacije koje mogu konvertirati znak  $s_1$  u  $s_2$ , ali duljina najkraćeg slijeda je udaljenost između dva znaka. Operacijama se mogu dodijeliti težine različite od jedan. Damerau je pokazao kako 80% tipografskih pogrešaka u ovom modelu imaju udaljenost 1 (Bard, 2007).

Konverzija prostornih koordinata pomoću naziva ulica iz zapisnika na osnovu Damerau-Levenshteinovog algoritma provedena je pomoću programskog jezika Python. Python je interaktivni objektno orijentirani programski jezik pod licencom otvorenog koda (*eng. open source licence*) što znači da se može slobodno koristiti i distribuirati, čak i u komercijalne svrhe. Razvoj je započeo 1989. godine, a objavljen je početkom 1991. godine. Uključuje module, iznimke, visokodinamične tipove podataka i klase. Iznimno je modern i ima vrlo jasnu sintaksu. Dostupan je za sve veće operativne sustave poput Windowsa, Linux/Unixa, OS/2, Maca i Amige. Osim standardnih biblioteka sadrži i module treće strane (*eng. third party modules*) za praktički sve zadatke. Ukoliko Python nije u stanju izvršiti neki zadatak ili su potrebne prednosti koda niske razine (*eng. low-level code*) mogu se napisati ekstenzije i moduli u C i C++ programskim jezicima, Javi za Jython i .NET-u za IronPython. Među najpoznatijim korisnicima Pythona su Google i YouTube (URL 4).

Napisana je skripta u Pythonu koja uključuje podatke o prometnim nesrećama i mrežu ulica Grada Zagreba preuzetu s OSM-a. Ako prometna nesreća (predstavljena točkom) nema među atributima naziv ulice na kojoj se dogodila, prostorne koordinate se konvertiraju u decimalne minute. Ukoliko je udaljenost točke do najbliže ceste manja od 100 m, nesreća se proglašava

pouzdana i ide u daljnju obradu. U suprotnom se zanemaruje. Cilj ovog dijela postupka je eliminacija to aka koje se nalaze izvan područja Grada Zagreba ili države, odnosno to aka optere enih grubim pogreškama.

Prometne nesre e koje me u atributima imaju bar jedan naziv ulice konvertiraju se u sva tri kutna formata. Drugim rije ima, nastaju tri to ke koje predstavljaju istu prometnu nesre u, ali se nalaze na razli itim lokacijama (Tablica 1). Oko svake to ke se kreira zona obuhvata (objekt tipa poligon) s radiusom od 1000 m. Stvara se unija tri zone obuhvata što u ve ini slu ajeva rezultira objektom tipa multipoligon. Multipoligon se preklapa se s mrežom ulica. Cilj ovog postupka je smanjenje obujma ra unanja. To se postiže izdvajanjem ulica koje se nalaze unutar multipoligona ili ga sijeku. Naziv ulice iz zapisnika se uspore uje s nazivima izdvojenih ulica iz OSM-a pomo u Damerau-Levenshteinovog algoritma udaljenosti. Prethodnim testiranjem je utvr eno da prihvatljiva grani na vrijednost algoritma iznosi 5. Ulice iz OSM-a za koje je vrijednost algoritma manja od ili jednaka 5 idu u daljnju obradu, a ostale se zanemaruju. Cilj primjene algoritma je daljnje smanjenje obujma ra unanja. To se postiže eliminacijom ulica iz multipoligona iji nazivi nimalo ne odgovaraju nazivu ulice iz zapisnika. Sljede i korak je ra unanje najmanje udaljenosti izme u to ke i preostalih linija, odnosno prometne nesre e i potencijalnih ulica. Uzevši u obzir da postoje tri to ke s okolnim linijama, uzima se u obzir par koji postiže najmanju udaljenost. Format te to ke se uzima kao ispravan, a ako je udaljenost manja od 150 m, to ka se ocjenjuje pouzdanom i ide u daljnju obradu. U suprotnom se zanemaruje. Tablica 1 prikazuje primjer ovog postupka. Sve tri to ke predstavljaju istu prometnu nesre u, samo su formati koordinata, a time i lokacije, druga iji. Izra unata je udaljenost svake to ke do najbliže linije. Treba napomenuti da ne mora biti udaljenost do iste linije jer se to ke esto nalaze na udaljenosti i do nekoliko kilometara. Budu i da najmanja udaljenost iznosi 3 metra, prometnoj nesre i se pridružuju koordinate DMS formata. Druge dvije to ke se zanemaruju.

*Tablica 1. Primjer odabira formata koordinata*

| <i>To ka br.</i> | <i>Format koordinata</i> | <i>Udaljenost do najbliže linije [m]</i> |
|------------------|--------------------------|------------------------------------------|
| 1                | DMS                      | 3                                        |
| 2                | Decimalne minute         | 5                                        |
| 3                | Decimalni stupnjevi      | 7                                        |

Uzevši u obzir današnju točnost i dostupnost GNSS tehnologije, vrijednosti od 100 m, odnosno 150 m za najmanju udaljenost djeluju velikodušno. Treba uzeti u obzir brojne okolišne faktore i strukturnu spremu osobe koja koristi prijemnik. Većina nesreća se događa u urbanim dijelovima grada gdje multipath, odnosno refleksija GNSS signala, igra presudnu ulogu. GNSS prijamnici su fiksno postavljeni u policijsko vozilo (Slika 2, Slika 3) što otežava situaciju ako se nesreća dogodila na prometnoj ulici. Automobili se moraju isključiti iz prometa što ih nerijetko udaljava od mjesta nesreće i kviri podatke (Slika 57). Prometne nesreće su i stresne situacije za sudionike i policijske službenike koji rade uviđaj. Jedan od najvažnijih faktora je strukturna spremu policijskog službenika koji koristi GNSS prijamnik. Ako im nije dobro objašnjena važnost i prednost spomenutog sustava, može doći do raznih propusta. Ukoliko se nesreća dogodila na sporednoj ulici ili prilazu većoj ulici, policijski službenik će unijeti naziv većije ulice. Uzevši u obzir količinu atributa koju moraju unijeti u kratkom vremenu, previdi su normalna i opravdana pojava. Ukoliko zaborave unijeti koordinate iz GNSS prijavnika na mjestu nesreće, postoji mogućnost da kasnije pokušavaju doći do njih na drugačiji način. Primjerice, vraćanjem na mjesto nesreće ili korištenjem aplikacija poput Google Eartha koje koriste drugačije kutne formate prostornih koordinata.

Tablica 2. Razlike u nazivima ulica između zapisnika i OSM-a

| <i>Ulica1/Ulica2<br/>(policijski zapisnik)</i> | <i>OpenStreetMap</i>             |
|------------------------------------------------|----------------------------------|
| Palmotićeva ulica                              | Ulica Junija Palmotićeva         |
| Boškovićeva ulica                              | Ulica Josipa Jurija Boškovića    |
| Škorićeva ulica                                | Ulica Ignjata Škorića            |
| Vlaška                                         | Vlaška ulica                     |
| Kačićeva ulica                                 | Ulica Fra Andrije Kačića Miošića |
| Klaićeva ulica                                 | Vjekoslava Klaića                |
| Mihanovićeva ulica                             | Antuna Mihanovića                |
| Heinzelova                                     | Ulica Vjekoslava Heinzela        |
| Podaupskog                                     | Ulica Eugena Podaupskog          |
| Jarunska ulica                                 | Jarunska                         |

Tablica 2 prikazuje razlike u nazivima ulica između zapisnika prometne policije i podataka OSM-a. Policijski zapisnici vode se analogno. Uzevši u obzir količinu atributa koje policijski

službenici trebaju unijeti, propusti su ova pojava. Primijećeno je velik broj nesreća u kojima policijski službenici ne upisuju puni naziv ulice. To rezultira zanemarivanjem tih opažanja, bez obzira na dobre prostorne koordinate.

### 3.3. Unos i obrada podataka pomoću PostgreSQL-a i PostGIS-a

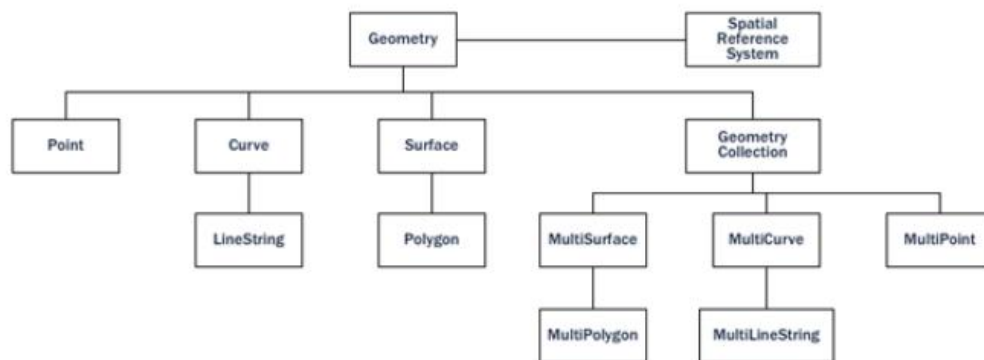
Rezultati obrade podataka zapisnika spremljeni su u CSV datoteku i uneseni u bazu podataka. Korištena je baza podataka PostgreSQL s prostornim proširenjem PostGIS. PostgreSQL je modernan sustav za upravljanje objektno-relacijskim bazama podataka (*eng. Object-Relational Database Management System - ORDBMS*) zasnovan na slobodnom kodu. Primarno je razvijen od strane Berkeley Computer Science Departmenta na University of California. Može biti korišten, modificiran i distribuiran u bilo koje svrhe, bilo da su privatne, komercijalne ili akademske. Aktivno se razvija već 15 godina i poznat je po pouzdanosti, integritetu podataka i točnosti. Radi na svim većim operativnim sustavima, uključujući i Linux, Unix i Windows. Zadovoljava ACID (*eng. Atomicity, Consistency, Isolation, Durability*) uvjete i ima punu podršku za strane ključeve, pridruživanje, poglede i ostale pohranjene procedure u brojnim jezicima. Uključuje većinu SQL:2008 tipova podataka i podržava pohranu velikih binarnih objekata poput slika, zvukova i videa. Ne postoji ograničenje na veličinu baze podataka, ali tablica može imati najviše 32 TB. PostgreSQL podržava složene, jedinstvene, djelomične i funkcionalne indekse koji mogu koristiti B-tree, R-tree, hash ili GiST metodu pohrane (URL 5).

GiST (*eng. Generalized Search Tree*) je napredan sustav indeksiranja koji uključuje brojne algoritme sortiranja i pretraživanja. Nudi fleksibilnost u određivanju načina pohrane te mogućnost definicije novih načina pretraživanja. To uvelike nadilazi mogućnosti standardnih B-tree, R-tree i ostalih generaliziranih algoritama. GiST je temelj mnogih javnih projekata poput OpenFITS-a i PostGIS-a koji koriste PostgreSQL. OpenFITS (*eng. Open Source Full Text Search Engine*) omogućuje online indeksiranje podataka i rangiranje po važnosti u pretraživanju baze podataka. PostGIS je projekt koji dodaje potporu za geografske objekte u PostgreSQL-u. Omogućuje korištenje PostgreSQL-a kao baze prostornih podataka za geografske informacijske sustave poput ESRI-jevog SDE-a ili Oracleove Spatial ekstenzije (URL 5).

PostGIS je baza prostornih podataka. Baze prostornih podataka pohranjuju i manipuliraju prostornim objektima kao i svim ostalim objektima u bazi podataka. Postoje tri aspekta koji povezuju prostorne podatke s bazama podataka (URL 6):

- *Tipovi prostornih podataka* (točka, linija, poligon, itd.);
- *Multidimenzionalno prostorno indeksiranje* za učinkovito procesiranje i prostorne operacije;
- *Prostorne funkcije* u SQL-u za upite o prostornim svojstvima i odnosima.

Baza prostornih podataka dodaje prostorne tipove podataka koji predstavljaju geometrijska svojstva i koji obuhvaćaju prostorne strukture poput granice i dimenzije. Tipovi prostornih podataka su organizirani hijerarhijski. Svaki podtip nasljeđuje strukturu i ponašanje svog nadtipa (URL 6).



Slika 4. Hijerarhija geometrijskih tipova (URL 6)

Baze podataka pružaju metode pristupa, poznate kao indeksi, koje omogućuju brz pristup podacima. Indeksiranje standardnih tipova se često izvodi pomoću B-tree indeksa, ali geometrijski se objekti mogu preklapati i sadržavati. U tom slučaju B-tree ne pomaže. Obuhvatni pravokutnici (eng. *bounding box*) se koriste zato što je rad unutar poligonima zahtjevniji, ali jako brz s pravokutnicima. Implementirani prostorni indeksi se razlikuju od baze do baze, a najčešće se koristi R-tree indeks kojeg koristi i PostGIS. Baze prostornih podataka pružaju skup funkcija za analizu geometrijskih komponenata, određivanje prostornih odnosa i manipulaciju geometrijom (URL 6).

Podaci prometnih nesreća su uneseni u tri tablice: *nesrece*, *sudionici* i *vozila*. Ključ prometne nesreće (oznaka iz zapisnika: ključ\_pn) u tablici *nesrece* je primarni ključ na osnovu kojeg se povezuju zapisi iz tablica *sudionici* i *vozila*. Svakoj prometnoj nesreći dodijeljeni su atributi

vremena nesre e tipa *timestamp without time zone* i geometrije nesre e s ciljem lakše obrade podataka.

Naizgled se isti rezultat mogao postići i učitavanjem CSV datoteke s prometnim nesre ama u GIS aplikaciju i spremanjem u SHP format. S pojavom prvih GIS aplikacija, SHP format je bio standardni na in pohrane i manipulacije prostornih podataka. Me utim, SHP-ovi imaju sljede e nedostatke (URL 6):

- *Datoteke zahtijevaju posebne aplikacije za učitavanje i obradu podataka.* SQL je apstrakcija za pristup i analizu podataka. Bez te apstrakcije, korisnici bi sami trebali napisati kodove za analizu i pristup podacima;
- *Istovremeno korištenje podataka od strane više korisnika može pokvariti podatke.* Iako je mogu e napisati dodatni kod koji omogu uje da višestruki unosi u datoteku ne pokvare podatke, dok se riješi taj problem i problemi vezani za izvršavanje napisat e se dobar dio sustava baze podataka;
- *Složeni upiti zahtijevaju složene aplikacije za odgovor.* Složeni i zanimljivi upiti (*eng. spatial joins, aggregations*) koji se mogu napisati u jednoj liniji SQL koda u bazi podataka su stotine linija specijaliziranog koda kad se programira s datotekama.

Ve ina korisnika PostGIS-a uspostavlja sustave gdje mnoge aplikacije imaju pristup podacima preko interneta. Primjena standardne SQL metode pristupa podacima pojednostavljuje razvoj. Korisnici koji rade s velikom koli inom podataka nisu ograni eni na razdvajanje podataka u više datoteka koji u bazi podataka mogu biti pohranjeni u jednu veliku tablicu. Kombinacija podrške mnogobrojnim korisnicima, složenih ad hoc upita i manipulacije velikog seta podataka odvaja baze prostornih podataka od sustava zasnovanih na datotekama (*eng. file-based systems*) (URL 6).

### 3.4. Analiza podataka pomo u hi-kvadrat testa

Sljede i korak u istraživanju je statisti ka obrada podataka prometnih nesre a. Prva metoda obrade je provo enje hi-kvadrat testova. Postoji nekoliko razli itih vrsta hi-kvadrat testova. Neke od njih se koriste pri procjeni nezavisnosti dviju varijabli, homogenosti varijabli ili procjeni dobrote prilagodbe podataka uzorka nekoj teorijskoj distribuciji. Struktura ulaznih i interpretacija dobivenih podataka ovisi o metodama, ali je postupak ra unanja u svakoj metodi mogu na sli an na in (Gibson i Dillard, 2012).

U radu je korišten hi-kvadrat test homogenosti s ciljem ispitivanja proporcionalnosti broja prometnih nesreća s različitim posljedicama po pojedinim atributima. Podaci koji se koriste pri ispitivanju moraju biti kategorizirani u tablicama kontingencije. Tablice kontingencije organiziraju rezultate istraživanja u dvije kategorije pomoću prikaza apsolutnih ili relativnih frekvencija elemenata svake skupine (URL 7).

Tablica 3. Primjer strukture tablice kontingencije (URL 7)

| Kategorija A                         | Kategorija B |           |       |           | Zbroj<br>(vrijednosti u<br>retcima) |
|--------------------------------------|--------------|-----------|-------|-----------|-------------------------------------|
|                                      | Skupina 1    | Skupina 2 | ..... | Skupina c |                                     |
| Skupina 1                            | $f_{11}$     | $f_{12}$  | ..... | $f_{1c}$  | Redak 1                             |
| Skupina 2                            | $f_{21}$     | $f_{22}$  | ..... | $f_{2c}$  | Redak 2                             |
| ...                                  | ...          | ...       | ..... | ...       | ...                                 |
| ...                                  | ...          | ...       | ..... | ...       | ...                                 |
| Skupina r                            | $f_{r1}$     | $f_{r2}$  | ..... | $f_{rc}$  | Redak r                             |
| Zbroj<br>(vrijednosti u<br>stupcima) | Stupac 1     | Stupac 2  | ..... | Stupac c  | n (ukupan<br>zbroj)                 |

Nakon što su podaci kategorizirani u tablici, slijedi postavljanje hipoteza. Potrebno je postaviti nul-hipotezu i alternativnu hipotezu. Hipoteze moraju biti postavljene tako da se međusobno isključuju. Pretpostavimo da su podaci prikupljeni od  $r$  skupina jedne kategorije i da postoji  $c$  skupina druge kategorije. Nul-hipoteza  $H_0$  nalaže da svaka skupina jedne kategorije ima jednake proporcije u frekvenciji po skupinama druge kategorije. Alternativna hipoteza  $H_a$  nalaže da barem jedna hipoteza nije prihvaćena (URL 8).

Provedba testa nastavlja se određivanjem broja stupnjeva slobode. Broj stupnjeva slobode  $DF$  jednak je (URL 8):

$$DF = (r - 1)(c - 1)$$

gdje je  $r$  broj skupina jedne kategorije (broj redaka), a  $c$  broj skupina druge kategorije (broj stupaca).

Potrebno je izračunati i očekivane frekvencije za svaki element svake skupine, prema formuli (URL 8):

$$E_{r,c} = \frac{n_r * n_c}{n}$$

gdje je  $E_{r,c}$  očekivana frekvencija za element na presjeku skupina  $r$  i  $c$ ,  $n_r$  je ukupan zbroj vrijednosti svih elemenata skupine  $r$ ,  $n_c$  ukupan zbroj vrijednosti svih elemenata skupine  $c$ , a  $n$  je ukupna veličina uzorka.

Sljedeći korak je računanje vrijednosti hi-kvadrata pomoću formule (URL 8):

$$\chi^2 = \sum \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}}$$

pri čemu je  $O_{r,c}$  opažena frekvencija, a  $E_{r,c}$  očekivana frekvencija određenog elementa u tablici kontingencije. Na temelju dobivene vrijednosti hi-kvadrata računa se *P-vrijednost*. *P-vrijednost* se pod pretpostavkom hipoteze  $H_0$ , definira kao vjerojatnost dobivanja jednakog ili ekstremnijeg rezultata od rezultata provedenog istraživanja. Zatim se ta vrijednost uspoređuje s razinom značajnosti. Najčešće se unaprijed definira razina značajnosti kao 0.05, odnosno 5%. Ukoliko je izračunata *P-vrijednost* manja od 0.05, nul-hipoteza se odbacuje (URL 8).

Primjena hi-kvadrat testa zahtijeva ispunjenje sljedećih uvjeta (URL 8):

- Varijable u tablici kontingencije moraju biti brojano određive;
- Suma očekivanih frekvencija mora biti jednaka sumi opažanih frekvencija;
- Ako tablica kontingencije ima više od 2 elije, manje od 20% očekivanih frekvencija smije biti manje od 5;
- elije u tablici kontingencije moraju biti međusobno nezavisne, odnosno jedna vrijednost se smije pojaviti isključivo u jednoj eliji.

U radu su provedeni hi-kvadrat testovi kod kojih su svi prethodno navedeni uvjeti ispunjeni.

### 3.5. Analiza podataka metodom normalizacije

Postoje brojni razlozi za mijenjanje vrijednosti atributa tijekom analize prije kartiranja skupa podataka. Jedan od razloga je što karta koja prikazuje zone s prostorno opsežnim atributima može biti varljiva. Primjerice, podjela prostora na općine koje prikazuju populaciju stanovništva. Očito se jedinstvena vrijednost predložena određenom bojom na karti ne može primijeniti na cijelu promatranu zonu. To nije ni smisleniji način je podjela područja na zone i prikaz prostorno intenzivnih atributa (*eng. spatially intensive attribute*). Primjerice, gustoća populacije kroz zone (Smith et al., 2007).



Ako su podaci organizirani po područjima, često je nužno stvaranje omjera. Cilj je prilagoditi kvantitativne podatke (jednostavan izraz, kontinuirana varijabla ili težina) vrijednosti zone u procesu koji se zove *normalizacija*. Termin se koristi u nekoliko prostornih konteksta, uključujući i prostornu analizu, matematičke izraze i topološke operacije. U ovom kontekstu, proces se odnosi na dijeljenje broja ili težine podataka na osnovu atributne vrijednosti koja će se pretvoriti u mjeru intenziteta za promatranu zonu. Ovaj proces uklanja posljedice povezane s podacima koji su povučeni iz zona različitih površina ili veličina populacije i omogućava usporedbu podataka temeljenih na zonama (*eng. zone based data*) (Smith et al., 2007).

Često su korištena tri tipa normalizacije: prosjeci, postoci i gustoće. Kod prosjeka se atributna vrijednost u svakoj zoni prilagođava dijeljenjem s drugom atributnom vrijednosti kako bi se dobila prosječna vrijednost koja se može pohraniti u novo polje ili kartirati. Primjerice, broj djece u zoni se dijeli s brojem kućanstava u istoj zoni. U slučaju postotaka ili pretvaranja u raspon [0, 1], odabrana vrijednost atributa se dijeli s maksimalnom vrijednošću ili ukupnim brojem u setu podataka. Primjerice, broj nezaposlenih osoba u zoni se dijeli s ukupnim brojem radno sposobnih ljudi (npr. 18-65). Kod gustoća, djelitelj je površina zone gdje je pronađen određeni, odnosno traženi atribut. Rezultat je skup vrijednosti koje prikazuju populaciju po jedinici površine, odnosno mjera gustoće. Primjerice, broj stanova/km<sup>2</sup>, tona pšenice/jutro i slično (Smith et al., 2007).

Još jedan način dijeljenja vrijednosti atributa u zoni s površinom zone je slučaj kad je zona podijeljena na velik broj vrlo malih kvadrata. Primjerice, svaki kvadrat sadrži određeni broj ljudi koji žive u tom kvadratu. Ukupna vrijednost zone je jednostavno suma vrijednosti svih kvadrata. Ovaj oblik reprezentacije odgovara rasteru ili mrežnom GIS-u. Zbog toga rasterske reprezentacije obično ne trebaju normalizaciju na osnovu površina (*eng. area-based normalisation*) jer se već nalaze u prikladnom formatu. Proces generiranja mrežne reprezentacije može uključiti izraz vrijednosti za svaku ćeliju iz vektorskih podataka, primjerice korištenjem uniformnog modela (Smith et al., 2007).

Spomenuta opažanja naglašavaju nekoliko nedostataka (Smith et al., 2007):

- Proces dijeljenja mora imati način rješavanja slučajeva gdje je djelitelj 0;
- Treba uzeti u obzir podatke koji nedostaju;
- Ako su atribut koji se analizira i/ili karta već normalizirani, ne smiju se više normalizirati jer to uglavnom daje besmislene rezultate;

- Vrlo promjenjivi podaci (*eng. highly variable data*) mogu dovesti do zbunjivanja ih rezultata zbog nestabilnosti varijance (*eng. variance instability*). Do nje dolazi kada se određuju pojave na osnovu iznimno visokih frekvencija. Primjerice, jedna zona može imati 10000 djece od kojih je u 10 godina 7 oboljelo od leukemije, a druga zona može imati 500 djece od kojih je samo jedno oboljelo u istom vremenskom periodu. Normalizirana stopa prve zone iznosi 0,7, a druge 2 (na 1000 djece);
- Rezultat normalizacije mora biti smislen u pogledu zadatka i dostupnih podataka. Korištenje neprikladnog djeljitelja ili izvornih podataka neizvjesne vrijednosti ili važnosti može dovesti do neprihvatljivih rezultata;
- Određivanje prikladnog djeljitelja može biti problematično. Primjerice, ako se traže obrasci pojedine bolesti u određenoj godini, podaci popisa stanovništva nisu prikladan izvor informacija s obzirom na veličinu ugrožene populacije.

Jedinstveni omjeri pružaju standardnu formu normalizacije, ali kompleksnije kombinacije atributa (iz više skupova podataka) mogu se kombinirati na različite načine kako bi se stvorile indeksne vrijednosti. Prednost ovakvih postupaka je što se indeksne vrijednosti mogu koristiti kao vodič u planiranju razvoja ili kao usporedna mjera. Kako potreba za uključivanjem više faktora u izgradnju takvih indeksa postaje sve važnija, jednostavne težinske i omjerne procedure nisu dovoljne. Traže se sofisticiranije metode poput raznih vrsta analiza grupiranja (Smith et al., 2007).

Rezultate dobivene normalizacijom podataka prometnih nesreća po mjesnim odborima Grada Zagreba potrebno je vizualno prikazati. To se postiže izradom tematskih karata. Tematska karta prikazuje podatke grupirane u različite klase. Na karti se različite klase prikazuju različitim bojama ili različitim nijansama iste boje. Kod prezentacije prikupljene i analizirane vrijednosti tematskih podataka su često grupirane u klase koje pojednostavljuju čitanje karte. Glavne metode klasifikacije podataka su (URL 9):

- Metoda jednakih intervala
- Metoda standardnog odstupanja
- Metoda kvantila (jednakih frekvencija)
- Metoda prirodnih granica

U radu je korištena metoda kvantila. U toj je metodi potrebno unaprijed definirati broj klasa koji će se koristiti. Slijedi rangiranje klasa pridruživanjem jednakog broja opažanja svakoj klasi. Broj opažanja u svakoj klasi se računa po sljedećoj formuli (URL 9):

$$\text{broj opažanja po klasi} = \frac{\text{ukupni broj opažanja}}{\text{broj klasa}}$$

Ukoliko formula ne daje cjelobrojno rješenje, potrebno je smjestiti približno jednak broj opažanja u svaku klasu. Prednost ove metode je jednostavno raunanje klasa i svaka je klasa približno jednako zastupljena na konačnoj karti. Nedostatak ove metode su potencijalne praznine u podacima. Te praznine dovode do pretjeranog ponderiranja (*eng. over-weighting*) pojedinog odvojenog opažanja blizu granice klase (URL 9).

Broj korištenih klasa u radu je 5. Pri izradi svake tematske karte korištena je gustoća kao tip normalizacije. Primjerice, broj prometnih nesreća određene vrste unutar mjesnog odbora je podijeljen s površinom odbora. Obzirom da je korištena metoda kvantila, svaka karta ima različite vrijednosti granica klasa zbog različitog broja prometnih nesreća određene vrste. Na kartama su prometnice prikazane zelenom bojom zbog boljeg kontrasta i orijentacije. Mjesni odbori u kojima nije bilo nesreća s određenim atributom prikazani su sivom bojom. Ostali mjesni odbori su prikazani različitim nijansama crvene boje, od svijetlih nijansi (mala gustoća prometnih nesreća) do tamnih nijansi (velika gustoća).

### 3.6. Analiza podataka statistikom metodom Kernel Density Estimation

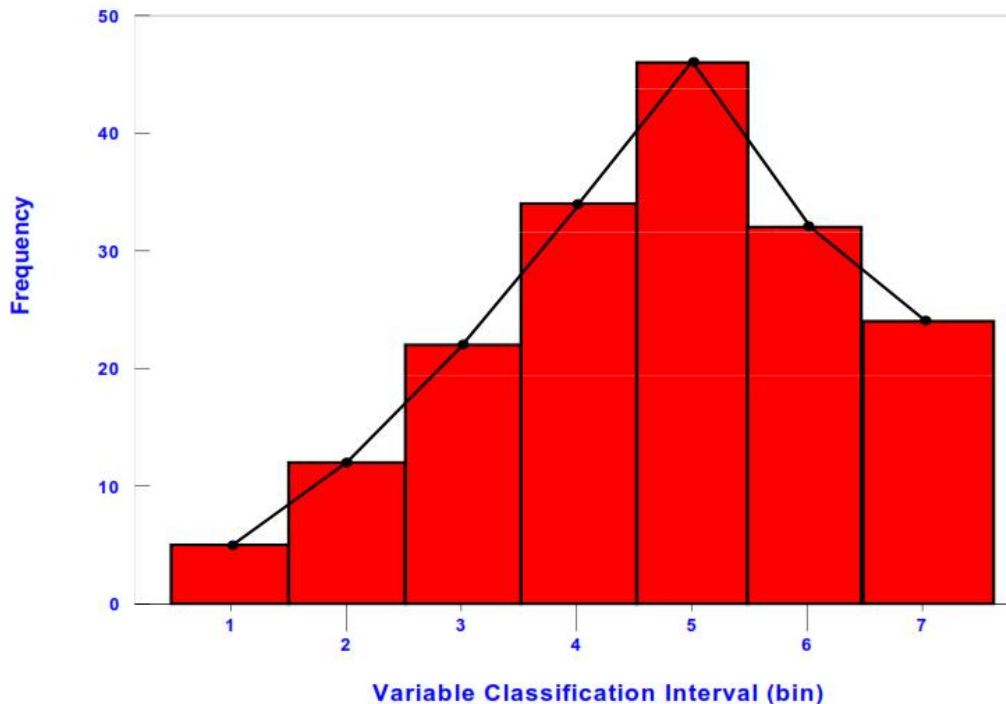
Prostorno opsežne varijable ne bi se smjele u normalnim uvjetima distribuirati direktno po zonama već standardizirati. Za varijable koje su podskup ukupnog skupa varijabli esto je poželjno izra unati i kartirati te podatke kao razmjer ukupnog broja, ali je za same ukupne skupove potrebno izra unati mjeru gusto e. Primjerice, dijeljenje ukupne populacije  $N$  s površinom zone  $A$ . To daje vrijednost gusto e  $N/A$  za svaku zonu, ali pretpostavlja se da je gusto a populacije konstantna kroz zonu i može se naglo promijeniti vrijednost na granicama zona. Što se ti e granica i površinske interpolacije, jednostavno ra unanje gusto e nije prikladno, ali ostaje u širokoj primjeni. Pretvaranje izvornog skupa podataka u rasterski model korištenjem neke vrste inteligentne interpolacije pruža zadovoljavaju e rezultate (Smith et al., 2007).

Ako skup podataka na raspolaganju nije temeljen na zonama već na to kama ili linijama, potrebne su alternativne metode odre ivanja gusto e. Najjednostavniji pristup je primjena podjele na zone i sumiranje broja opažanja unutar zone. Gusto a opažanja je jednostavno ukupan broj opažanja podijeljen s površinom zone na koju se odnose opažanja. Nedostatak ovog pristupa je što gusto e zna ajno variraju ovisno o odabranom obliku zone ili mreže te u veli ini i pridruženim atributima. Ukupni izra un gusto e za sve zone ima sli ne probleme, posebno u pogledu definiranja granica grada. Navedeni problemi uvelike otežavaju analizu podataka (Smith et al., 2007).

Alternativni pristup ra unanju gusto e dvodimenzionalnih skupova to kastih podataka se temelji na tehnikama zasnovanim na jednoparametarskoj (*eng. univariate*) statisti koj analizi (Smith et al., 2007). Obrada podataka je zasnovana na kreiranju žarišta prometnih nesre a odre enih karakteristika pomo u Kernel Density Estimationa (KDE). KDE podrazumijeva postavljanje simetri ne površine preko svake to ke i procjene udaljenosti izme u to ke i referentne lokacije zasnovane na matemati koj funkciji. Zatim slijedi sumiranje vrijednosti svih površina za tu referentnu lokaciju. Spomenuti postupak se ponavlja za uzastopne to ke [(Anderson, 2009), (Levine, 2013)].

Tehnika je razvijena kasnih 50-ih godina 20. stolje a kao alternativna metoda procjene gusto e histograma. Histogram je grafi ka prezentacija distribucije frekvencija. Kontinuirana varijabla se dijeli na intervale odre ene širine. Broj opažanja u svakom intervalu se zbraja i

prikazuje kao blok dijagram. Pretpostavlja se da histogram predstavlja glatku, osnovnu distribuciju (funkciju gusto  $e$ ). S ciljem procjene funkcije glatke gusto  $e$  iz histograma (*eng. smooth density function*), tradicionalno se susjedni intervali varijabli povezuju spajanjem središnjih to aka intervala nizom linija (Slika 5) (Levine, 2013).



Slika 5. Metoda spajanja središnjih to aka intervala (Levine, 2013)

Nažalost, ovaj postupak rezultira sljede im problemima (Levine, 2013):

- Informacije su odba ene zato što su sva opažanja unutar intervala pridružena središnjoj to ki. Gubitak informacija je proporcionalan širini intervala;
- Tehnika povezivanja središnjih to aka vodi diskontinuitetu funkcije gusto  $e$  koja nije glatka iako se pretpostavlja da je osnovna funkcija gusto  $e$  glatka. Jedno od mogu ih rješenja je smanjenje širine intervala. Na taj na in funkcija gusto  $e$  s manjim širinama intervala postaje gla a, iako još uvijek nije glatka. Postoje ograni enja u ovom pristupu jer pretjerano smanjenje širine intervala može dovesti do nepouzdanih rezultata;
- Tehnika je ovisna o proizvoljno definiranoj veli ini intervala. Pove anjem intervala procjena postaje grublja i obrnuto, smanjenjem intervala procjena postaje gla a. Pretpostavlja se da je osnovna distribucija gusto  $e$  glatka, kontinuirana i neovisna o širini intervala histograma.

### 3.6.1. Kernel Density Estimation (KDE)

Potrebno je dobiti procjenu gustoće vjerojatnosti (*eng. probability density estimate*) na osnovu dostupnog skupa podataka (*eng. sample*). Pretpostavimo da je nasumičan skup podataka  $(x_1, x_2, \dots, x_n)$  dobiven na osnovu nepoznate kontinuirane distribucije vjerojatnosti  $f(x)$ . Cilj je napraviti procjenu ove distribucije vjerojatnosti, koja se obilježava s  $\hat{f}(x)$ . Postoje brojni načini primjene ovakvih procjena. Primjerice, možemo biti zainteresirani za poznavanje oblika  $f(x)$ , je li bliže normalnoj ili bimodalnoj te koliko se razlikuje od ostalih distribucija. Takve primjene zadovoljava graf od  $\hat{f}(x)$ . S druge strane,  $\hat{f}(x)$  može biti iskorišten u procjeni vjerojatnosti. Primjerice, koliki se dio populacije nalazi ispod granice siromaštva ili kolika je vjerojatnost da će više od 10 cm padalina pasti u mjesec dana. Takve primjene zahtijevaju numeričke operacije nad  $\hat{f}(x)$ . Operacije mogu biti jednostavne, poput pronalaženja površine između dvije vrijednosti od  $X$ , ili mogu biti jako komplicirane. Bez obzira je li primjena grafička ili numerička, potreban je način procjene  $f$  za svaku vrijednost od  $X$  (Burt i Barber, 1996).

Postoje tradicionalne metode koje se mogu koristiti. Može se primijeniti parametarski pristup ako se pretpostavi da je  $X$  normalno distribuiran i koristi skup podataka da bi se procijenila srednja vrijednost  $\bar{x}$  (*eng. mean*) i standardno odstupanje  $s$  (*eng. standard deviation*). Budući da su ovo samo parametri od  $f(x)$ , ovo je jednako procjeni samog  $f(x)$ . Ako se koriste  $\bar{x}$  i  $s$  kao procjene, dobije se (Burt i Barber, 1996):

$$\hat{f}(x) = \frac{1}{s\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x - \bar{x}}{s}\right)^2\right]$$

Funkcija se može iskoristiti u cilju kreiranja grafa ili integrirati s ciljem pridruživanja vjerojatnosti. Iako je metoda jednostavno primjenjiva, problem kod ovog pristupa je što u slučaju da je  $f(x)$  daleko od normale, graf i pridružene vjerojatnosti mogu dovesti do krivih zaključaka (Burt i Barber, 1996).

Još jedna mogućnost je neparametarski pristup koji koristi skup podataka pri konstrukciji histograma relativnih frekvencija. Ako se svaka vrijednost histograma podijeli širinom intervala, dobije se procjena gustoće vjerojatnosti za svaki interval. Iako se ne postavljaju pretpostavke o formi  $f(x)$ , postoje proizvoljne odluke o broju i položaju intervala klasa. Ovisno o skupu podataka, male promjene mogu rezultirati velikim posljedicama u vizualnom

prikazu i izra unatim vjerojatnostima. Još jedan nedostatak je robusnost histograma gdje se  $\hat{f}(x)$  naglo mijenja na granicama klasa (Burt i Barber, 1996).

Kernel estimation je jednostavan pristup koji izbjegava prethodno spomenute nedostatke (Burt i Barber, 1996). Kernel je, u doslovnom smislu riječi, jezgra ili središnji dio predmeta. Isto se u ra unalnim znanostima odnosi na središnji dio operativnog sustava dok se u geoprostornim analizama odnosi na metode koje uključuju ra unanja pomoću dobro definiranog lokalnog susjedstva. Primjerice, modeliranje gustoće ili analiza lokalne mreže pomoću blokova elija ili radijalno simetrične funkcije (Smith et al., 2007).

Kernel Density Estimation (u daljnjem tekstu: KDE) je tehnika generalizacije lokacija opažanja na cijelo područje. Dok prostorna distribucija i statistika žarišta pružaju statistički pregled samih podataka opažanja, tehnike interpolacije generaliziraju podatke opažanja na itavo područje. Tehnike interpolacije pružaju procjene gustoće za sve dijelove promatranog područja. Procjena gustoće je varijabla intenziteta (*eng. intensity variable*),  $z$ - vrijednost, koja je procijenjena na određenoj lokaciji. Može se prikazati kao karta površina ili reljefna karta koja prikazuje intenzitet na svim lokacijama. KDE je tehnika interpolacije prikladna za pojedinačne točkaste lokacije (Levine, 2013).

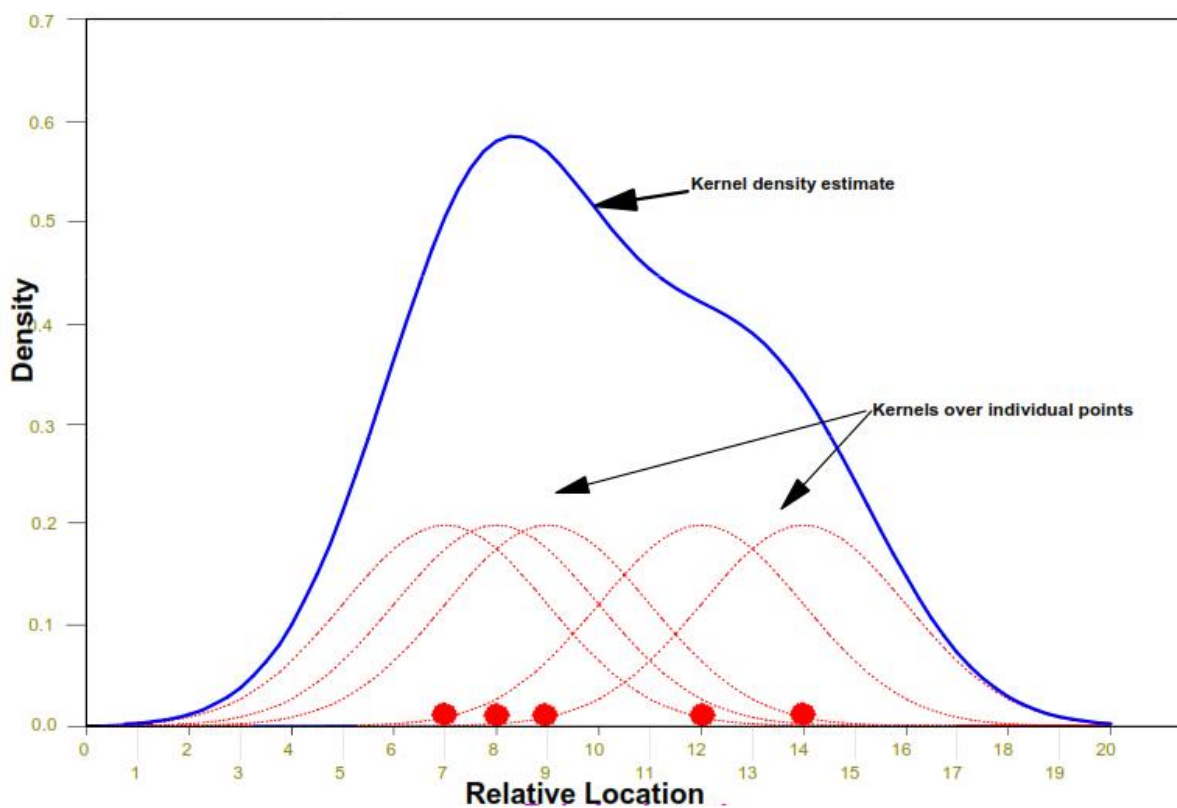
Metoda kernel estimation pruža glatke procjene bez potrebe za eksplicitnim pretpostavkama o formi  $f(x)$ . U slučaju histograma, svako opažanje se tretira kao točka. Ako točka leži u intervalu vjerojatnosti, cijela masa točke  $1/n$  se dodaje vjerojatnosti. S druge strane, kernel metode raspoređuju masu svakog opažanja oko opažane vrijednosti. Količina širenja je određena funkcijom pod nazivom kernel (Burt i Barber, 1996).

Kernel funkcija je često korištena kao normalna distribucija, odnosno kao zvonolika krivulja, koja u svakom smjeru teži u beskonačnost, ali ima konačnu (jediničnu) površinu ispod krivulje (Smith et al., 2007). U većini slučajeva kernel funkcija je simetrična tako da je točka jednako proširena prema višim i nižim vrijednostima (Burt i Barber, 1996). Oblik kernel funkcije nema velik utjecaj na skup vrijednosti gustoće koji je dodijeljen linearnom segmentu ili površini u 2D aplikacijama. Puno značajniji utjecaj ima izbor parametra širenja, odnosno širine pojasa (*eng. bandwidth*) (Smith et al., 2007).

Da bi se procijenila gustoća vjerojatnosti, dodaju se vrijednosti kernela za svaku točku seta podataka (Burt i Barber, 1996). Slika 6 prikazuje navedeni proces s pet točaka, odnosno na pet lokacija. Iznad svake lokacije postavlja se simetrična kernel funkcija. Pod simetričnom

funkcijom podrazumijeva se funkcija koja opada jednakom stopom u svim smjerovima proporcionalno udaljenosti od svake točke. U ovom je slučaju riječ o normalnoj distribuciji, ali se mogu koristiti i ostali tipovi simetričnih distribucija. Osnovna distribucija gustoće se procjenjuje sumiranjem individualnih kernel funkcija na svim lokacijama kako bi se napravila glatka cjelokupna funkcija gustoće (*eng. smooth cumulative density function*). Treba primijetiti da su funkcije sumirane na svakoj točki duž ljestvice, a ne samo na lokacijama točaka. Prednosti ovog pristupa su sljedeće (Levine, 2013):

- Svaka točka jednako doprinosi površini gustoće;
- Konačna funkcija gustoće je kontinuirana na svim točkama duž ljestvice



Slika 6. Sumiranje normalne kernel funkcije za pet točaka (Levine, 2013)

Da bi se operacija matematički izrazila, neka  $K$  predstavlja kernel funkciju,  $h$  stupanj širenja, odnosno širinu pojasa. S tim oznakama, izraz za KDE glasi (Burt i Barber, 1996):

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right)$$

Gledajući prvo argument  $K$ , vidi se da je udaljenost između  $x$  i svake vrijednosti skupa podataka izražena kao višekratnik od  $h$ . Budući da je  $h$  iste mjerne jedinice kao i  $x$ , omjer je



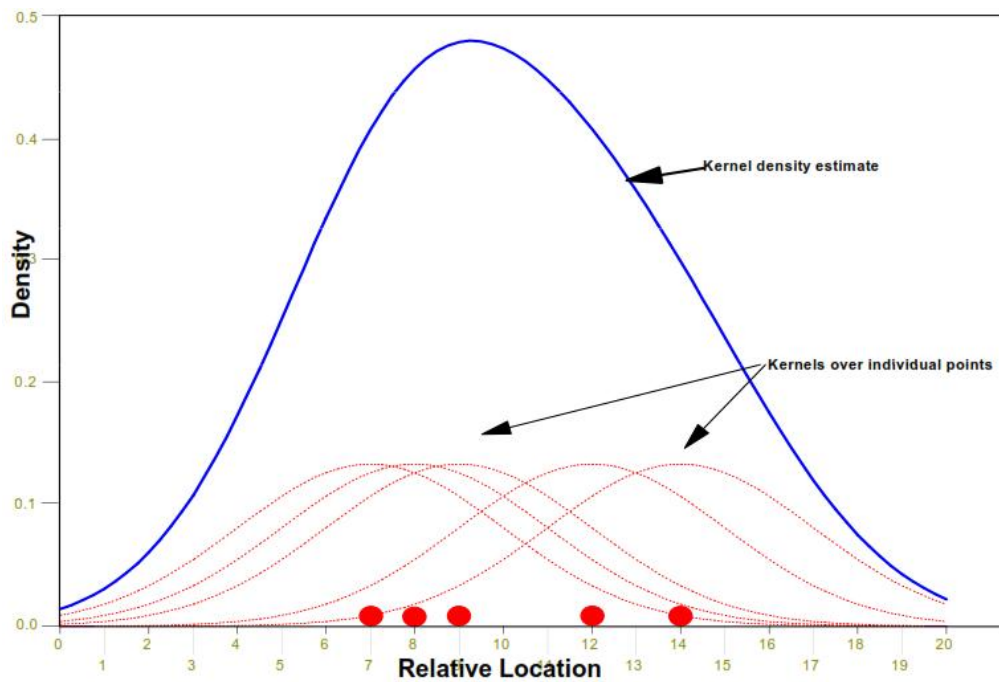
bezdimezionalan. Ova se skalirana udaljenost koristi u kernel funkciji kako bi se pronašao doprinos svakog opažanja. S  $K$  kao opadajućom funkcijom udaljenosti, doprinos točke koja se nalazi na velikoj udaljenosti od  $x$  će biti mali, dok će bliže točke imati veći doprinos. Suma svih doprinosa podijeljena s  $nh$  postaje procjena gustoće. Dijeljenje s  $h$  je nužno kako bi  $\hat{f}(x)$  bila gustoća s dimenzijama vjerojatnosti po jedinici  $X$  (Burt i Barber, 1996).

Općenito, ako funkcija zadovoljava sljedeće pretpostavke može se koristiti kao kernel (Guidoum, 2013):

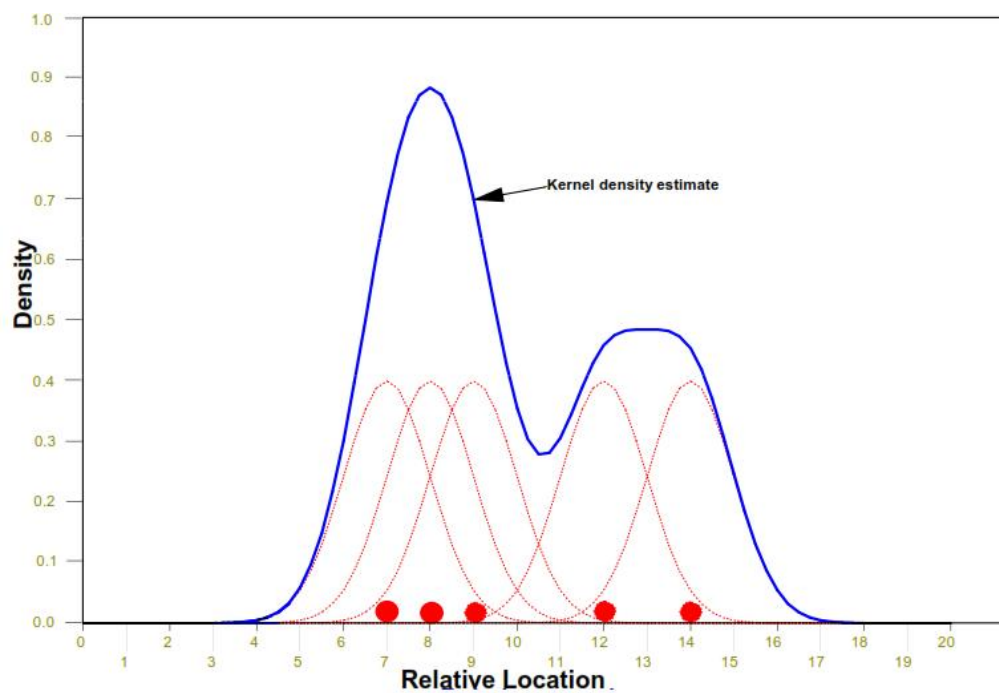
- $K(x) \geq 0$  i  $\int_{\mathbb{R}} K(x)dx = 1$ ;
- Simetrična oko ishodišta  $\int_{\mathbb{R}} xK(x)dx = 0$ ;
- Konačan drugi moment  $\mu_2(K) = \int_{\mathbb{R}} x^2K(x)dx < \infty$  s oznakom  $R(K) = \int_{\mathbb{R}} (K(x))^2 dx$ .

Može se donijeti nekoliko općenitih zaključaka o ovom pristupu. S jedne strane,  $\hat{f}(x)$  nasljeđuje mnoga svojstva od kernel funkcije. Ako je  $K$  funkcija vjerojatnosti,  $\hat{f}(x)$  će također biti funkcija vjerojatnosti (u potpunosti nenegativna, s ukupnom površinom jednakom jedinici). Ako je  $K$  glatka i  $\hat{f}(x)$  će biti glatka funkcija. S druge strane, razina detaljnosti predstavljena u  $\hat{f}(x)$  varira direktno u ovisnosti o širini pojasa  $h$ . Velika vrijednost vodi velikom proširenju svake točke skupa podataka. To rezultira širokom i glatkom funkcijom gustoće bez fine strukture (Burt i Barber, 1996).

Širina kernela je prilagodljiva. Uski kerneli tijesno koncentriraju masu oko središnjih (opažanih) vrijednosti, a široki kerneli rezultiraju većim širenjem (Burt i Barber, 1996). Manja vrijednost širine područja vodi do gušće mreže procjene gustoće s mnoštvom 'vrhova i dolina' (Slika 8). S druge strane, velika vrijednost širine područja vodi do glatke distribucije i manje varijabilnosti između područja (Slika 7). Dok manje vrijednosti širine područja prikazuju veću razliku između područja, treba uzeti u obzir statističku preciznost procjene. Ako veličina uzorka nije prevelika, manja širina područja vodi do nepreciznijih procjena, a 'vrhovi i doline' mogu prikazivati samo slučajne varijacije. Ako je na raspolaganju već uzorak, može se napraviti bolja procjena gustoće. Općenito, dobra je praksa isprobavati različite fiksne vrijednosti da bi se vidjelo koji rezultati imaju smisla (Levine, 2013).



Slika 7. Funkcija gustoće s velikom širinom pojasa (Levine, 2013)



Slika 8. Funkcija gustoće s malom širinom pojasa (Levine, 2013)

Simetrija je atribut koji se ne nasljeđuje iz kernela. Procjena gustoće zahtijeva  $n$  procjena od  $K$ . Uzevši u obzir da velik broj procjena može biti potreban pri stvaranju grafa ili izraza una vjerojatnosti, ova metoda otkriveno nije prikladna za ručno rađivanje (Burt i Barber, 1996).

Ne postoje stroga pravila kod izbora kernel funkcije, iako se obično bira simetri na distribucija vjerojatnosti. Unutar obitelji simetričnih kernela, izbor ovisi o računovim zahtjevima i uglađenosti. Triangular kernel je vjerojatno najjednostavniji izbor (Burt i Barber, 1996):

$$K(z) = \begin{cases} 1 - |z|, & |z| \leq 1 \\ 0, & |z| > 1 \end{cases}$$

gdje je

$$z = \left( \frac{x - \bar{x}}{h} \right)$$

Ovaj je kernel samo trokut jednu jedinicu visok, dvije jedinice širok i ima područje jedinstva (*eng. area of unity*). Njegova glavna prednost leži u jednostavnosti izraza, samo su potrebne operacije oduzimanja i dijeljenja. Nema potrebe za razmatranjem točaka aka skupa podataka koje su više od  $h$  jedinica udaljene od bilo kojeg  $x$ . Ne pruža glatku procjenu poput ostalih kernela. Iako je kernel kontinuiran, ima prekid u nagibima na centru i na rubovima koji prenose diskontinuitete nagiba u  $\hat{f}$ . Prekidi se mogu reducirati korištenjem funkcije višeg reda u  $z$ . Primjerice, polinomom drugog stupnja izbjeći će se diskontinuitet središnjeg nagiba (Burt i Barber, 1996):

$$K(z) = \begin{cases} \frac{3}{4}(1 - z^2), & |z| \leq 1 \\ 0, & |z| > 1 \end{cases}$$

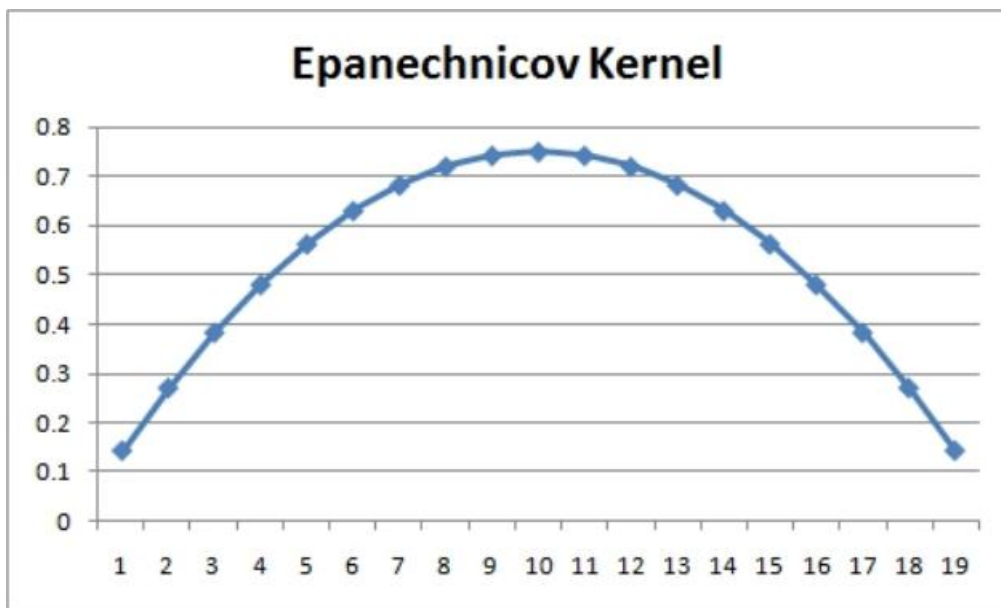
Sada  $K$  varira glatko preko svoje središnje točke i  $\hat{f}$  je glatki. Još jednom  $K$  postaje udaljenost koja teži nuli daleko od opažanja, tako da se samo obližnja opažanja trebaju razmatrati. Međutim, postoje diskontinuiteti nagiba na rubovima ovog kernela. Još glatke procjene stvaraju se korištenjem normalnog (Gaussovog) kernela (Burt i Barber, 1996):

$$K(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right)$$

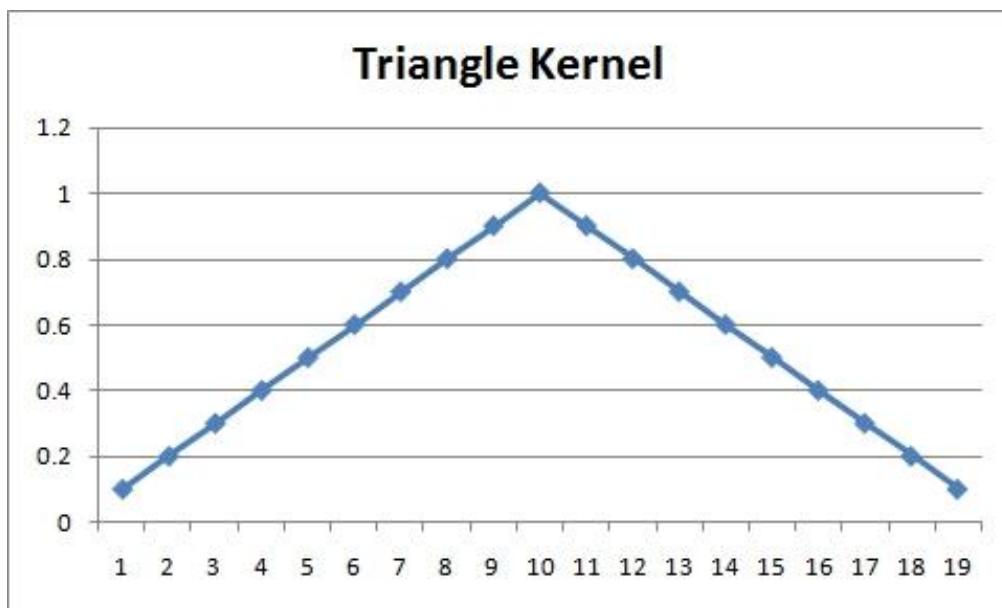
Ovo rezultira vrlo glatkom funkcijom gustoće, ali na trošak više aritmetike. Ovdje je dodan teret eksponencijala i  $K$  je tehnički svugdje ne-nula vrijednost. Budući da se normalna krivulja brzo približava nuli, opažanje udaljeno više od nekoliko višekratnika od  $h$  ne može puno doprinijeti  $\hat{f}$ . Stoga se računovanja mogu ograničiti na obližnje točke (Burt i Barber, 1996).

Tablica 4. Kernel funkcije primijenjene u obradi prometnih nesre a KDE metodom  
( $r$  je najve a derivacija kernela) (Guidoum, 2013)

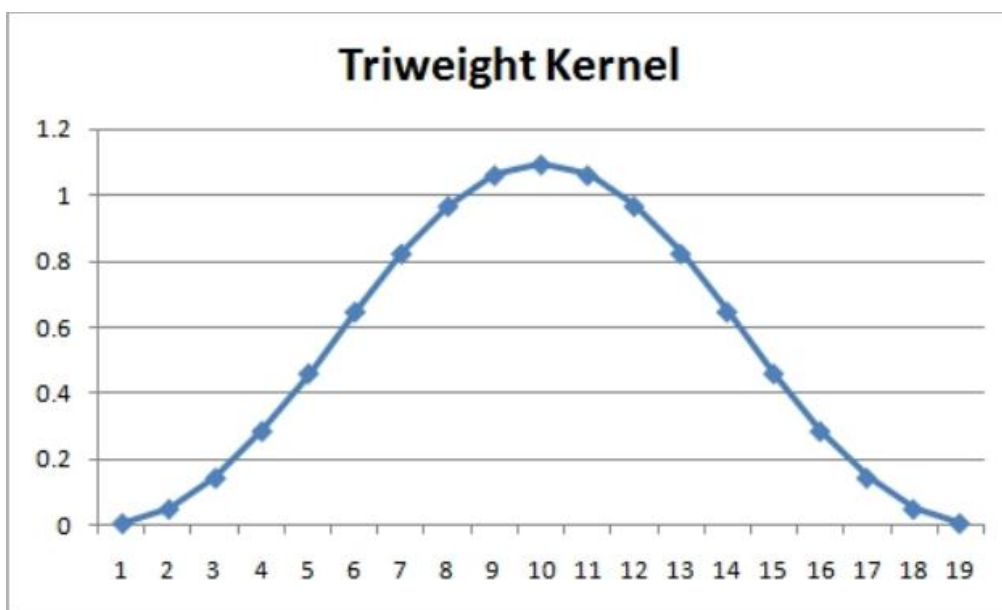
| Kernel                      | $K(z;r)$                                               | $R(K)$  | $\mu_2(K)$ |
|-----------------------------|--------------------------------------------------------|---------|------------|
| Epanechnikov<br>(Quadratic) | $K(z; 2) = \frac{3}{4} (1 - z^2) 1_{( z  \leq 1)}$     | 3/5     | 1/5        |
| Uniform                     | $K(z; 0) = \frac{1}{2} 1_{( z  \leq 1)}$               | 1/2     | 1/3        |
| Triangular                  | $K(z; 1) = (1 -  z ) 1_{( z  \leq 1)}$                 | 2/3     | 1/6        |
| Triweight                   | $K(z; 6) = \frac{35}{32} (1 - z^2)^3 1_{( z  \leq 1)}$ | 350/429 | 1/9        |
| Biweight (Quartic)          | $K(z; 4) = \frac{15}{16} (1 - z^2)^2 1_{( z  \leq 1)}$ | 5/7     | 1/7        |



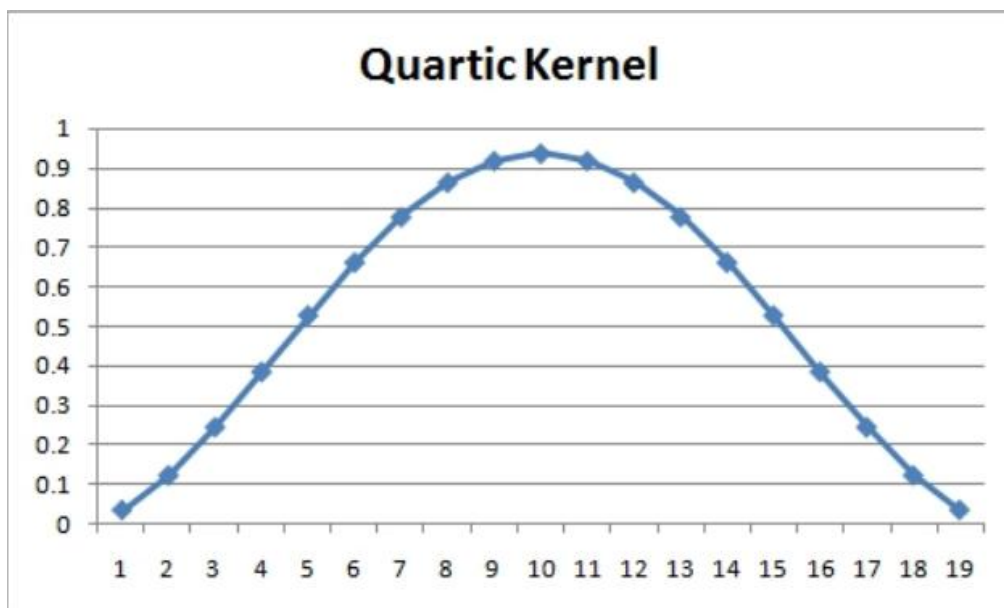
Slika 9. Graf Epanechnikov (Quadratic) kernel funkcije (URL 10)



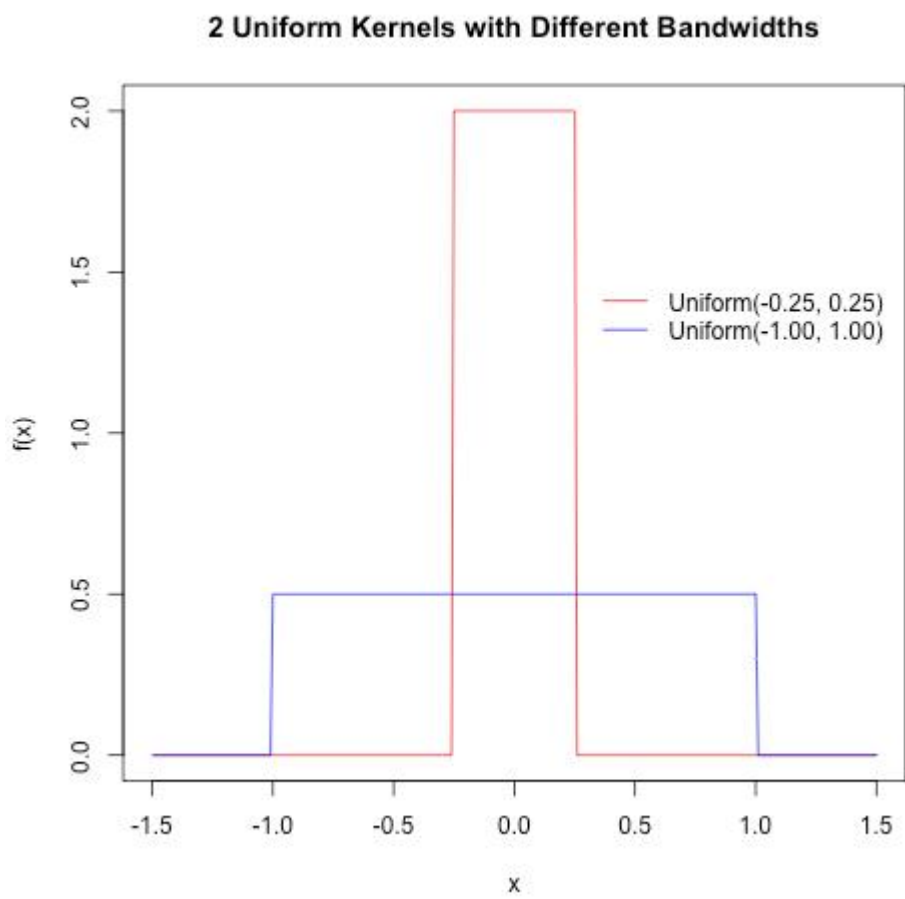
Slika 10. Graf Triangular kernel funkcije (URL 10)



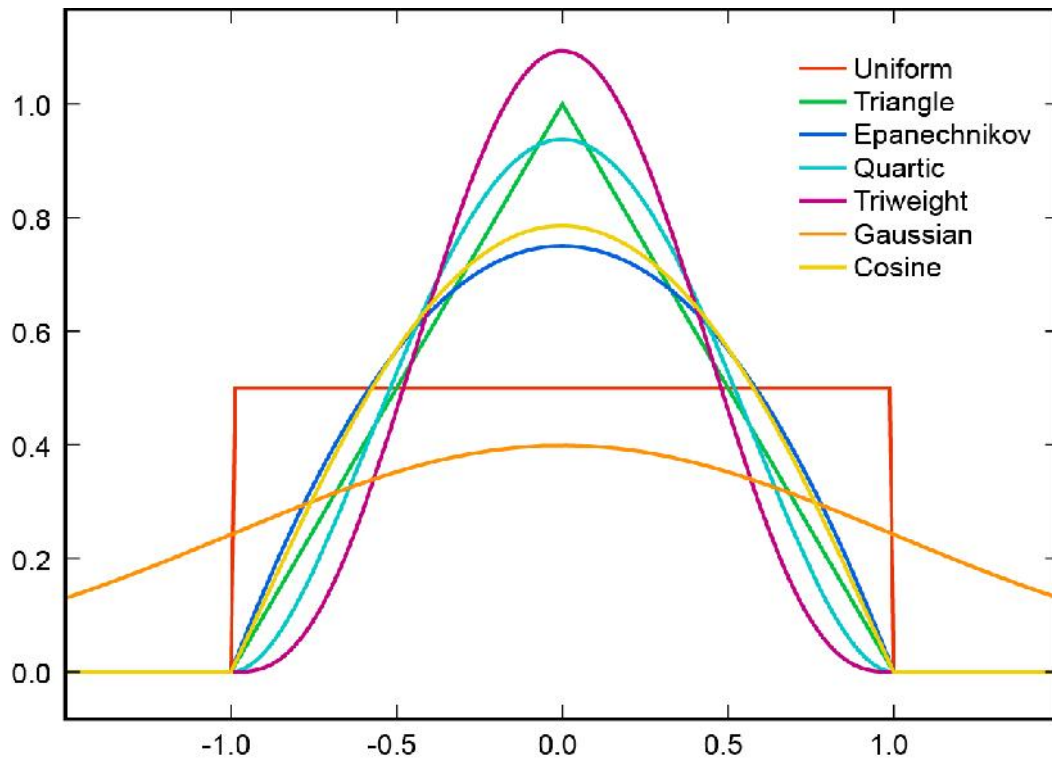
Slika 11. Graf Triweight kernel funkcije (URL 10)



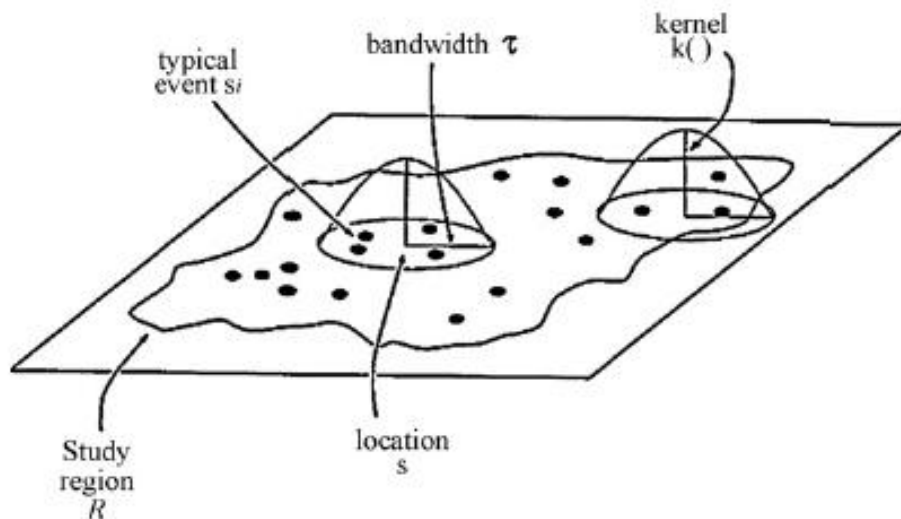
Slika 12. Graf Biweight (Quartic) kernel funkcije (URL 10)



Slika 13. Grafovi Uniform kernel funkcija razli itih širina pojasa (URL 11)



Slika 14. Usporedba grafova svih kernel funkcija (URL 12)



Slika 15. Dijagram koji prikazuje kako funkcionira metoda Quadratic Kernel Density (Anderson, 2009)

Izbor kernel funkcije ovisi o tome koliko korisnik želi naglasiti bliske točke u odnosu na daleke. Korištenje kernel funkcije koja ima veliku težinsku razliku između bliskih i udaljenih točaka (npr. triangular kernel) rezultira finijim varijacijama unutar površine, nego što je to

služaj kod funkcija kod kojih su manje razlike u težinama (npr. quartic kernel). Potonje funkcije nastoje izgladiti distribuciju (Levine, 2013).

U radu je testirano svih pet navedenih metoda. Utvrđeno je da dvije metode, *quartic* (*biweight*) i *triweight*, daju najprihvatljivije rezultate. Procjena gustoće vjerojatnosti kod obje metode prikazana je različitim bojama, od svijetlih nijansi žute (manja gustoća) do tamnih nijansi crvene (veća gustoća, središnji dio žarišta). Vizualna značajka quartic metode je ravnomjerna raspodjela gustoće vjerojatnosti oko središnje točke žarišta. Približno jednakom širinom prikazana je svaka nijansa boje. Kod triweight metode su središnji dio žarišta i njegova neposredna okolina jače istaknuti, ali s puno manjom širinom. Triweight metoda je vrlo korisna kada se želi naglasiti linearno protezanje opažanih veličina.

### 3.6.2. Određivanje širine pojasa $h$

Nakon odabira kernel funkcije, ostaje pitanje širine pojasa,  $h$ . Prevelika vrijednost sprječava isticanje varijacija sitnih razmjera iz  $f$  u  $\hat{f}$ , a premala vrijednost daje procjene koje su jako varijabilne. U tom slučaju treba eksperimentirati s velikim rasponom vrijednosti širine pojasa i uspoređivati rezultirajuće procjene gustoće jednu s drugom. Prvi problem je odabrati razumnu vrijednost od  $h$  za eksperimentiranje. Intuicija predlaže da ukoliko je skup podataka jako raspršen u  $X$ , potrebna je velika širina pojasa. Budući da veliki uzorci sadrže više informacija, širina pojasa se treba smanjivati povećanjem broja opažanja  $n$ , dok sve ostalo ostaje nepromijenjeno. Kao prvi korak u odabiru optimalnog  $h$ , potrebno je izračunati standardno odstupanje uzorka  $s$  i raspon interkvartila  $R$ . Širina pojasa je manji iznos od sljedećih vrijednosti (Burt i Barber, 1996):

$$h_1 = 0.90 \frac{s}{\sqrt[5]{n}}$$

$$h_2 = 0.67 \frac{R}{\sqrt[5]{n}}$$

Disperzija je puno važnija od veličine uzorka. Ako se uzme da je  $h$  mjera koliko se informacija može izvući iz uzorka, potreban je jako velik  $n$  za kompenzaciju široko razmaknutih točaka skupa podataka. Gornje jednadžbe pružaju jednostavan način izbora  $h$  za svaki skup podataka i daju vrijednosti koje odgovaraju velikom broju funkcija gustoće. Međutim, ovise samo o skupnom statističkom uzorku ( $s$ ,  $R$ ), a ne detaljima rasporeda točaka (Burt i Barber, 1996).



Kod fiksne širine područja, korisnik određuje interval koji će se koristiti i mjerne jedinice. Ovisno o tipu korištene kernel funkcije, interval ima različite znaenke. Kod funkcije normalnog kernela širina područja je standardna devijacija normalne distribucije. Kod funkcije quartic i triangular kernela širina područja je radijus područja pretraživanja u kojem se vrši interpolacija (Levine, 2013).

Odabir širine područja je više umjetnost nego znanost, ali se može podvrgnuti formalnoj analizi i procjeni. Primjerice, primjenom KDE procedura nad skupovima podataka s poznatim gustoćama. Alternativa fiksnom odabiru širine područja je adaptivni odabir gdje korisnik određuje kriterij odabira. Korisnik može definirati broj točaka opažanja (*eng. event point*), koje će se uključiti u krug centriran na svakoj opažanoj točki i uzeti radijus tog kruga kao širinu područja oko te točke (Smith et al., 2007). Naučniji pristupi, moguće su dva poboljšanja koja povećavaju iskoristivost informacija iz uzorka (Burt i Barber, 1996).

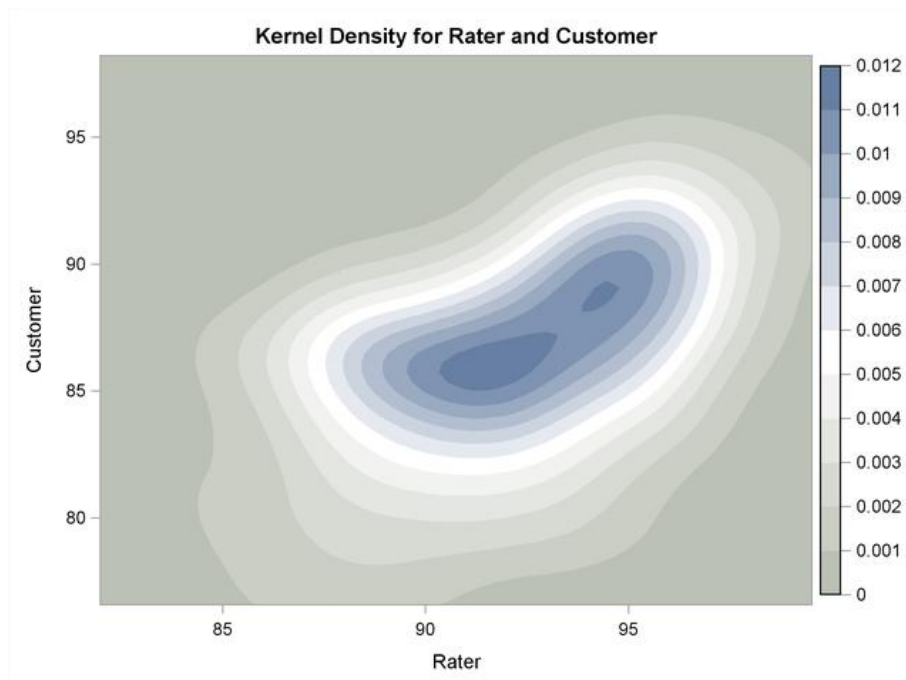
U prvom slučaju, moguće je optimirati izbor širine pojasa odabiranjem vrijednosti koja umanjuje pretpostavljenu pogrešku u  $\hat{f}$ . Tehnika se zove *cross-validation* i koristi samo podatke u identifikaciji optimalnog  $h$ . Osnovna ideja je da je dobar izbor za  $h$  dati dobru procjenu za  $\hat{f}$ . Cross-validation koristi dostupne podatke pri procjeni stupnja kvalitete određenog  $h$ . Odabire se vrijednost koja daje najniži procijenjeni broj pogrešaka u  $\hat{f}$ . (Burt i Barber, 1996).

Drugo poboljšanje je upotreba *adaptivnog* kernela. U ovom slučaju širina pojasa smije varirati s  $X$ . Male vrijednosti se koriste na mjestima gdje su podaci zgusnuti, a velike vrijednosti gdje su opažanja raspršena. Na ovaj se način kernel prilagođava lokalnim varijacijama gustoće, pružaju i visoku rezoluciju bez stvaranja lažnih svojstava u regijama od  $X$  niske gustoće. Pristup zahtijeva dva prolaska kroz podatke gdje prvi koristi fiksni kernel i daje početne procjene gustoće. Ove početne procjene koriste se u određivanju jedinstvene širine pojasa za svako opažanje. Iako zahtjevne, adaptivne procjene su posebno korisne kada postoje velike varijacije gustoće u skupu podataka (Burt i Barber, 1996).

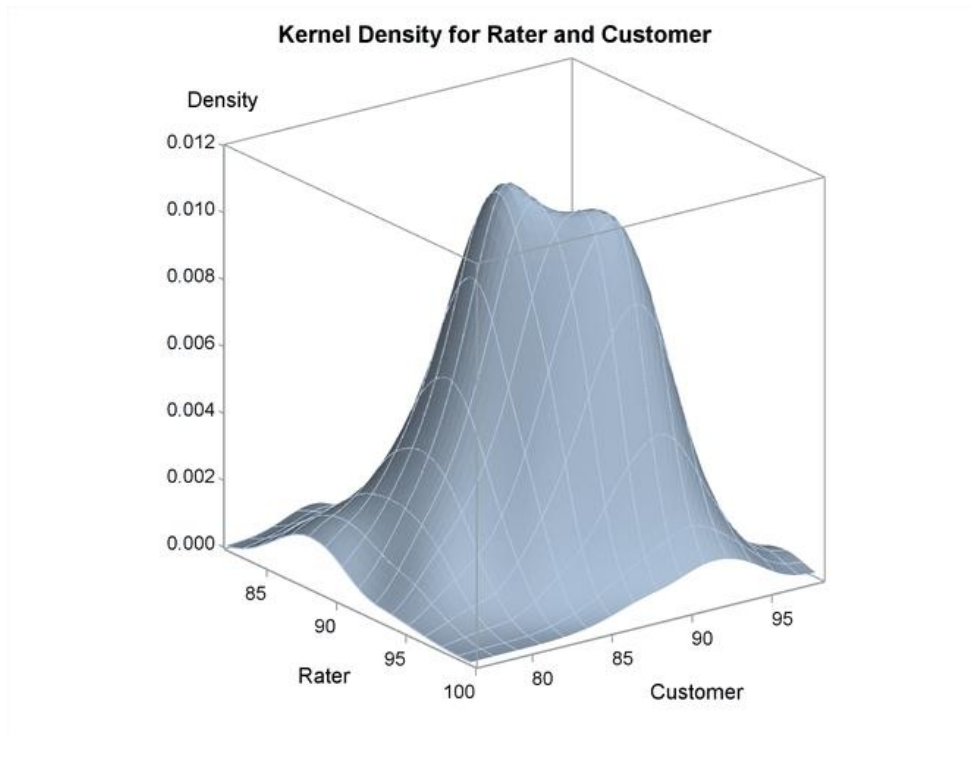
### 3.6.3. KDE u 2D i 3D

Dosad su opisivane kernel metode u procjeni funkcije  $\hat{f}(X)$  gdje opažanja leže na koordinatnoj osi. Domena od  $f$  je očitto jednodimenzionalna, a lokacija unutar domene je određena jednom koordinatom (Burt i Barber, 1996). Prenos procesa u dvodimenzionalni skup podataka se svodi na jednostavno preuzimanje jednoparametarskih procedura i

dodavanje druge dimenzije, odnosno rotacija funkcije oko svake točke. Primjerice, ako bi se koristila normalna distribucija kao funkcija izglanjanja, iznad svake točke bi se nalazila dvodimenzionalna zvonolika krivulja. Postavlja se kernel funkcija iznad svake točke u promatranom području i računa se vrijednost koju je doprinijela ta točka na nacrtanoj mreži. Ni u jednom trenutku rezolucija mreže ne utječe na oblik rezultirajuće površine. Ukoliko je moguće, treba biti postavljena na razumnu vrijednost s obzirom na kontekst skupa podataka koji se analizira. To uključuje poznate prostorne pogreške, primijenjena zaokruživanja i stvaranje dozvole za isključenje određenih područja iz računanja. Vrijednosti svih točaka na presjeku svake mreže ili za svaku ćeliju mreže se računaju i dodaju jedna drugoj kako bi dale složenu površinu gustoće (*eng. composite density surface*). Ona može biti iscrtana u 2D (npr. konture gustoće) ili kao 3D ploha (Smith et al., 2007).

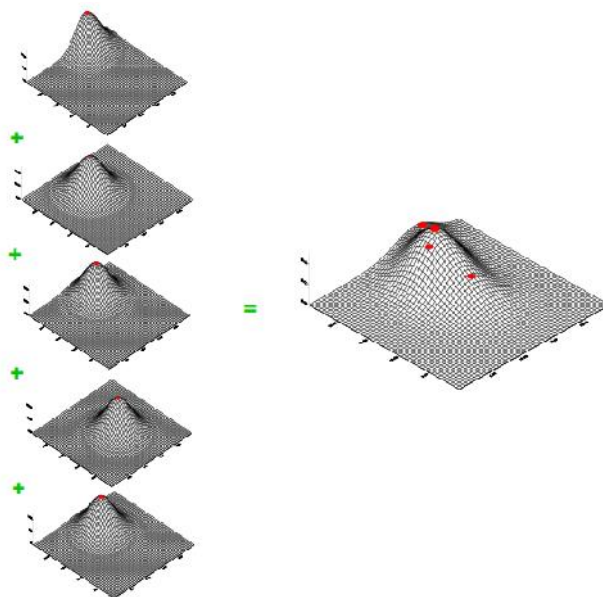


Slika 16. Vizualizacija 2D plohe KDE-a (URL 13)



Slika 17. Vizualizacija 3D plohe KDE-a (URL 13)

Slika 18 prikazuje trodimenzionalnu normalnu distribuciju položenu preko svih pet prethodno navedenih to aka i rezultiraju u površinu gusto e dobivenu kao sumu svih pet individualnih površina. Ova je metoda posebno prikladna za geografske podatke. Može povezati dvije ili više varijabli primjenom procjene kernela za svaku varijablu posebno i zatim dijeljenjem jedne s drugom kako bi se napravila trodimenzionalna procjena utjecaja (Levine, 2013).



Slika 18. Sumiranje normalnih kernel površina za pet to aka (Levine, 2013)

Vrijednosti rezultiraju e mreže mogu biti dobivene kao (Smith et al., 2007):

- *Relativne gusto e* koje daju vrijednosti opažanja po jedinici površine. One se prilago avaju veli ini mreže (broj opažanja/m<sup>2</sup> ili broj opažanja/ha). To je zadana ili jedina opcija u mnogim GIS aplikacijama;
- *Apsolutne gusto e* koje daju vrijednosti broja opažanja po eliji mreže. Ne mogu se prilago avati veli inom elije. Suma vrijednosti svih elija mora biti jednaka broju opažanja u analizi;
- *Vjerojatnosti* koje daju vrijednosti broja opažanja podijeljenog s ukupnim brojem opažanja.

Budu i da su ova tri tipa ra unanja u direktnom me uodnosu, kona na se mreža ne e razlikovati u varijabilnosti. Izbor ovisi o ra unanjima koja mogu biti iskorištena za procjenu apsolutnih gusto a, relativnih gusto a ili vjerojatnosti. Za usporedbu razli itih vrsta opažanja s razli itim vremenskim periodima esto se koriste apsolutne gusto e. Da bi se odredila vjerojatnost ponavljanja odre ene vrste opažanja na nekoj lokaciji, ispis rezultata kao vjerojatnosti ima više smisla (Levine, 2013).

Koli ina doprinosa kernela pojedine opažane to ke vrijednosti mreže ovisi o (Smith et al., 2007):

- *Tipu kernel funkcije* (ome ena ili neome ena);
- *Parametru k* kojeg može definirati korisnik ili može biti automatski odre en;
- *Širini podru ja h* koje se odabire (ve a širina podru ja širi utjecaj opažane to ke na ve u udaljenost);
- *Težinama* koje se mogu dodijeliti pojedinim to kama, što zna i da neke to ke imaju ve i doprinos od drugih.

Slijedi generalizacija kernel metode u svrhu primjene iste nad skupom dvoparametarskih podataka. Domena je dvodimenzionalna i traži se dvoparametarska funkcija  $\hat{f}(X, Y)$  u kojoj to ke skupa podataka leže na ravnini (Burt i Barber, 1996).

Na raspolaganju je skup to aka u prostoru i potrebno je odrediti varijacije gusto e to aka na traženom podru ju. Kernel metode su privla ne iz razloga što daju procjene na svakoj lokaciji. Ako se koristi glatki kernel, kona na površina gusto e e biti glatka. Neka je  $d_i$  udaljenost izme u to ke  $i$  i neke lokacije  $(x, y)$  (Burt i Barber, 1996):

$$d_i = \sqrt{(x - x_i)^2 + (y - y_i)^2}$$

Kao i u prethodnom slučaju, potrebno je odabrati širinu pojasa kernela  $h$  i koristiti udaljenosti skalirane za  $h$  u kernelu. Primjerice, normalni kernel je (Burt i Barber, 1996):

$$K\left(\frac{d_i}{h}\right) = \frac{1}{2\pi} \exp\left[-\frac{1}{2}\left(\frac{d_i}{h}\right)^2\right]$$

S ovim obilježavanjem, procjena gustoće glasi (Burt i Barber, 1996):

$$\hat{g}(x, y) = \frac{1}{h^2} \sum_{i=1}^n K\left(\frac{d_i}{h}\right)$$

Treba primijetiti dijeljenje s  $h^2$  u gornjoj jednadžbi. Budući da je  $h$  izražen u metri kim jedinicama, procjena gustoće biti izražena u broju točaka po jedinici površine. Primjerice, broj točaka po kilometru kvadratnom ili broj točaka po kvadratnoj milji. Gornja jednadžba može biti primijenjena za bilo koju  $x, y$  lokaciju, što znači da metoda pruža kontinuiranu površinu gustoće (Burt i Barber, 1996).

Koristi se  $g$  umjesto  $f$  kako bi se naznačilo da jednadžba daje gustoću točaka, a ne gustoću vjerojatnosti. Gustoća točaka je prikladna kada se traže informacije o prostornim varijacijama u gustoći opažanja. Primjerice, mogu se opažati epicentri potresa i potrebno je kartirati gustoću epicentara. U tom slučaju iscrtava funkcija gustoće  $g$ . S druge strane, mogu se kartirati klimatske varijable na temelju vrijednosti opažanih na mreži meteoroloških stanica. Gustoća stanica igra ključnu ulogu u određivanju moći i odlučivanju mreže. Potrebna je karta gustoće stanica kako bi se mogla procijeniti učinkovitost mreže, a kernel metoda može dati tu procjenu (Burt i Barber, 1996).

Jedan od mogućih problema je procjena dvoparametarske funkcije gustoće vjerojatnosti. Domena je još uvijek dvodimenzionalna, ali ovog puta  $x$  i  $y$  nisu prostorne varijable. Tražena funkcija  $\hat{f}(x, y)$  je površina koja daje gustoću vjerojatnosti, a volumen ispod površine je vjerojatnost. Da bi se gustoća točaka pretvorila u gustoću vjerojatnosti, potrebno je samo dijeljenje s brojem točaka skupa podataka (Burt i Barber, 1996):

$$\hat{f}(x, y) = \frac{1}{n} \hat{g}(x, y) = \frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{d_i}{h}\right)$$

Dok je god kernel funkcija distribucija vjerojatnosti, procjena  $\hat{f}$  će biti distribucija vjerojatnosti. Ako  $x$  i  $y$  imaju različite jedinice, potrebno ih je standardizirati kako bi izražavanje udaljenosti imao smisla. Prirodan izbor je izražavanje obje vrijednosti kao  $z$ -vrijednost kodira unanja  $d$ . Na taj se način  $h$  interpretira kao širina kernela u jedinicama standardnog odstupanja umjesto u mjernim jedinicama poput kilometara i milja. U svim ostalim slučajevima dvoparameterska metoda je direktno proširenje jednodimenzionalne tehnike (Burt i Barber, 1996).

U radu je korištena širina pojasa od 200 metara. Vrijednost je izabrana zbog karakteristika korištenih kernela i zadovoljavaju ih rezultata provedenih analiza. Površina analiziranog područja Grada Zagreba iznosi oko 600 km<sup>2</sup>, što je približno jednako protezanju od 30x20 km. Na taj način vrijednost od 200 metara iznosi oko 1% vrijednosti linearnog protezanja analiziranog područja. Uzevši u obzir prostornu distribuciju prometnih nesreća, vrijednost od 200 metara pokazala se prikladnom za identifikaciju žarišta.

### 3.6.4. Primjene KDE-a

Postoje brojne prednosti KDE-a u odnosu na jednostavne statističke metode određivanja žarišta i tehnike grupiranja poput K-meansa. Glavna prednost KDE-a je određivanje širenja rizika prometne nesreće. Širenje rizika može se definirati kao površina oko definiranog grupiranja (*eng. cluster*) u kojem postoji visoka vjerojatnost ponavljanja prometne nesreće na osnovu prostorne zavisnosti. Korištenjem ove metode definira se proizvoljna prostorna jedinica za analizu koja je homogena za cijelo područje. To omogućuje usporedbu i taksonomiju (Anderson, 2009).

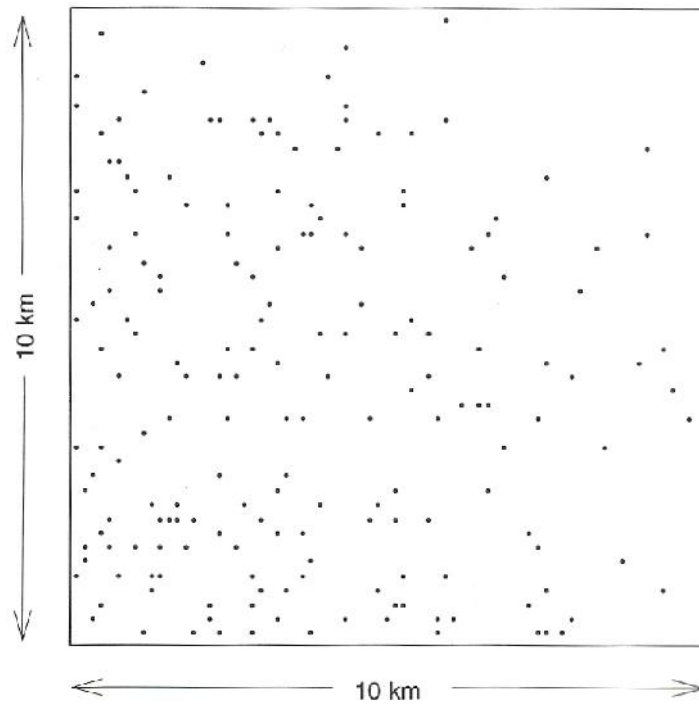
KDE metode imaju razne primjene (Smith et al., 2007):

- Izglavljavanje skupa točkastih podataka;
- Kreiranje kontinuiranih površina iz skupa točkastih podataka. Cilj je preklapanje kontinuiranih površina s drugim skupovima podataka koji su u kontinuiranom/rasterskom obliku;
- Procjena distribucije vjerojatnosti (*eng. probability distribution estimation*);
- Interpolacija i otkrivanje žarišta;
- Vizualizacija i analiza vremenskih obrazaca.

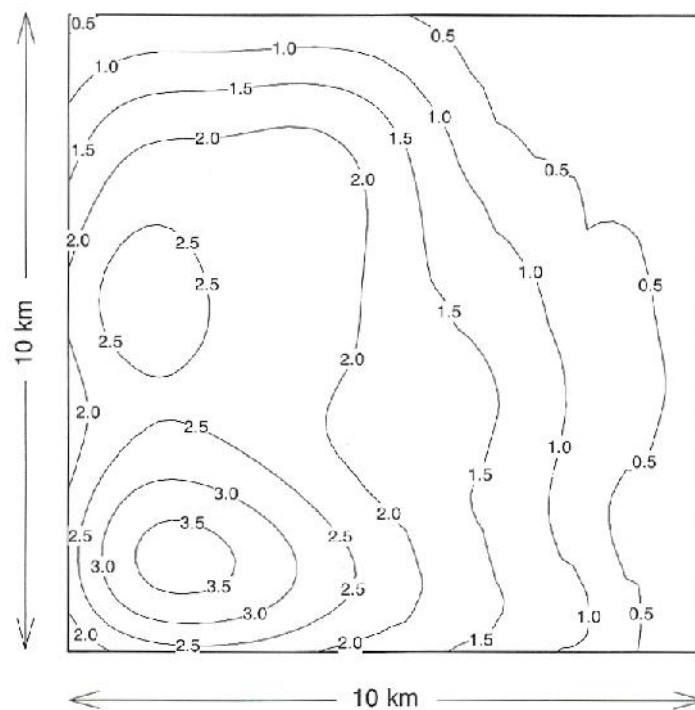
KDE se također može primijeniti nad više od jednog skupa točkastih podataka. Rezultat takve dvostruke analize skupova podataka je često omjer primarnog i sekundarnog skupa. Cilj je obrada primarnih obrazaca s uklonjenim ili smanjenim pozadinskim efektima. Pritom treba paziti da rezultati nisu pod utjecajem distorzije vrlo malih vrijednosti ili vrijednosti blizu nule u drugoj površini gustoće. Korištenje kernel funkcija kao oblika za dodjeljivanje težina točkama omogućuje stvaranje lokalnih težinskih srednjih vrijednosti i varijanci (Smith et al., 2007).

U nastavku je opisan način primjene KDE metode koji su opisali autori Burt i Barber (Burt i Barber, 1996). Iako je riječ o nasumičnom skupu točkastih podataka, postupak je identičan obradi prometnih nesreća. Slika 19 prikazuje uzorak od 200 točaka unutar područja od 100 km<sup>2</sup>. Vizualnim pregledom utvrđeno je da sjeveroistok ima manju, a jugozapad relativno veću gustoću točaka. Da bi potvrdili pretpostavku, autori traže numeričku informaciju o varijacijama gustoće uzorka. Iako se može podijeliti regiju na ćelije i brojati točke unutar svake ćelije, taj pristup uključuje sve nedostatke povezane s histogramima. Traži se metoda koja daje izglednu procjenu gustoće i stoga se odabire dvoparameterska kernel metoda. Odabire se širina pojasa  $h$  u iznosu od 1 km ili 10% raspona  $x$  i  $y$  koordinata. To ograničava razinu detaljnosti na konačnu površinu gustoće. Kernel funkcije se polažu preko svih 200 točaka i sumiraju na svim lokacijama. Slika 20 predstavlja reljefnu kartu (*eng. contour map*) koja prikazuje rezultat iscrtanih vrijednosti. Najveća je gustoća na jugozapadu, ali se može uočiti drugi maksimum sjeverno od primarnog maksimuma. Ako bi se oslonili samo na točke skupa podataka (Slika 19), vjerojatno se ne bi uočio drugi maksimum. S 200 točaka na području od 100 km<sup>2</sup>, srednja gustoća je 2 točke/km<sup>2</sup>. Na karti je vidljivo da su ispodprosječne gustoće na istočnim i najsjevernijim dijelovima promatranog područja. Može se uočiti nagli pad gustoće na južnim i zapadnim rubovima karte. Riječ je o rubnom efektu uzrokovanom neuzimanjem uzoraka izvan granice karte. Da su uzeti uzorci izvan karte, procjene blizu rubova bile bi veće. Budući da normalna krivulja naglo pada, jako je malen doprinos točkama koje su na većoj udaljenosti od  $2h$ . Bilo bi poželjno imati točke u pojasu širine  $2h$  oko promatranog područja, ali ne više od toga. Također se rubni efekti ne šire više od  $2h$  u područje promatranja. Ne vidi se rubni efekt duž istočne granice zato što su gustoće toliko niske da točke koje nedostaju nemaju veliki doprinos. Vidi se da kontura od 0.5 ima male izbojine manje od širine pojasa od 0.1 km. Te su izbojine rezultat metode, odnosno indikacija da je  $h$  premala vrijednost za regije male gustoće. Može se koristiti veća širina pojasa, ali u tom se

slučajne gube detalji i područjima visoke gustoće. Bolji izbor je adaptivni kernel koji se automatski prilagođava varijacijama gustoće.



Slika 19. Uzorak od 200 točaka (Burt i Barber, 1996)



Slika 20. Obrisi gustoće skupa podataka iz slike 19 (Burt i Barber, 1996)



### 3.7. Vizualizacija rezultata

Rezultati analiza su vizualizirani u tri GIS aplikacije. Rezultati normalizacije i KDE metode su vizualizirani u QGIS 2.0.1 Dufour aplikaciji. QGIS je GIS aplikacija otvorenog koda (*eng. open source*) za upravljanje prostornim vektorskim i rasterskim podacima. Projekt je započet u svibnju 2002. godine s ciljem kreiranja aplikacije koja će biti pouzdana alternativa u odnosu na komercijalne aplikacije sličnog tipa. QGIS je multiplatformska aplikacija koju je moguće instalirati na većinu Linux, Windows, OS X i Android operativnih sustava. Razvijen je pomoću Qt toolkit-a i C++-a. Svrha QGIS-a je pružiti jednostavan GIS softver s funkcijama i opcijama koje se najčešće koriste u praksi (URL 14). Kao podloga za prikaz rezultata dobivenih primjenom metode KDE korištena je platforma za razvoj web kartata OpenLayers. OpenLayers je JavaScript biblioteka za prikaz kartografskih podataka u većini suvremenih web preglednika. Implementira JavaScript API za razvoj geografskih aplikacija za internet. Sličan je Google Maps API-u, ali je slobodan i razvijan od strane zajednice (URL 15)

Prikazan je primjer vizualizacije rezultata normalizacije KDE-a u 3D. U tu su svrhu korištene dvije aplikacije, SAGA GIS i ILWIS. SAGA GIS je GIS aplikacija otvorenog koda za automatizirane geoznanstvene analize. Razvijen je za jednostavnu i uinkovitu implementaciju prostornih algoritama. Nudi opsežan i rastući skup geoznanstvenih metoda. Pruža pristup na grafičke slike i mnoge mogućnosti vizualizacije. Radi na Linux i Windows operativnim sustavima. Napisana je u C++ programskom jeziku i ima objektno orijentiran dizajn sustava. Aplikacija je razvijena na Zavodu za fizikalnu geografiju Sveučilišta u Gottingenu (URL 16).

ILWIS je GIS aplikacija otvorenog koda razvijena u Nizozemskoj od strane Instituta za aerometriju i geoznanosti. Kombinira operacije nad rasterskim, vektorskim i tematskim podacima u jednu opsežnu aplikaciju namijenjenu desktop korisnicima. Pristupan je korisnicima i dokazano pouzdan. Potpuno integrira GDAL, knjižnicu koja pruža brojne rasterske formate. Omogućuje pristup PostgreSQL i GIS bazama podataka i kartama (URL 17).

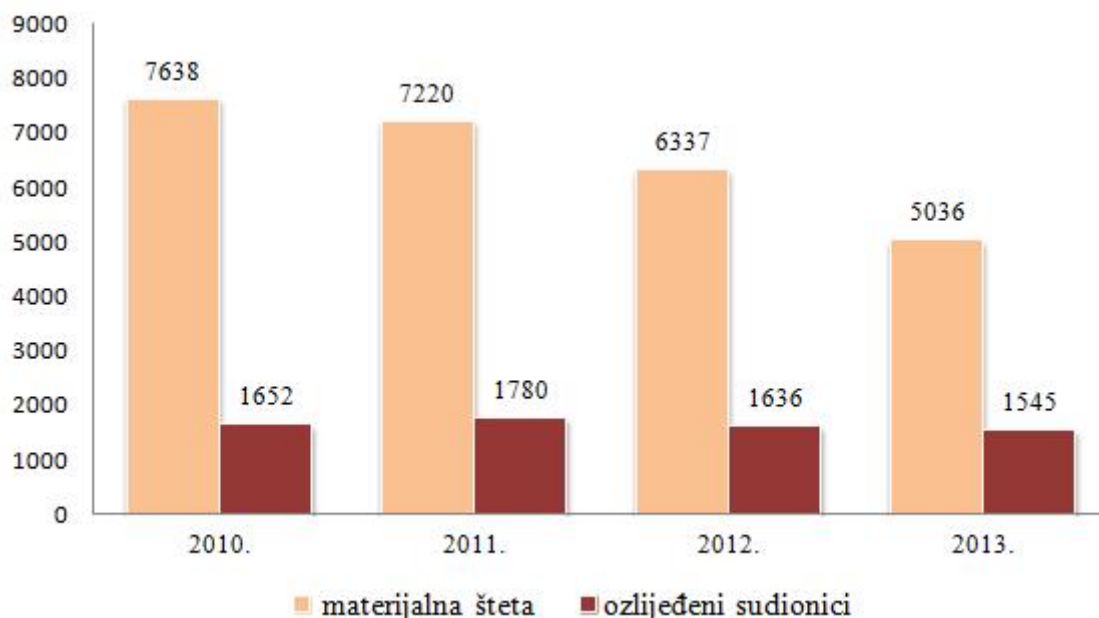
## 4. Rezultati

Podaci o prometnim nesrećama su prikupljeni od strane MUP-a za područje grada Zagreba u razdoblju od 2010. do 2013. godine. Svaka prometna nesreća ima prostorne koordinate i atribute koji opisuju okolnosti u kojoj su se dogodile. Cilj je utvrditi postoji li korelacija između određenih tipova prometnih nesreća i pojedinih lokacija.

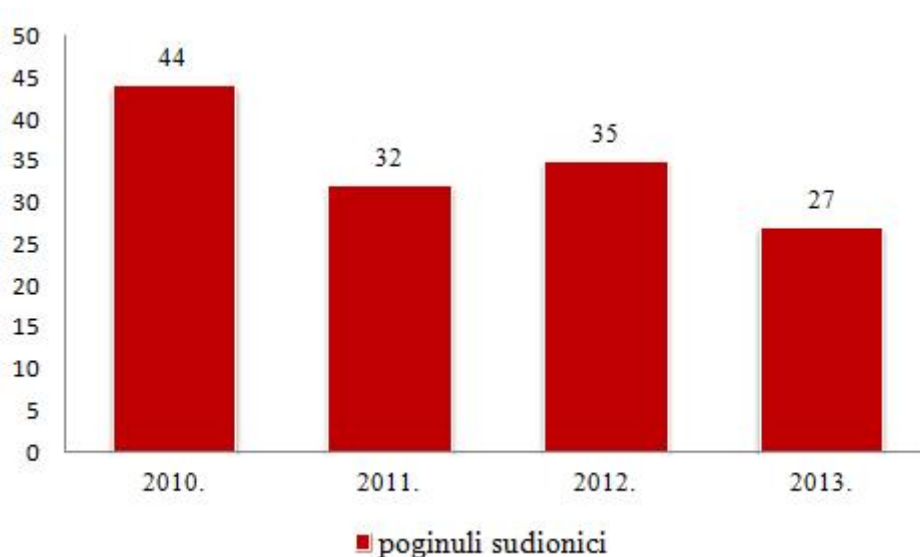
### 4.1. Osnovni podaci za obradu

Tablica 5. Broj prometnih nesreća po godinama s obzirom na posljedicu nesreće

| Prometne nesreće          | 2010. (%)    | 2011. (%)    | 2012. (%)    | 2013. (%)    | Ukupno (%)    |
|---------------------------|--------------|--------------|--------------|--------------|---------------|
| S materijalnom štetom     | 7638 (81.83) | 7220 (79.94) | 6337 (79.13) | 5036 (76.21) | 26231 (79.53) |
| S ozlijeđenim sudionicima | 1652 (17.70) | 1780 (19.71) | 1636 (20.43) | 1545 (23.38) | 6613 (20.05)  |
| S poginulim sudionicima   | 44 (0.47)    | 32 (0.35)    | 35 (0.44)    | 27 (0.41)    | 138 (0.42)    |
| Ukupno                    | 9334         | 9032         | 8008         | 6608         | 32982         |



Slika 21. Grafika usporedba broja prometnih nesreća s materijalnom štetom i ozlijeđenim sudionicima po godinama s obzirom na posljedicu nesreće



Slika 22. Grafi ka usporedba broja prometnih nesre a s poginulim sudionicima po godinama

Tablica 5 prikazuje broj odre enih posljedica prometnih nesre a po godinama. Rije je o nesre ama s materijalnom štetom, s ozlije enim sudionicima i s poginulim sudionicima. Nesre e s materijalnom štetom ine 80%, s ozlije enim sudionicima 20%, a nesre e s poginulim sudionicima ine oko 0.40% ukupnog broja prometnih nesre a. Uo eno je zna ajno smanjenje broja prometnih nesre a kroz etiri godine. To se prvenstveno odnosi na nesre e s materijalnom štetom. Broj nesre a s ozlije enim sudionicima se tako er smanjio, ali se njihov udio u ukupnom broju nesre a pove ao.

Slika 21 prikazuje odnos nesre a s materijalnom štetom i s ozlije enim sudionicima kroz etiri godine. Slika 22 prikazuje promjenu broja prometnih nesre a s poginulim sudionicima kroz etiri godine odvojeno od ostalih vrsta zato što je broj ovih nesre a puno manji od prethodnih. Grafovi su uvedeni iz razloga što bolje vizualiziraju tabli ne podatke.

Tablica 6. Broj prometnih nesreća po okolnostima koje su im prethodile i karakteristikama ceste

| Okolnosti prometne nesreće        | Križanje          |                      |                    | Ravni cestovni potez |                      |                    | Zavoj             |                      |                    | Ukupno            |                      |                    |
|-----------------------------------|-------------------|----------------------|--------------------|----------------------|----------------------|--------------------|-------------------|----------------------|--------------------|-------------------|----------------------|--------------------|
|                                   | Materijalna šteta | Ozlijeđeni sudionici | Poginuli sudionici | Materijalna šteta    | Ozlijeđeni sudionici | Poginuli sudionici | Materijalna šteta | Ozlijeđeni sudionici | Poginuli sudionici | Materijalna šteta | Ozlijeđeni sudionici | Poginuli sudionici |
| Brzina neprimjerena uvjetima      | 1325              | 620                  | 13                 | 1447                 | 696                  | 20                 | 724               | 358                  | 4                  | 3496              | 1674                 | 37                 |
| Nepoštivanje prednosti prolaska   | 2879              | 935                  | 7                  | 110                  | 64                   | 3                  | 4                 | 2                    | 0                  | 2993              | 1001                 | 10                 |
| Nepoštivanje svjetlosnog znaka    | 638               | 391                  | 8                  | 6                    | 5                    | 0                  | 2                 | 0                    | 0                  | 646               | 396                  | 8                  |
| Nepropisne radnje vozača          | 2779              | 620                  | 8                  | 4410                 | 666                  | 18                 | 260               | 64                   | 1                  | 7449              | 1350                 | 27                 |
| Vožnja na nedovoljnoj udaljenosti | 1229              | 277                  | 0                  | 1453                 | 433                  | 5                  | 24                | 5                    | 0                  | 2706              | 715                  | 5                  |
| Ostale pogreške i propusti vozača | 1181              | 329                  | 5                  | 3316                 | 377                  | 16                 | 215               | 23                   | 0                  | 4712              | 729                  | 21                 |
| Pogreške i propusti pješaka       | 29                | 175                  | 6                  | 41                   | 191                  | 13                 | 2                 | 5                    | 1                  | 72                | 371                  | 20                 |
| <b>Ukupno</b>                     | <b>10060</b>      | <b>3347</b>          | <b>47</b>          | <b>10783</b>         | <b>2432</b>          | <b>75</b>          | <b>1231</b>       | <b>457</b>           | <b>6</b>           | <b>22074</b>      | <b>6236</b>          | <b>128</b>         |

Tablica 7. Broj prometnih nesreća po vrstama nesreća i karakteristikama ceste

| Vrsta prometne nesreće                 | Križanje          |                      |                    | Ravni cestovni potez |                      |                    | Zavoj             |                      |                    | Ukupno            |                      |                    |
|----------------------------------------|-------------------|----------------------|--------------------|----------------------|----------------------|--------------------|-------------------|----------------------|--------------------|-------------------|----------------------|--------------------|
|                                        | Materijalna šteta | Ozlije eni sudionici | Poginuli sudionici | Materijalna šteta    | Ozlije eni sudionici | Poginuli sudionici | Materijalna šteta | Ozlije eni sudionici | Poginuli sudionici | Materijalna šteta | Ozlije eni sudionici | Poginuli sudionici |
| Me usobni sudar vozila                 | 8693              | 1874                 | 12                 | 8089                 | 1233                 | 25                 | 640               | 156                  | 3                  | 17422             | 3263                 | 40                 |
| Slijetanje vozila s ceste              | 352               | 112                  | 5                  | 627                  | 289                  | 12                 | 386               | 211                  | 2                  | 1365              | 612                  | 19                 |
| Nalet na biciklista                    | 125               | 249                  | 2                  | 27                   | 73                   | 1                  | 2                 | 6                    | 0                  | 154               | 328                  | 3                  |
| Nalet na pješaka                       | 63                | 758                  | 18                 | 55                   | 471                  | 28                 | 2                 | 20                   | 1                  | 120               | 1249                 | 47                 |
| Nalet na motociklista                  | 91                | 119                  | 4                  | 48                   | 64                   | 2                  | 4                 | 7                    | 0                  | 143               | 190                  | 6                  |
| Nalet na životinju                     | 19                | 1                    | 0                  | 245                  | 2                    | 0                  | 7                 | 0                    | 0                  | 271               | 3                    | 0                  |
| Udar vozila u objekt na ili kraj ceste | 494               | 45                   | 5                  | 1033                 | 80                   | 5                  | 161               | 38                   | 0                  | 1688              | 163                  | 10                 |
| Ostale vrste                           | 223               | 189                  | 1                  | 659                  | 220                  | 2                  | 29                | 19                   | 0                  | 911               | 428                  | 3                  |
| <i>Ukupno</i>                          | 10060             | 3347                 | 47                 | 10783                | 2432                 | 75                 | 1231              | 457                  | 6                  | 22074             | 6236                 | 128                |

Tablica 6 prikazuje odnos izme u okolnosti prometnih nesre a i karakteristika cesta na kojima su se dogodile. Okolnosti prometne nesre e uklju uju neprimjerena ponašanja sudionika u prometu. Karakteristike cesta uklju uju križanja, ravne cestovne poteze, zavoje i ostali oblici. Ostali oblici nisu analizirani jer uklju uju oblike poput mostova, pješa kih staza i prijelaza preko pruge. Rije je o mjestima na kojima se dogodio mali broj nesre a. Najve i broj nesre a s poginulim sudionicima se dogodio na ravnim cestovnim potezima, naj eš e zbog brzine neprimjerene uvjetima i nepropisnih radnji voza a. Najve i broj nesre a s ozlije enim sudionicima se dogodio na križanjima zbog nepoštivanja prednosti prolaska i nepropisnih radnji voza a. Nesre e s materijalnom štetom su podjednako zastupljene na križanjima i ravnim cestovnim potezima. Kod križanja je rije o nepoštivanju prednosti prolaska, a kod ravnih cestovnih poteza o nepropisnim radnjama voza a. Na zavojima dogodilo malo nesre a, oko 5% ukupnog broja nesre a. Najve i broj se dogodio zbog brzine neprimjerene uvjetima. Tablica daje pregled koji su dijelovi cesta bili opasniji u pojedinim okolnostima.

Tablica 7 prikazuje odnos vrsta prometnih nesre a i karakteristika cesta na kojima su se dogodile. Vrsta prometne nesre e uklju uje me usobni sudar vozila, nalet na pješaka, biciklista i sli no. Najve i broj nesre a s poginulim sudionicima dogodio se na ravnim cestovnim potezima. Na križanjima se dogodio najve i broj nesre a s ozlije enim sudionicima. U oba su slu aja naj eš e vrste prometnih nesre a nalet na pješaka i me usobni sudar vozila. Tablica daje pregled koli ine pojedinih vrsta prometnih nesre a na razli itim dijelovima cesta i može pomo i pri donošenju odluka o detaljnijoj analizi karakteristi nih lokacija.

## 4.2. Rezultati obrade hi-kvadrat testovima

Tablica 8. Distribucija prometnih nesreća po posljedicama i karakteristikama ceste

( $P = 0.000$ , hi-kvadrat = 200.66, broj stupnjeva slobode = 4, razina značajnosti = 5%)

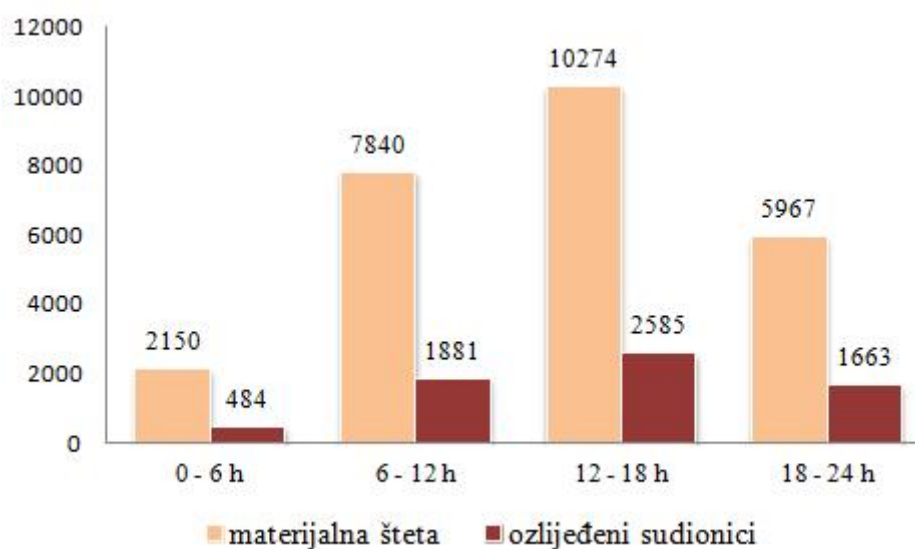
| Prometne nesreće          | Križanje (%)  | Ravni cestovni potez (%) | Zavoj (%)    | Ukupno (%)    |
|---------------------------|---------------|--------------------------|--------------|---------------|
| S materijalnom štetom     | 10060 (74.77) | 10783 (81.14)            | 1231 (72.67) | 22074 (77.62) |
| S ozlijeđenim sudionicima | 3347 (24.88)  | 2432 (18.30)             | 457 (26.98)  | 6236 (21.93)  |
| S poginulim sudionicima   | 47 (0.35)     | 75 (0.56)                | 6 (0.35)     | 128 (0.45)    |
| <b>UKUPNO</b>             | <b>13454</b>  | <b>13290</b>             | <b>1694</b>  | <b>28438</b>  |

Tablica 8 prikazuje distribuciju posljedica prometnih nesreća s obzirom na karakteristike cesta. S ciljem lakše interpretacije rezultata, apsolutnim su vrijednostima u tablici kontingencije pridruženi pripadaju i postotci. Nesreće s materijalnom štetom češće se događaju na ravnim cestovnim potezima, a nesreće s ozlijeđenim sudionicima su puno češće na zavojima. Najveći broj prometnih nesreća s poginulim sudionicima dogodio se na ravnim cestovnim potezima. Provedbom hi-kvadrat testa homogenosti, odbijena je nul-hipoteza. Dakle, broj prometnih nesreća s različitim posljedicama nije zastupljen u jednakim omjerima na različitim dijelovima cesta.

Tablica 9. Distribucija prometnih nesreća po posljedicama i dijelovima dana

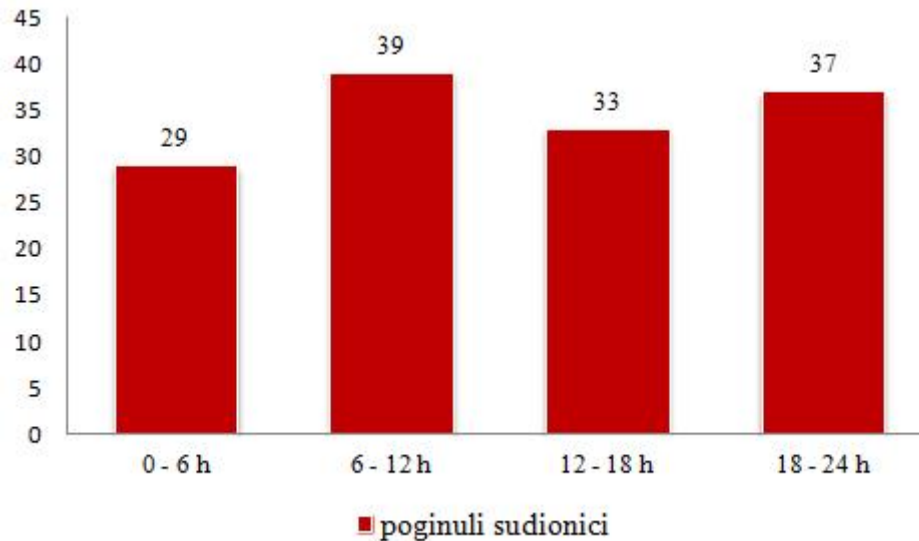
( $P=0.000$ , hi-kvadrat = 59.60, broj stupnjeva slobode = 6, razina značajnosti = 5%)

| Prometne nesreće          | 0–6 h (%)    | 6–12 h (%)   | 12–18 h (%)   | 18–24 h (%)  | Ukupno (%)    |
|---------------------------|--------------|--------------|---------------|--------------|---------------|
| S materijalnom štetom     | 2150 (80.74) | 7840 (80.33) | 10274 (79.69) | 5967 (77.83) | 26231 (79.53) |
| S ozlijeđenim sudionicima | 484 (18.17)  | 1881 (19.27) | 2585 (20.05)  | 1663 (21.69) | 6613 (20.05)  |
| S poginulim sudionicima   | 29 (1.09)    | 39 (0.40)    | 33 (0.26)     | 37 (0.48)    | 138 (0.42)    |
| <b>UKUPNO</b>             | <b>2663</b>  | <b>9760</b>  | <b>12892</b>  | <b>7667</b>  | <b>32982</b>  |



Slika 23 Grafički prikaz broja prometnih nesreća s materijalnom štetom i ozlijeđenim sudionicima po dijelovima dana





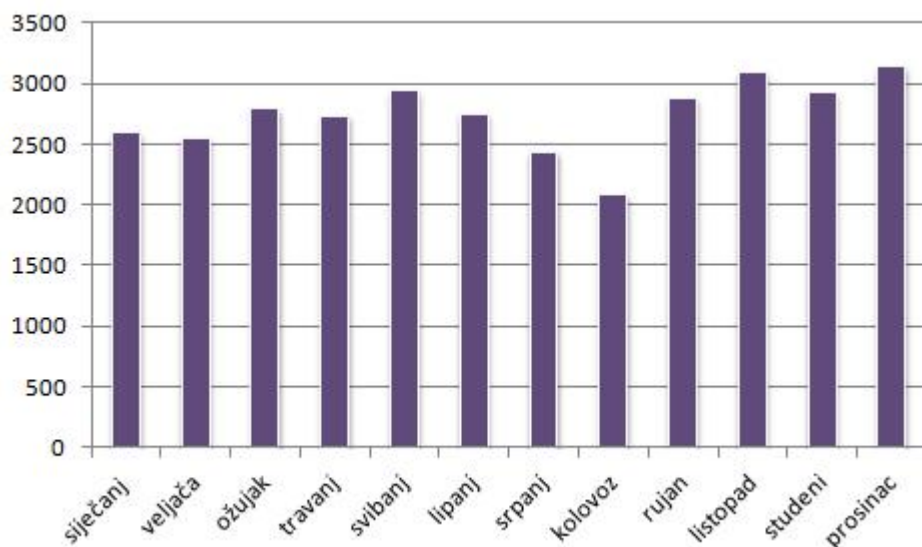
*Slika 24. Grafi ki prikaz broja prometnih nesre a s poginulim sudionicima po dijelovima dana*

Tablica 9 prikazuje tablicu kontingencije za provedbu hi-kvadrat testa homogenosti za provjeru postojanja omjera u raspodjeli prometnih nesre a s razli itim posljedicama po dijelovima dana. Nul-hipoteza nalaže da ne postoji razlika izme u udjela posljedica prometnih nesre a po dijelovima dana. Provedbom testa nul-hipoteza se odbacila. Slika 23 i Slika 24 grafi ki prikazuju podatke iz tablice kontingencije. Grafi ki prikaz broja prometnih nesre a s poginulim sudionicima prikazan je odvojeno zbog manjeg broja nesre a, kako bi slike bile jasnije. Rezultat dobiven provedbom hi-kvadrat testa može se na ovaj na in i vizualno provjeriti. Kod nesre a s materijalnom štetom i ozlije enim sudionicima omjeri u broju nesre a imaju sli ne karakteristike (najviše nesre a izme u 12 h i 18 h, a najmanje izme u 0 h i 6 h), ali kod prometnih nesre a s poginulim sudionicima to nije tako (najviše ovakvih prometnih nesre a zabilježeno je od 6 h do 12 h, a zna ajno je manji njihov udio izme u 12 h i 18 h).

Tablica 10. Distribucija prometnih nesreća po mjesecima tijekom godine (2010.-2013.)

(P = 0.000, hi-kvadrat = 144.07, broj stupnjeva slobode = 22, razina značajnosti = 5%)

| <i>Prometne nesreće</i>          | <i>Sijeanj (%)</i> | <i>Veljača (%)</i> | <i>Ožujak (%)</i> | <i>Travanj (%)</i> | <i>Svibanj (%)</i> | <i>Lipanj (%)</i> | <i>Srpanj (%)</i> | <i>Kolovoz (%)</i> | <i>Rujan (%)</i> | <i>Listopad (%)</i> | <i>Studeni (%)</i> | <i>Prosinac (%)</i> | <i>Ukupno (%)</i> |
|----------------------------------|--------------------|--------------------|-------------------|--------------------|--------------------|-------------------|-------------------|--------------------|------------------|---------------------|--------------------|---------------------|-------------------|
| <i>S materijalnom štetom</i>     | 2196<br>(84.43)    | 2118<br>(82.90)    | 2273<br>(81.41)   | 2164<br>(78.98)    | 2267<br>(76.93)    | 2135<br>(77.47)   | 1841<br>(75.36)   | 1602<br>(76.61)    | 2230<br>(77.32)  | 2490<br>(80.27)     | 2347<br>(80.21)    | 2568<br>(81.65)     | 26231<br>(79.53)  |
| <i>S ozlijeđenim sudionicima</i> | 397<br>(15.26)     | 426<br>(16.67)     | 508<br>(18.19)    | 567<br>(20.69)     | 666<br>(22.60)     | 606<br>(21.99)    | 586<br>(23.99)    | 481<br>(23.00)     | 639<br>(22.16)   | 600<br>(19.34)      | 572<br>(19.55)     | 565<br>(17.97)      | 6613<br>(20.05)   |
| <i>S poginulim sudionicima</i>   | 8<br>(0.31)        | 11<br>(0.43)       | 11<br>(0.39)      | 9<br>(0.33)        | 14<br>(0.48)       | 15<br>(0.54)      | 16<br>(0.65)      | 8<br>(0.38)        | 15<br>(0.52)     | 12<br>(0.39)        | 7<br>(0.24)        | 12<br>(0.38)        | 138<br>(0.42)     |
| <i>Ukupno</i>                    | 2601               | 2555               | 2792              | 2740               | 2947               | 2756              | 2443              | 2091               | 2884             | 3102                | 2926               | 3145                | 32982             |

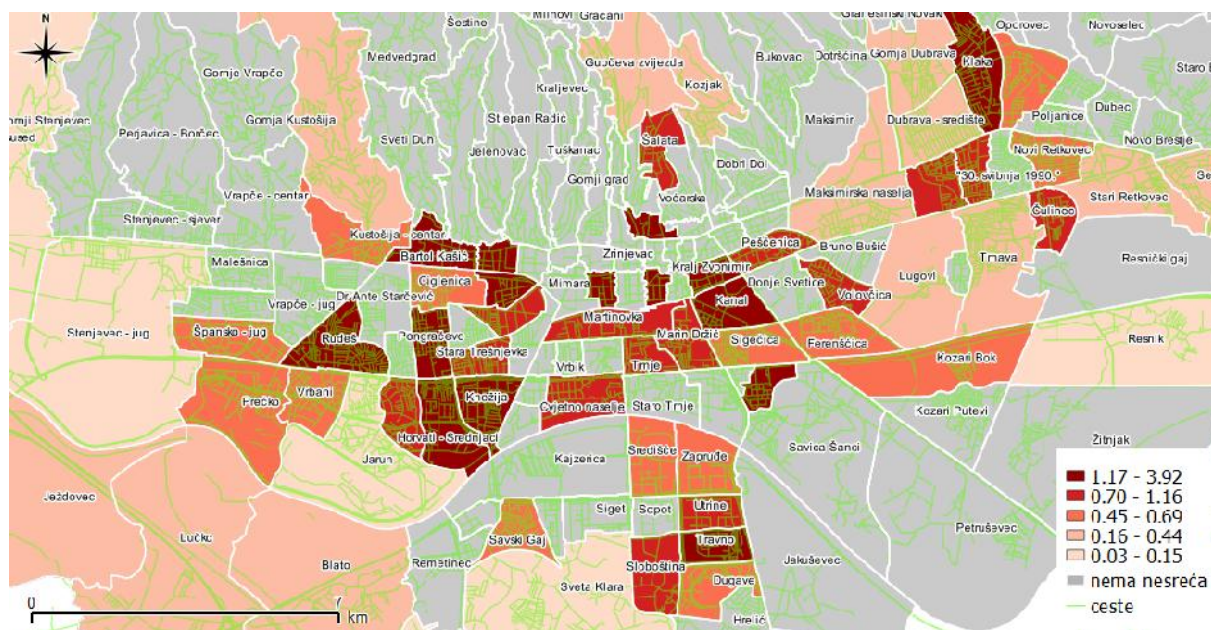


Slika 25. Broj prometnih nesreća po mjesecima u godini (2010.-2013.)

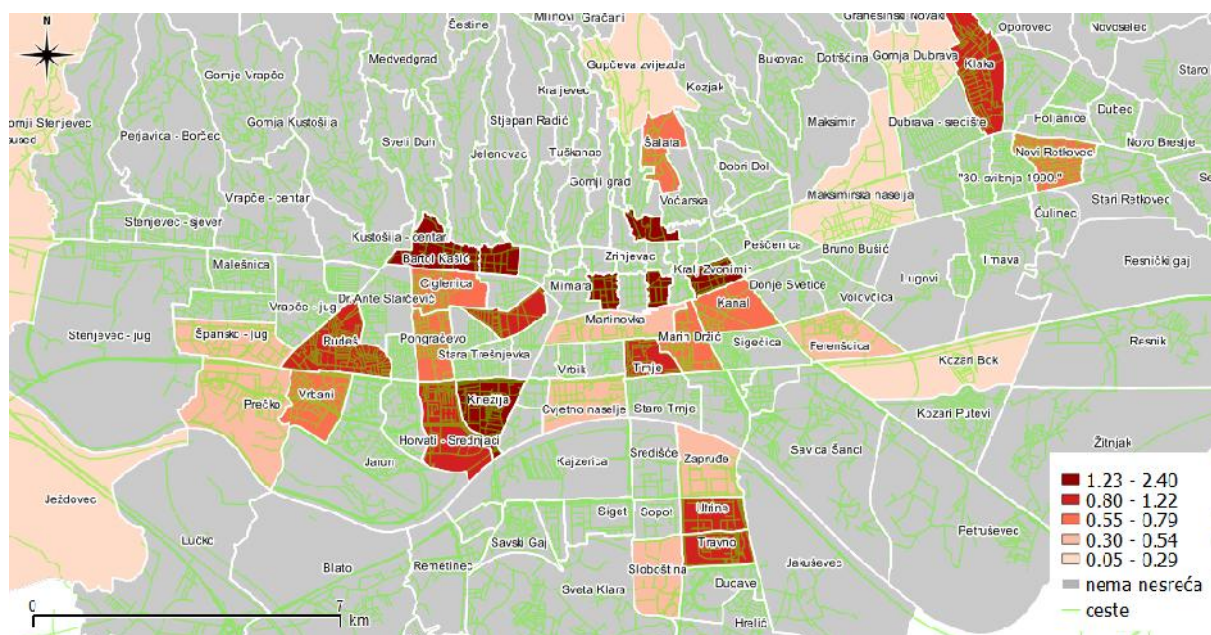
Tablica 10 prikazuje distribuciju prometnih nesreća po mjesecima tijekom četiri godine. Poslužila je kao tablica kontingencije u hi-kvadrat testu homogenosti. Postavljena je nul-hipoteza da ne postoji razlika u omjerima broja prometnih nesreća i različitih posljedica po mjesecima u godini. Na temelju dobivenog rezultata, navedena nul-hipoteza je odbijena.

Slika 25 prikazuje ukupni broj prometnih nesreća po mjesecima tijekom četiri godine. Utvrđeno je da se najveći broj prometnih nesreća događa u jesenskim mjesecima. Slika omogućuje jasniji uvid u raspodjelu broja prometnih nesreća tijekom godina i istovremeno razdoblja s većom učestalošću nesreća.

### 4.3. Rezultati obrade statisti kom metodom normalizacije



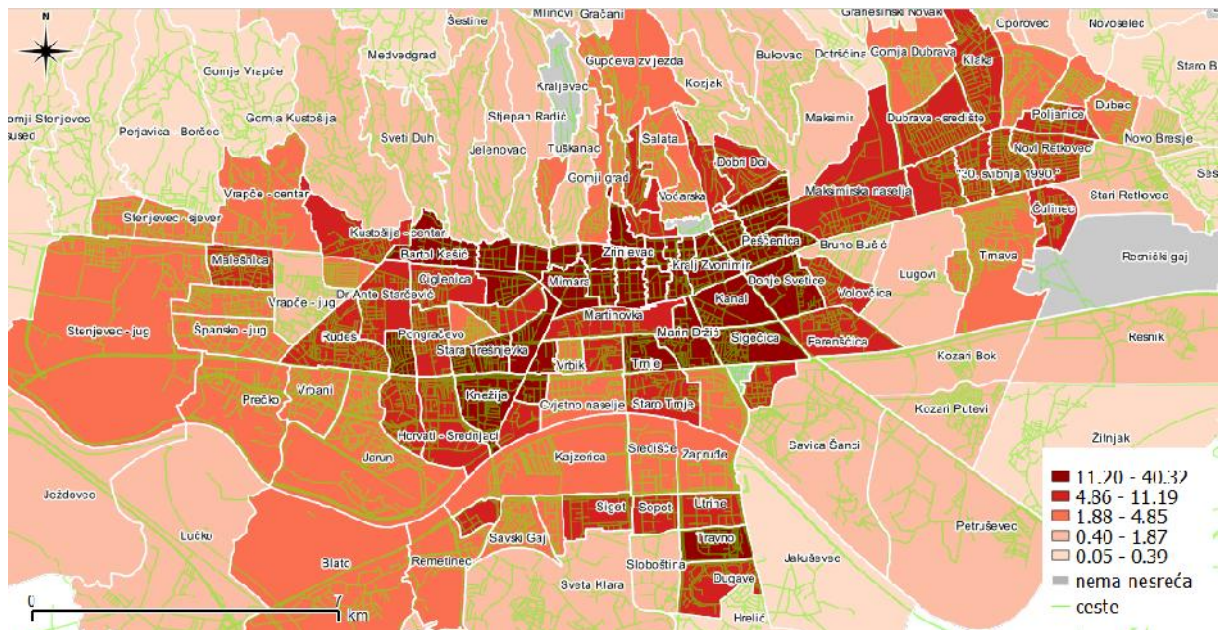
Slika 26. Broj prometnih nesre a s poginulim osobama po km<sup>2</sup>



Slika 27. Broj prometnih nesre a s poginulim pješacima po km<sup>2</sup>

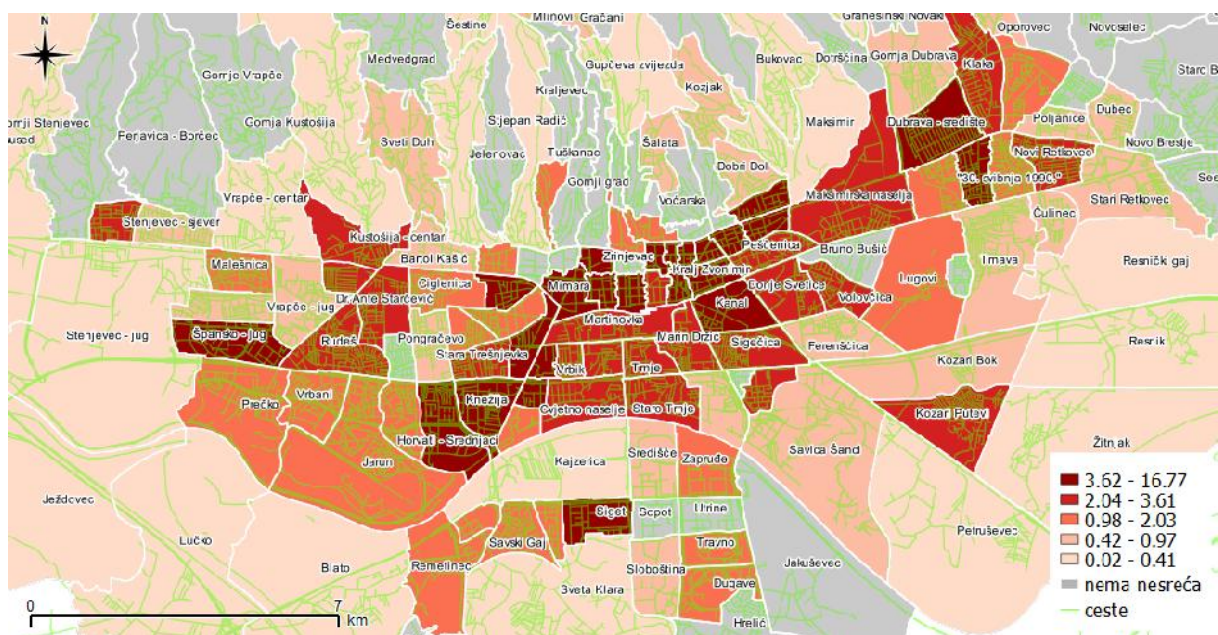
Vizualni prikaz rezultata normalizacije prometnih nesre a s poginulim sudionicima nije najprikladniji za analizu zbog malog broja nesre a (Slika 26, Slika 27). Ve ina mjesnih odbora s najve om gusto om je zastupljena u obje slike što ukazuje da se ve i dio nesre a s poginulim sudionicima odnosi na pješake. Postoji velik broj odbora bez ijedne registrirane

nesre e što onemogu uje daljnju analizu ovog skupa podataka. Slike prikazuju utjecaj malog broja opažanja na metodu normalizacije.



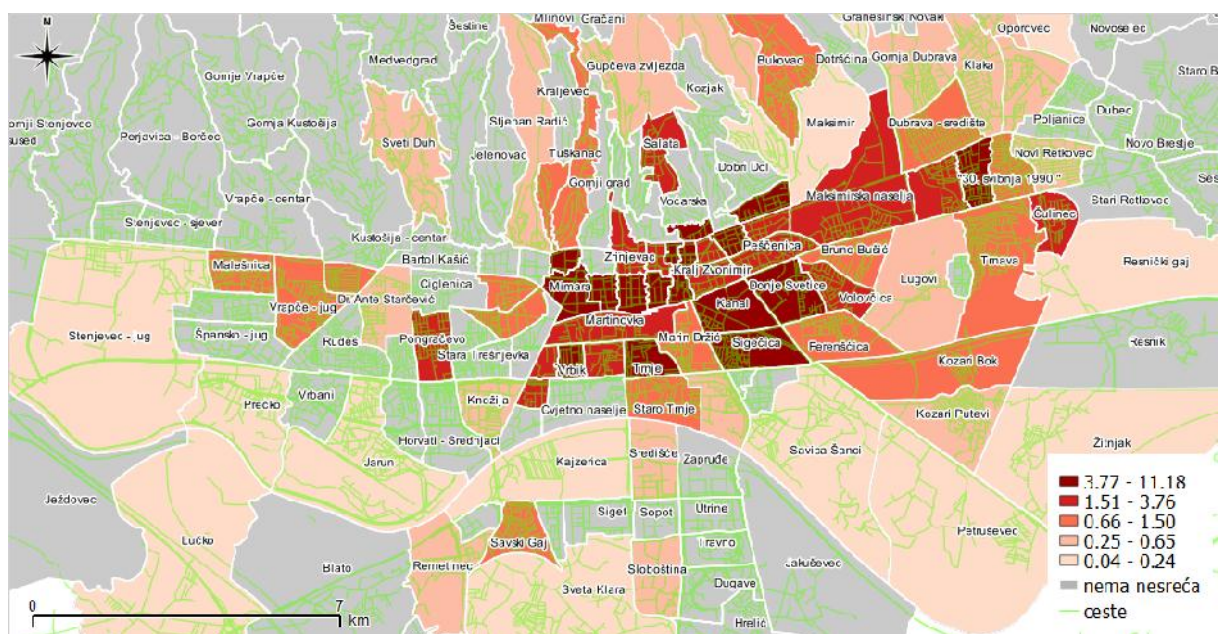
Slika 28. Broj prometnih nesreća s teško ozlijeđenim sudionicima po km<sup>2</sup>

Slika 28 prikazuje distribuciju gustoće prometnih nesreća s teško ozlijeđenim sudionicima. Postoji grupiranje mjesnih odbora s najvećom gustoćom u centru Zagreba i duž glavnih prometnica (Ilica, Savska cesta i Ulica Vjekoslava Heinzela). Ostali mjesni odbori posjeduju znatno manju gustoću. Slikom je prikazan primjer pravilne primjene metode kad su skoro sve prostorne jedinice uključene u analizu.



Slika 29. Broj prometnih nesreća s poginulim i ozlijeđenim biciklistima po km<sup>2</sup>

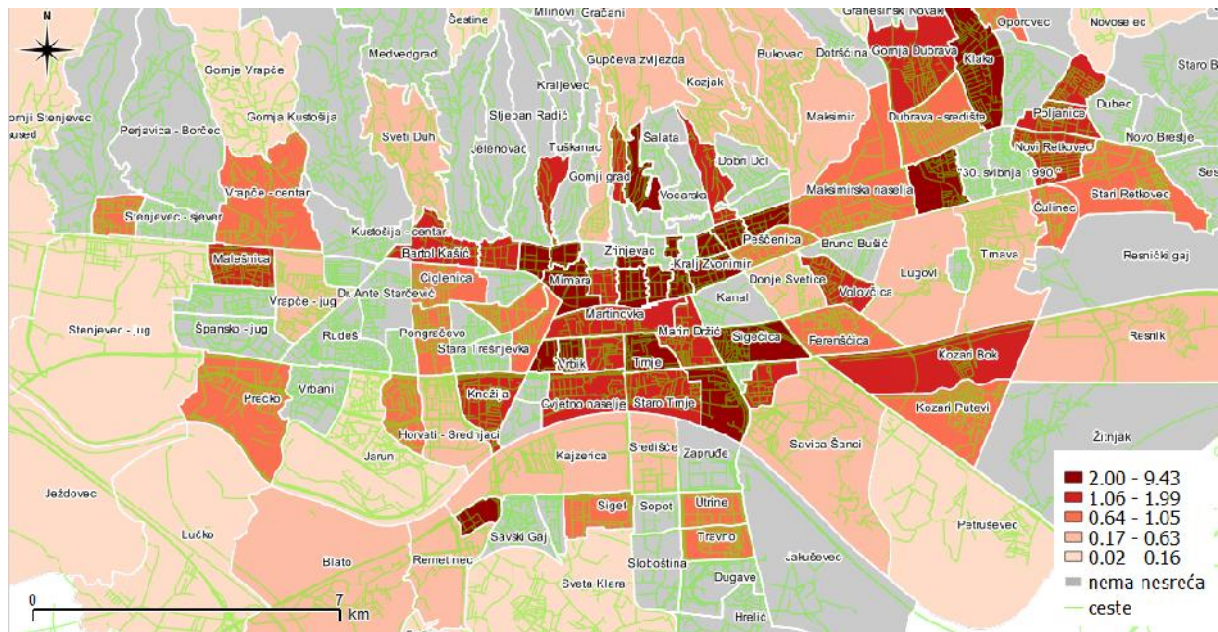
Slika 29 prikazuje distribuciju gustoće prometnih nesreća s poginulim i ozlijeđenim biciklistima. Uočeno je grupiranje mjesnih odbora s najvećom gustoćom duž Savske ceste, Vlaške ulice, Maksimirske ceste i Ulice Vjekoslava Heinzela. Slika ističe opasne zone za bicikliste.



Slika 30. Broj prometnih nesreća s poginulim i ozlijeđenim motociklistima po km<sup>2</sup>

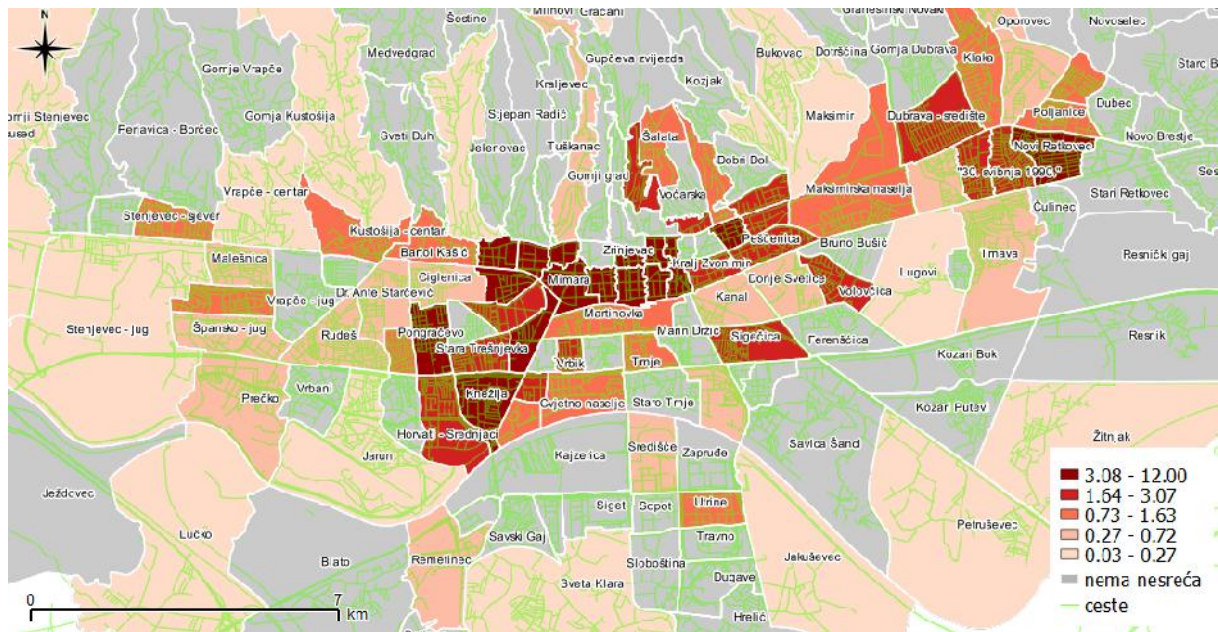
Slika 30 prikazuje distribuciju gustoće prometnih nesreća s poginulim i ozlijeđenim motociklistima. Na slici su uočena dva grupiranja s najvećom gustoćom. Prvo se odnosi na

mjesne odbore izme u Savske ceste i Trga žrtava fašizma (Ulica baruna Trenka i Ulica Andrije Hebranga). Drugo grupiranje se nalazi izme u Avenije Marina Drži a i Ulice Vjekoslava Heinzela. Slika naglašava da se nesre e s motociklistima doga aju pretežito na isto nom dijelu grada.



*Slika 31. Broj prometnih nesreća po km<sup>2</sup> zbog nepropisne brzine*

Slika 31 prikazuje distribuciju gustoće prometnih nesreća uzrokovanih nepropisnom brzinom. Uključene su sve posljedice prometnih nesreća. Uočena su dva grupiranja mjesnih odbora s najvećom gustoćom. Prvo se odnosi na mjesne odbore između Ulice kralja Zvonimira i Ulice kneza Branimira te duž Maksimirske ceste. Drugo grupiranje čine mjesni odbori između Ulice i Ulice Valentina Vodnika. Slika naglašava u kojim zonama treba smanjiti najveću u dozvoljenu brzinu.



*Slika 32. Broj prometnih nesreća po km<sup>2</sup> koje su se dogodile zbog nekorištenja obilježenih pješačkih prijelaza*

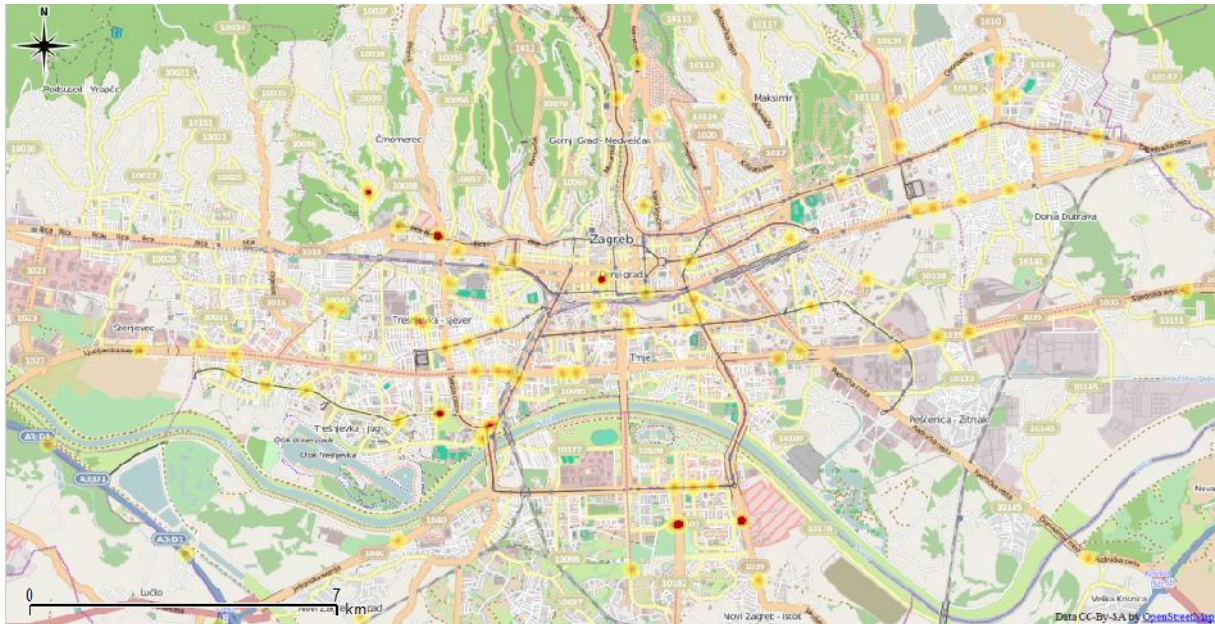
Slika 32 prikazuje distribuciju gustoće prometnih nesreća uzrokovanih nekorištenjem obilježenih pješačkih prijelaza. Uključene su sve posljedice prometnih nesreća. Uoeno je grupiranje mjesnih odbora s najvećom gustoćom u centru Zagreba, duž Savske ceste i Ilice. Slika naglašava na kojim zonama treba popraviti signalizaciju za pješake.



#### 4.4. Rezultati obrade statisti kom metodom KDE

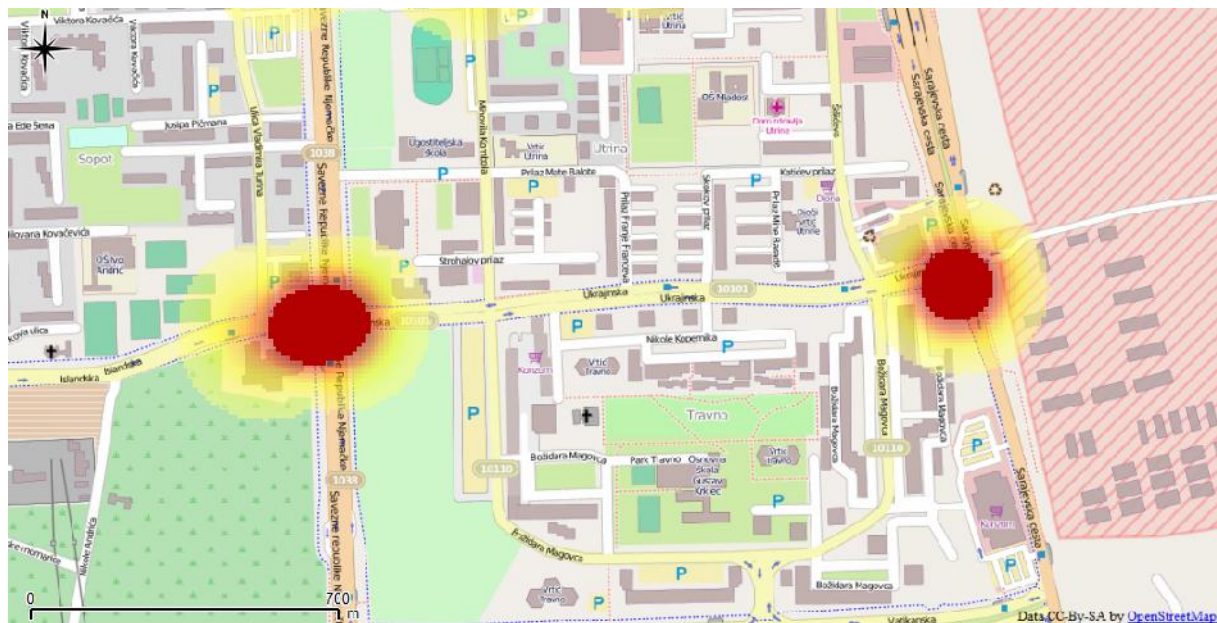
U nastavku su prikazani rezultati KDE metode. Odabrane su vrste prometnih nesre a koje su od najve eg zna aja MUP-u. Prva slika potpoglavlja daje pregled svih žarišta na promatranom podru ju, a zatim slijedi obrada pojedina nih žarišta.

##### 4.4.1. Prometne nesre e s poginulim sudionicima



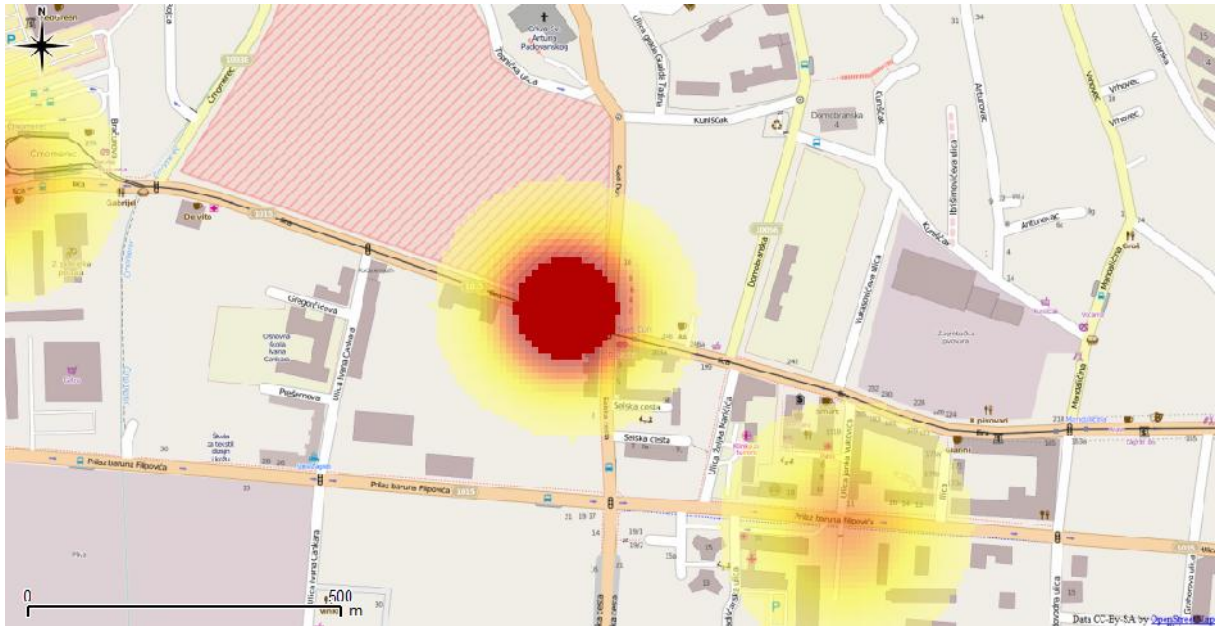
*Slika 33. Prikaz žarišta najtežih prometnih nesre a od 2010. do 2013. godine*

Slika 33 prikazuje žarišta najtežih prometnih nesre a, odnosno prometnih nesre a s poginulim sudionicima. Rije je o malom broju prometnih nesre a, ali su ipak uo ena odre ena žarišta koja su u nastavku detaljnije analizirana. Najteže prometne nesre e su jako važne MUP. Iz tog je razloga ova slika bitna jer prikazuje raspored žarišta. U nastavku su pojedina žarišta detaljnije obra ena.



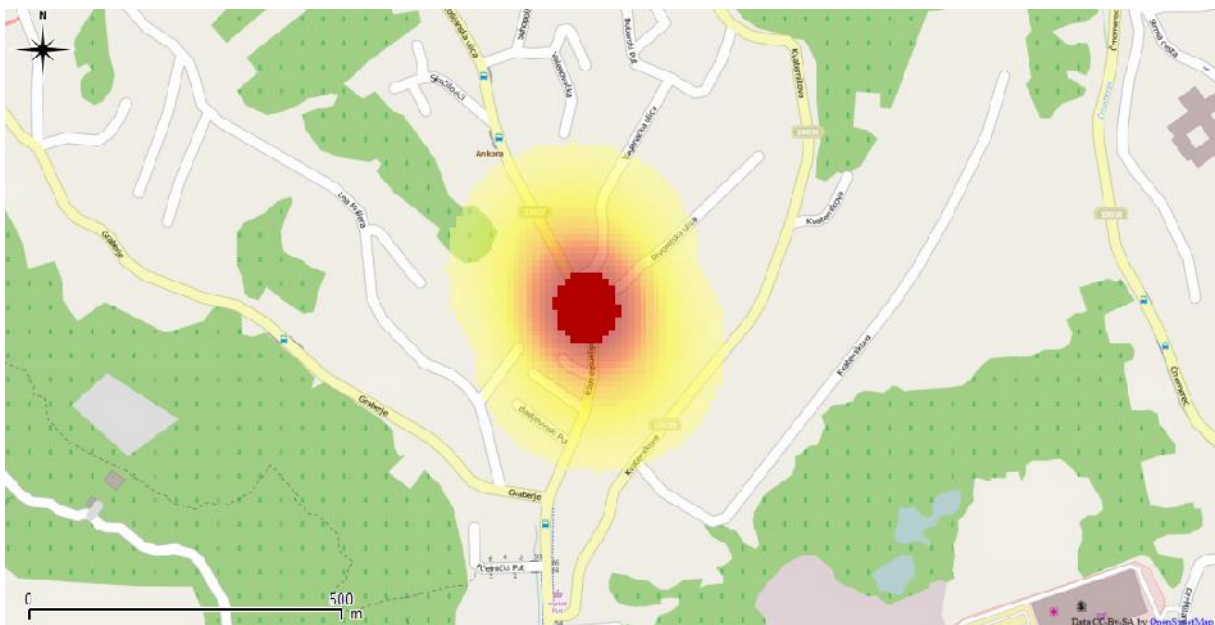
*Slika 34. Prikaz žarišta najtežih prometnih nesre a  
(lijevo žarište - 3 nesre e; desno žarište - 2 nesre e)*

Slika 34 prikazuje dva žarišta. Prvo se nalazi na križanju Ulice Savezne Republike Njema ke i Islanjske ulice, odnosno Ukrajinske ulice (lijevo); a drugo na križanju Ukrajinske ulice i Sarajevske ceste (desno). Na lijevom su se žarištu dogodile tri nesre e ovog tipa. Dvije uklju uju nalet na pješaka, a tre a je slijetanje vozila s ceste. Okolnosti su razli ite u svim nesre ama. Dogodile su se u popodnevnim satima u stabilnim atmosferskim prilikama. Na desnom su se žarištu dogodila dva bo na sudara unutar mjesec dana. U oba slu aja je rije o nepoštivanju prometnih pravila. Nesre e su se dogodile tijekom dana, ali pri razli itim vremenskim uvjetima. Ova su žarišta važna zato što isti u mjesta na kojima se najteže nesre e ponavljaju.



*Slika 35. Prikaz žarišta najtežih prometnih nesre a (2 nesre e)*

Slika 35 prikazuje žarište na križanju Ilice i Svetog duha. Rije je o dva naleta na pješaka zbog neprimjerene brzine. Dogodile su se u ranijim popodnevnim satima i u stabilnim atmosferskim prilikama. Potrebno je istaknuti ovo žarište zbog sli nosti prometnih nesre a koje su se dogodile na istom mjestu.



*Slika 36. Prikaz žarišta najtežih prometnih nesre a (2 nesre e)*

Slika 36 prikazuje žarište na križanju Kustošijanske ulice i Vagana ke ulice. Rije je o dvije prometne nesre e uzrokovane pogreškama voza a. Dogodile su se tijekom no i u stabilnim

atmosferskim prilikama. Nužno je istaknuti ovo žarište iz razloga što na toj lokaciji nema adekvatne horizontalne signalizacije i ne postoji vertikalna signalizacija.

### 4.4.2. Prometne nesre e s poginulim i ozlije enim sudionicima



*Slika 37. Prikaz žarišta prometnih nesre a s poginulim i ozlije enim sudionicima*

Uzevši u obzir mali broj prometnih nesre a s poginulim sudionicima, skup podataka je proširen na prometne nesre e s poginulim i ozlije enim sudionicima. Cilj je pove anje veli ine uzorka radi otkrivanja potencijalnih obrazaca (Slika 37). Slika prikazuje raspored žarišta koja se u nastavku detaljnije obra uju.



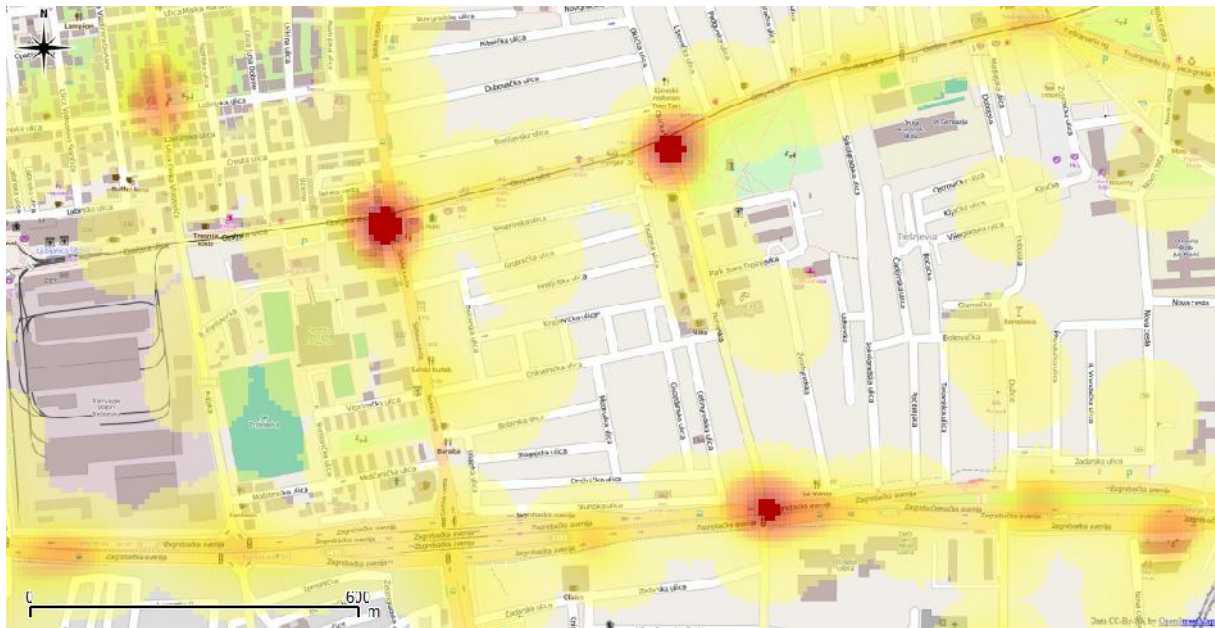
*Slika 38. Prikaz žarišta prometnih nesre a s poginulim i ozlije enim sudionicima (lijevo žarište - 10 nesre a; desno žarište - 11 nesre a)*

Slika 38 ponovo prikazuje žarišta na križanjima Ulice Savezne Republike Njemačke i Islandske, odnosno Ukrajinske ulice (lijevo) te Ukrajinske ulice i Sarajevske ceste (desno). Na lijevom se žarištu prevladavaju nesreće uzrokovane nepoštivanjem prednosti prolaska, nepropisnim uključivanjem u promet i pogreškama pješaka. Najčešće rezultiraju bodnim sudarima, sudarima iz suprotnih smjerova i naletima na pješaka ili biciklista. Već dio nesreća se dogodio tijekom dana u povoljnim atmosferskim prilikama. Na desnom žarištu prevladavaju prometne nesreće uzrokovane nepoštivanjem prednosti prolaska. One rezultiraju bodnim sudarima i sudarima iz suprotnih smjerova. Dogodaju se u stabilnim atmosferskim uvjetima tijekom cijelog dana. Nesreće su na ovoj lokaciji sve češće. Iako nije riječ o velikim žarištima, sliku je potrebno istaknuti zato što su ove lokacije posebno problematične.



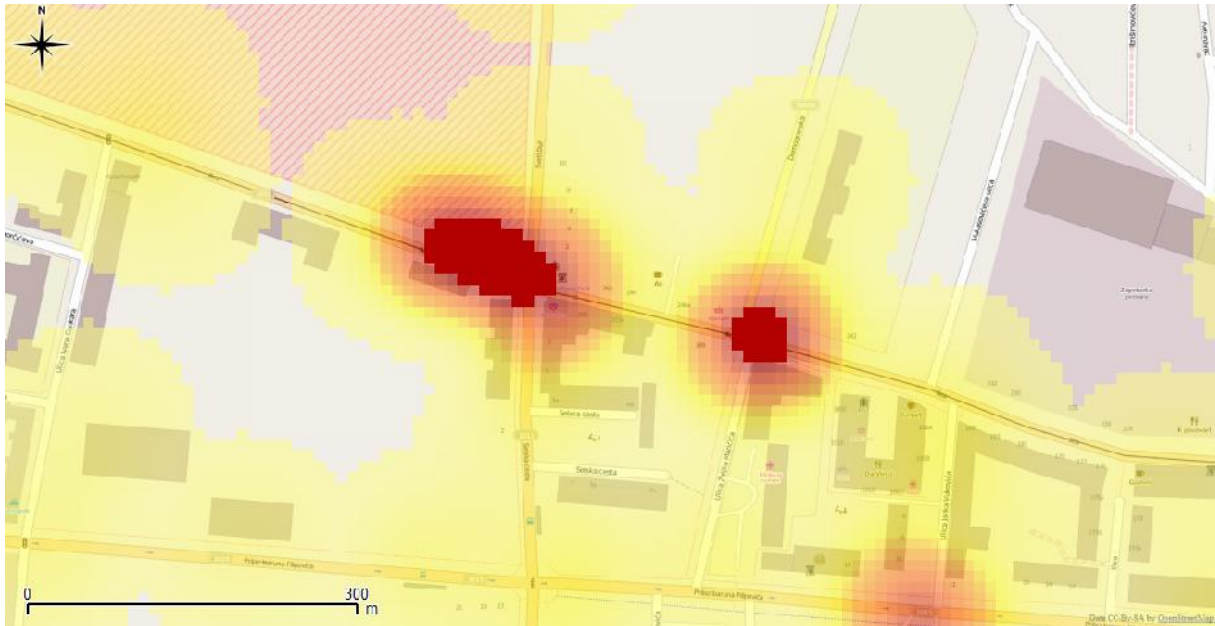
*Slika 39. Prikaz žarišta prometnih nesre a s poginulim i ozlije enim sudionicima  
(lijevo žarište - 9 nesre a; desno žarište - 14 nesre a)*

Slika 39 prikazuje dva žarišta. Prvo je na križanju Jadranske avenije i Blata (lijevo), a drugo na južnom dijelu Rotor (desno). Na lijevom žarištu prevladavaju bo ni sudari uzrokovani nepoštivanjem svjetlosnog znaka. Druga vrsta nesre a su sudari vozila u slijedu uzrokovani neprimjerenom brzinom. Ve ina nesre a se dogodila tijekom ljetnih mjeseci u ranijim poslijepodnevним satima i u stabilnim atmosferskim prilikama. Na desnom žarištu prevladavaju sudari kod vožnje u slijedu koji su uzrokovani nepropisnom brzinom i vožnjom na nedovoljnoj udaljenosti. Druga vrsta nesre a su bo ni sudari zbog nepoštivanja prednosti prolaska. Ve ina se nesre a dogodila tijekom kasnih poslijepodnevних sati, u radnim danima i pri stabilnim atmosferskim prilikama. Žarišta su površinski mala, ali je o ito rije o velikoj gusto i nesre a iji se uzroci trebaju istražiti.



*Slika 40. Prikaz žarišta prometnih nesre a s poginulim i ozlije enim sudionicima (gornje lijevo žarište - 12 nesre a; gornje desno žarište - 9 nesre a; donje žarište - 8 nesre a)*

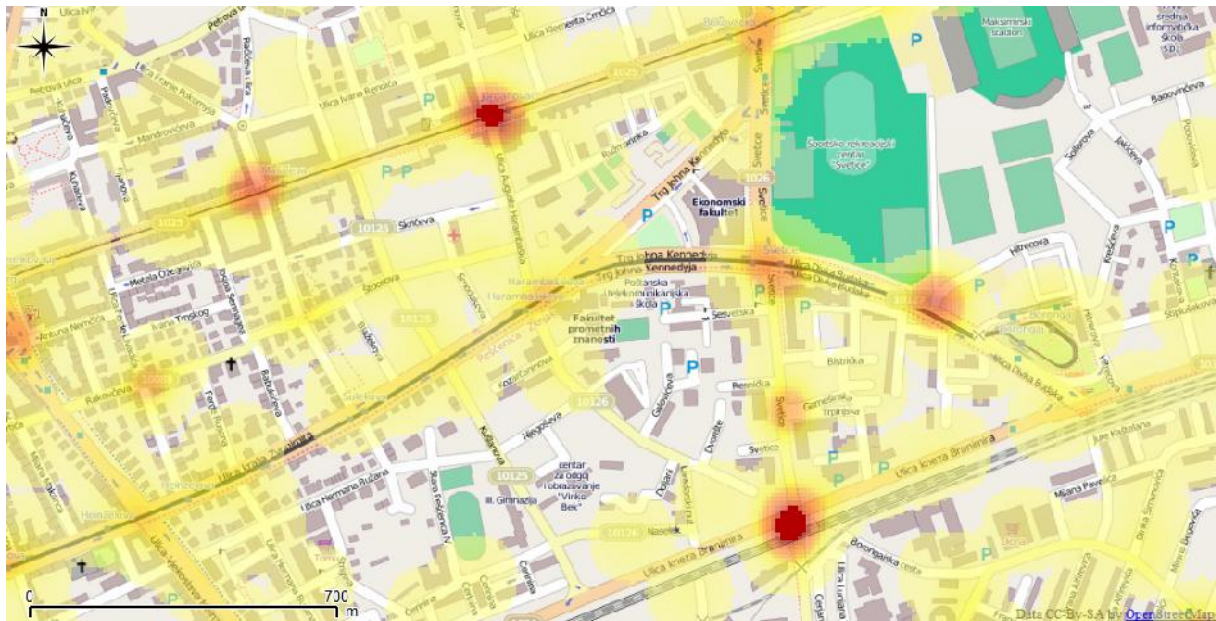
Slika 40 prikazuje tri žarišta. Prvo je na križanjima Selske ceste i Ozaljske ulice (gore lijevo), drugo na križanju Ozaljske ulice, Oki ke ulice i Nehajske (gore desno), a treće na križanju Nehajske i Zagreba ke avenije (dolje). Na gornjem lijevom žarištu prevladavaju prometne nesreće uzrokovane nepoštivanjem svjetlosnog znaka. Najčešće rezultiraju bodnim sudarima i naletima na motocikliste. Naleti na pješake su većinom uzrokovani pogreškama od strane pješaka. Većina nesreća se dogodila u kasnijim poslijepodnevним i večernjim satima te u stabilnim atmosferskim prilikama. Na gornjem desnom žarištu prevladavaju nesreće uzrokovane nepoštivanjem svjetlosnog znaka koje rezultiraju bodnim sudarima i sudarima iz suprotnih smjerova. Nesreće su se dogodile pri stabilnim atmosferskim prilikama u priepodnevним satima. Na donjem žarištu prevladavaju naleti na bicikliste, a najčešći su uzroci brzina neprimjerena uvjetima i nepoštivanje svjetlosnog znaka. Nesreće su podjednako zastupljene tijekom dana, a atmosferske prilike uglavnom stabilne. Treba napomenuti da su nesreće na toj lokaciji sve češće. Kao i u prethodnom slučaju, riječ je o koncentraciji velikog broja nesreća na malim područjima. Također, bitno je istaknuti da je većina nesreća na ovim lokacijama uzrokovana nepoštivanjem svjetlosnog znaka.



*Slika 41. Prikaz žarišta prometnih nesre a s poginulim i ozlije enim sudionicima (lijevo žarište - 12 nesre a; desno žarište - 15 nesre a)*

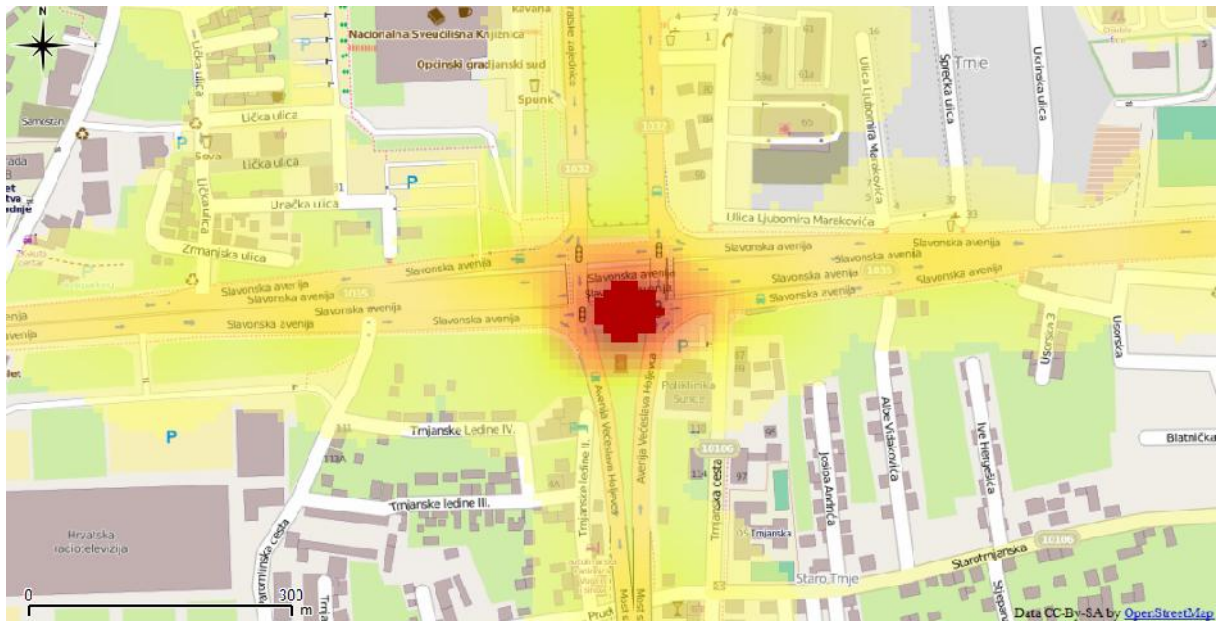
Slika 41 prikazuje dva žarišta. Prvo je na križanju Ilice, Svetog Duha i Selske ceste (lijevo), a drugo na križanju Ilice i Domobranske ulice (desno). Na lijevom su žarištu naj eš e prometne nesre e uzrokovane brzinom neprimjerenom uvjetima. Naj eš e rezultiraju naletom na pješaka i sudarom vozila u slijedu. Ve ina nesre a se dogodila tijekom dana u stabilnim atmosferskim prilikama. Na desnom žarištu prevladavaju naleti na pješake uzrokovani brzinom neprimjerenom uvjetima. Dogodili su se tijekom radnih dana, uglavnom u poslijepodnevnim satima i pri stabilnim atmosferskim prilikama. Na toj lokaciji nema semafora. Potrebno je istaknuti žarišta u kojima ve inom stradaju pješaci.





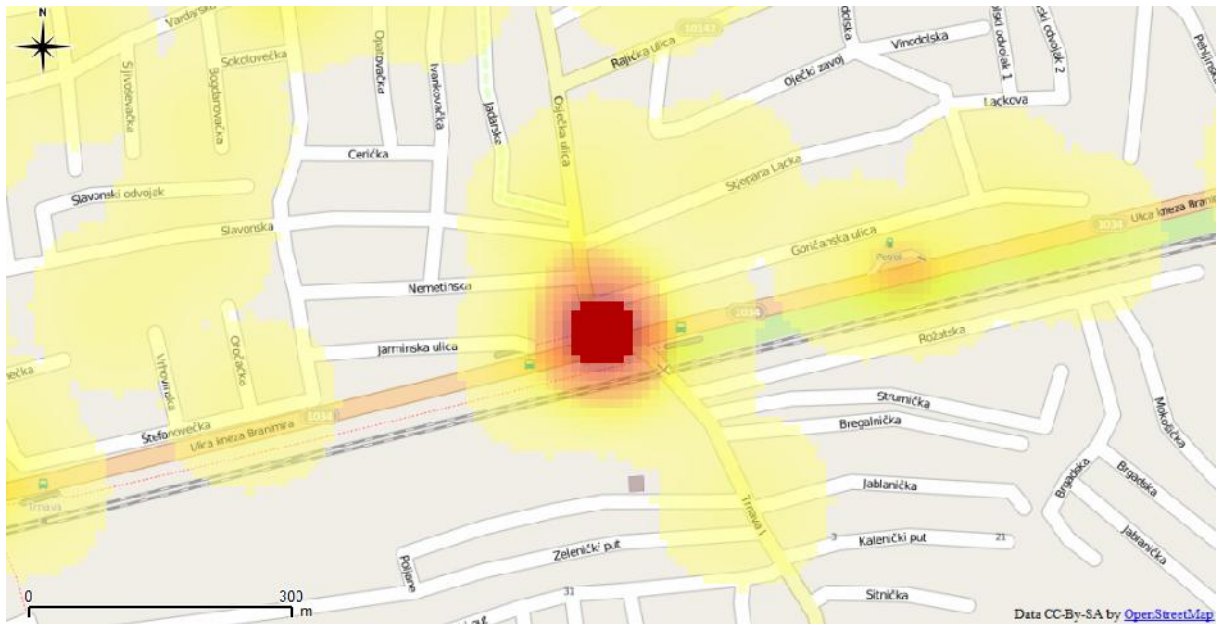
*Slika 42. Prikaz žarišta prometnih nesre a s poginulim i ozlije enim sudionicima  
(gornje žarište - 12 nesre a; donje žarište - 18 nesre a)*

Slika 42 prikazuje dva žarišta. Prvo je na križanju Maksimirske ceste s Jordanovcem i Ulicom Augusta Harambašića (gore), a drugo na križanju Ulice kneza Branimira sa Sveticama i Donjim Sveticama (dolje). Na gornjem žarištu prevladavaju bo ni sudari uzrokovani nepoštivanjem svjetlosnog znaka i naleti na pješake uzrokovani nepoštivanjem svjetlosnog znaka, uglavnom od strane pješaka. Ve ina nesre a se dogodila u poslijepodnevni satima pri stabilnim atmosferskim prilikama. Na donjem žarištu prevladavaju nesre e uzrokovane nepoštivanjem prednosti prolaska i nepoštivanjem svjetlosnog znaka. Naj eš e rezultiraju bo nim sudarima i sudarima iz suprotnih smjerova. Doga aju u poslijepodnevni satima i stabilnim atmosferskim prilikama. Uo eno je znatno smanjenje broja navedenih vrsta nesre a, ali se pove ao broj naleta na pješake koji ne poštuju svjetlosne znakove. Slika jasno prikazuje da je uzrok ve ine prometnih nesre a na ovom podru ju nepoštivanje svjetlosnog znaka.



*Slika 43. Prikaz žarišta prometnih nesreća s poginulim i ozlijeđenim sudionicima (24 nesreće)*

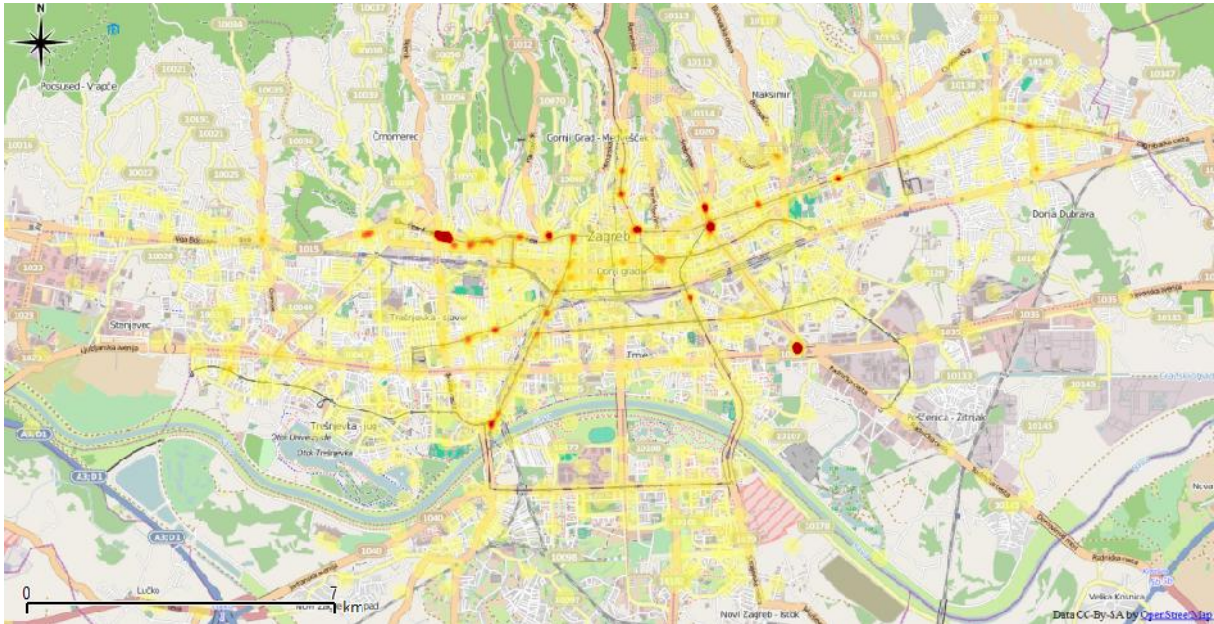
Slika 43 prikazuje žarište na križanju Ulice Hrvatske bratske zajednice i Slavonske avenije. Prevladavaju nesreće uzrokovane nepoštivanjem svjetlosnog znaka i nepropisnim uključivanjem u promet. Najčešće rezultiraju običnim sudarima i naletima na bicikliste. Nesreće se događaju tijekom cijelog dana i u stabilnim atmosferskim prilikama. Ovo je žarište potrebno istaknuti zbog velike koncentracije prometnih nesreća.



*Slika 44. Prikaz žarišta prometnih nesre a s poginulim i ozlije enim sudionicima (19 nesre a)*

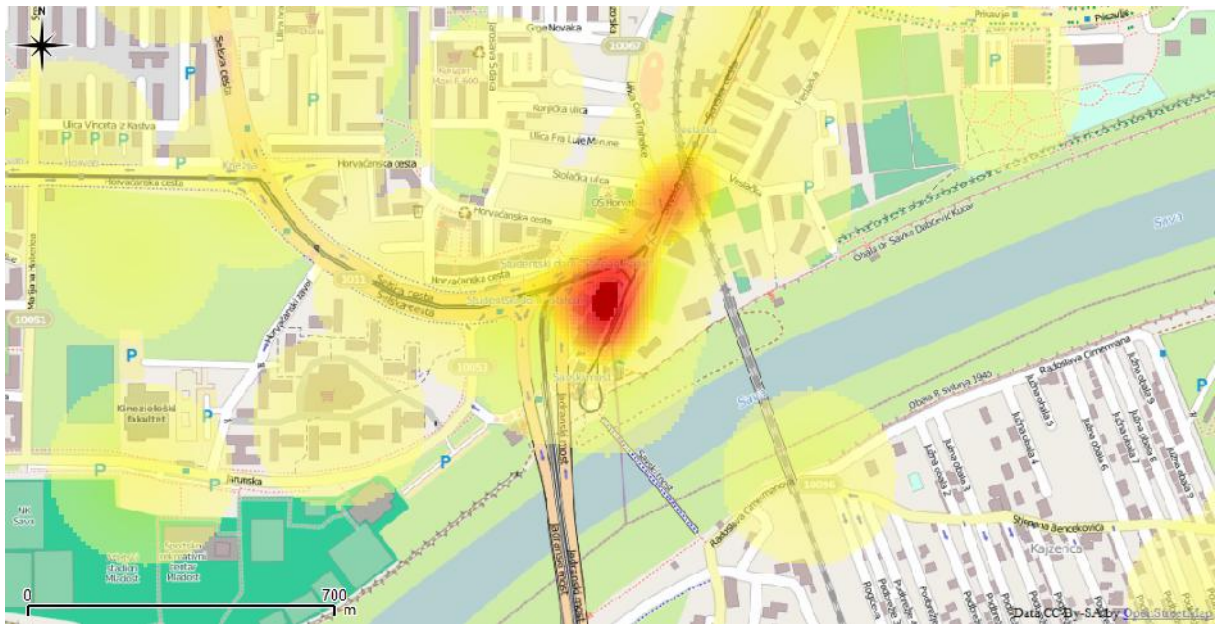
Slika 44 prikazuje žarište na križanju Osječke ulice, Ulice kneza Branimira i Trnava I. Ovdje prevladavaju nesreće uzrokovane nepoštivanjem svjetlosnog znaka, nepoštivanjem prednosti prolaska i brzine neprimjerene uvjetima. Najčešće rezultiraju sudarom vozila u slijedu. Većina nesreća se dogodila tijekom noćnih i stabilnim atmosferskim prilikama. Kao i u prethodnom slučaju, potrebno je istaknuti žarište zbog velike koncentracije prometnih nesreća.

## 4.4.3. Prometne nesre e s poginulim i ozlije enim pješacima



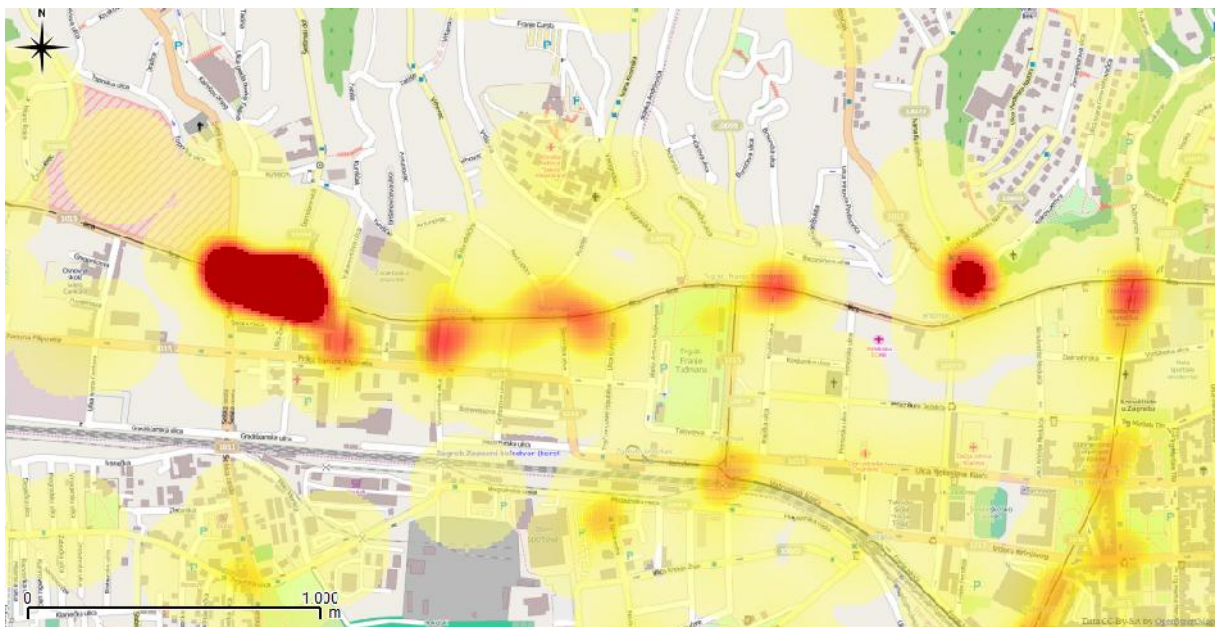
*Slika 45. Prikaz žarišta prometnih nesre a s poginulim i ozlije enim pješacima*

U prethodnoj je analizi uo eno da su naleti na pješaka naj eš e uzrokovani brzinom neprimjerenom uvjetima i nepoštivanjem svjetlosnog znaka od pješaka. Budu i da su pješaci jedna od najugroženijih skupina sudionika u promet, sljede i je korak obrada prometnih nesre a s poginulim i ozlije enim pješacima (Slika 45). Slika prikazuje koja su podru ja najopasnija za pješake.



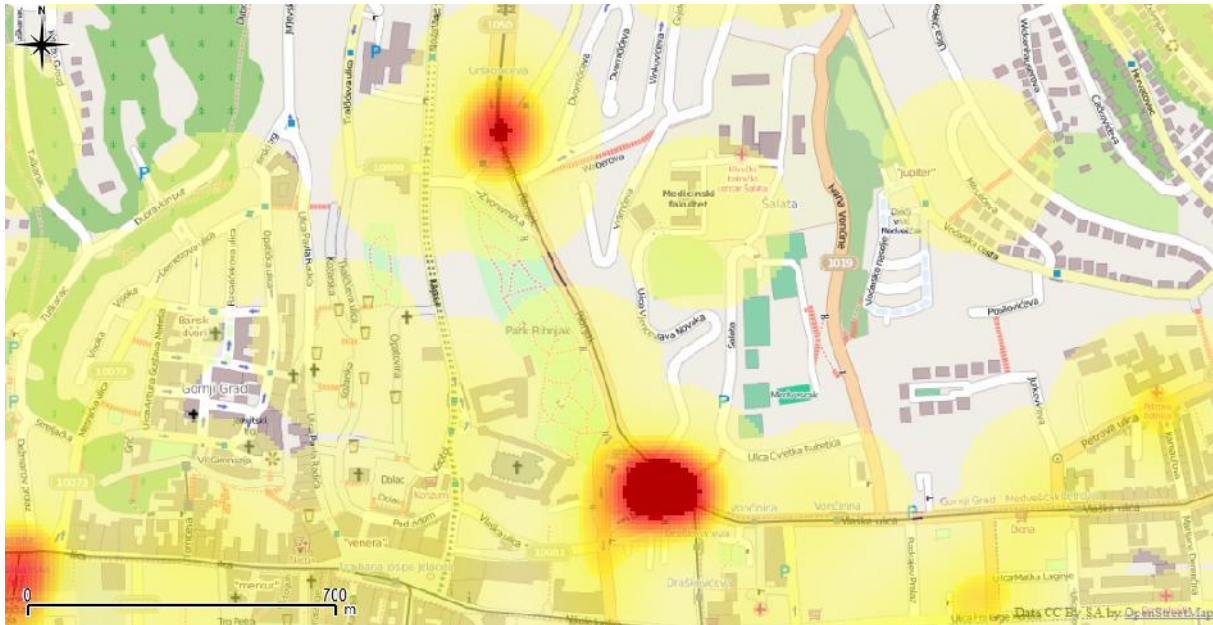
*Slika 46. Prikaz žarišta prometnih nesre a s poginulim i ozlije enim pješacima (11 nesre a)*

Slika 46 prikazuje žarište na križanjima Selske ceste i Jadranskog mosta. Na donjem žarištu prevladavaju nesre e uzrokovane nepoštivanjem pravila prometa od strane pješaka i brzinom neprimjerenom uvjetima. Ve ina nesre a se dogodila tijekom no i u stabilnim atmosferskim prilikama. Potrebno je istaknuti žarište zbog velike koncentracije nesre a.



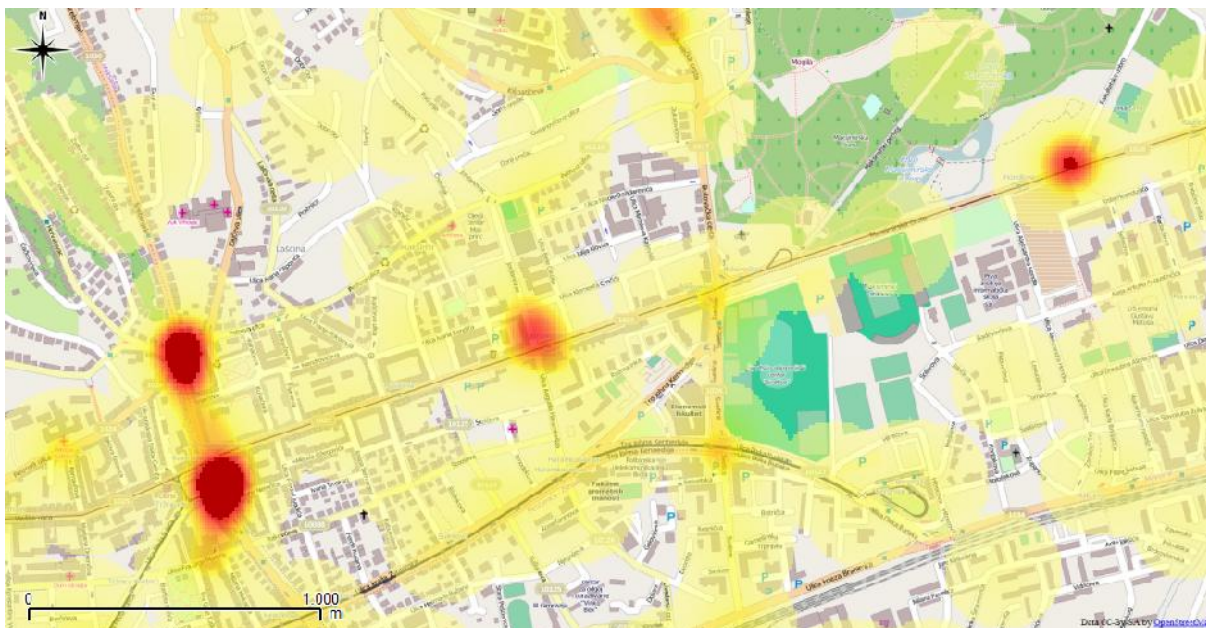
*Slika 47. Prikaz žarišta prometnih nesre a s poginulim i ozlije enim pješacima (lijevo žarište - 24 nesre a; desno žarište - 9 nesre a)*

Slika 47 prikazuje dva žarišta. Prvo na križanju Ilice i Svetog Duha (lijevo), a drugo na Britanskom trgu (desno). Na lijevom žarištu prevladavaju nesre e uzrokovane brzinom neprimjerenom uvjetima koje su se dogodile tijekom dana i u stabilnim atmosferskim prilikama. Na desnom žarištu prevladavaju nesre e uzrokovane pogreškama voza a koje su se tako er dogodile tijekom dana i stabilnim atmosferskim prilikama. Potrebno je izdvojiti ovu sliku koja prikazuje žarišta uzrokovana pogreškama voza a.



*Slika 48. Prikaz žarišta prometnih nesre a s poginulim i ozlije enim pješacima  
(donje žarište - 11 nesre a; gornje žarište - 8 nesre e)*

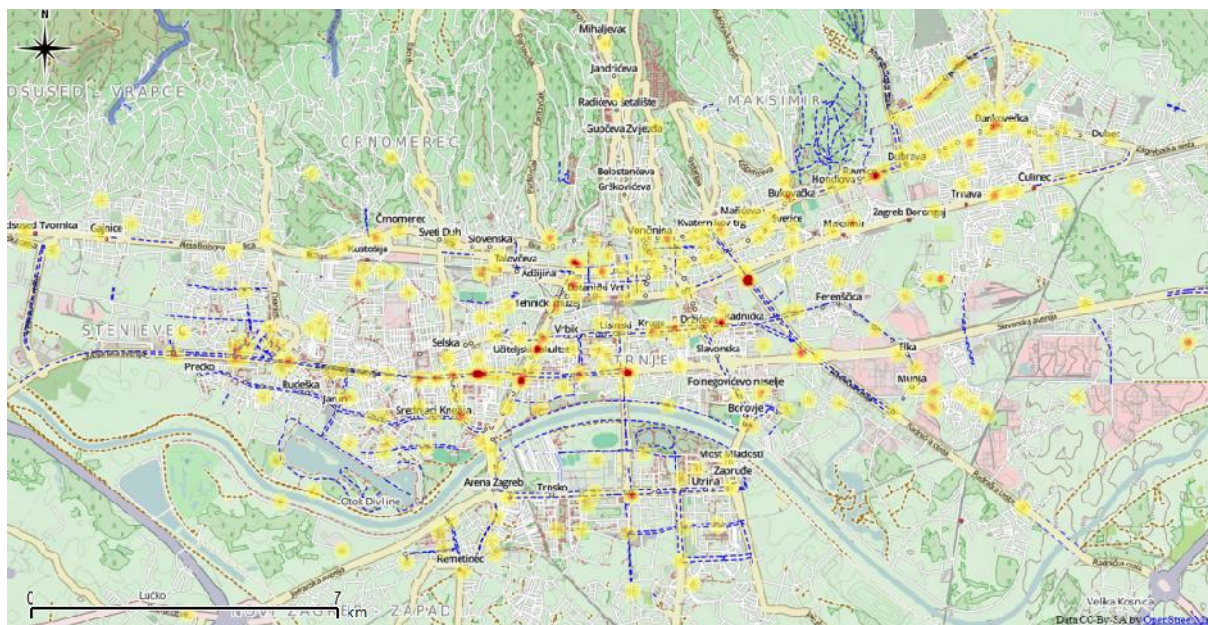
Slika 48 prikazuje dva žarišta. Prvo na Trgu Josipa Langa (dolje), a drugo na križanju Medveška i Ulice Nike Grškovi a (gore). Na donjem žarištu prevladavaju nesre e uzrokovane brzinom neprimjerenom uvjetima i nepropisnim skretanjem. Doga aju se tijekom cijelog dana i u stabilnim atmosferskim prilikama. Na gornjem žarištu prevladavaju nesre e uzrokovane pogreškama pješaka. Najviše nesre a se dogodilo u poslijepodnevnim satima, a sve su se dogodile u stabilnih atmosferskih prilika. Slika je izdvojena zbog žarišta s velikim koncentracijama nesre a.



*Slika 49. Prikaz žarišta prometnih nesre a s poginulim i ozlije enim pješacima  
(donje lijevo žarište - 11 nesre a; gornje lijevo žarište - 11 nesre a;  
gornje desno žarište – 7 nesre a)*

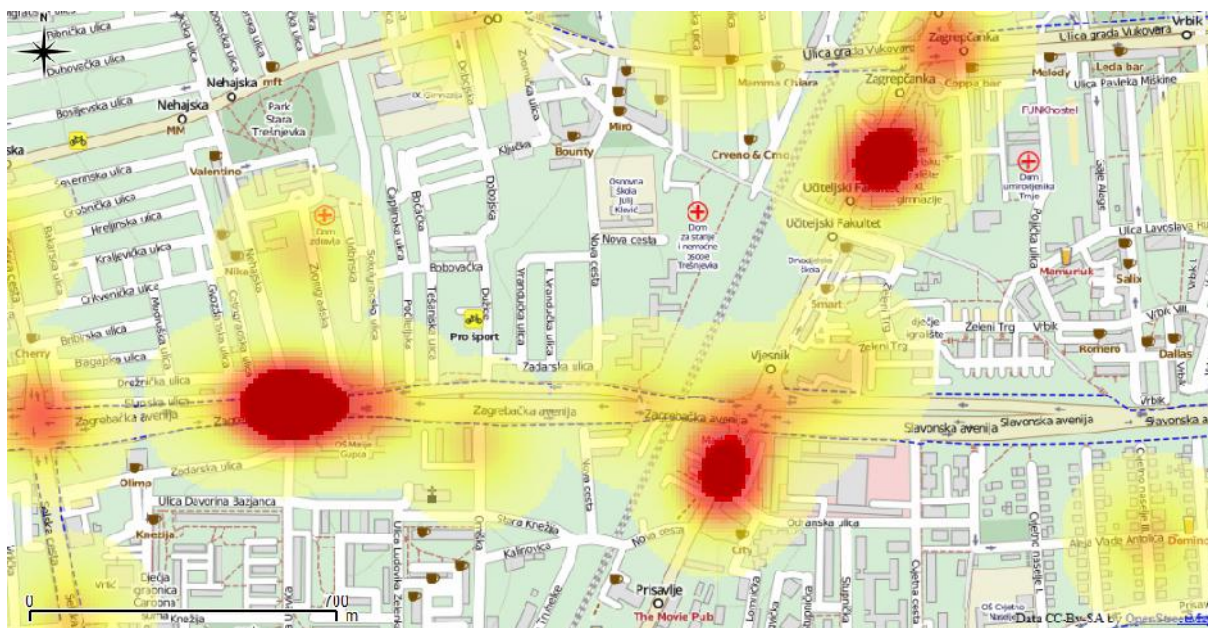
Slika 49 prikazuje tri žarišta. Prvo na kružnom toku gdje se spajaju Ulica Dragutina Domjanića, Petrova ulica, Voarska ulica i Srebrnjak (gore lijevo), drugo na Trgu Eugena Kvaternika (dolje lijevo), a treće na križanju Maksimirske ceste i Ulice Fakultetskog dobra (gore desno). Na gornjem lijevom žarištu prevladavaju nesreće uzrokovane brzinom neprimjerenom uvjetima, a veina se dogodila tijekom 2013. godine. Tijekom 2013. godine sve se nesreće dogodile na mokrom kolniku i u pretežitom većernjim satima. Na donjem lijevom žarištu prevladavaju nesreće uzrokovane pogreškama vozača. Dogodile su se tijekom dana i u stabilnim atmosferskim prilikama. Na gornjem desnom žarištu prevladavaju nesreće uzrokovane brzinom neprimjerenom uvjetima. Veina nesreća se dogodila tijekom dana na mokrom kolniku. Na toj lokaciji unutar zadnje dvije godine nije zabilježena nijedna nesreća ove vrste. Slika je izdvojena jer prikazuje utjecaj vremenskih uvjeta na prometne nesreće i prikazuje napredak u regulaciji prometa.

## 4.4.4. Prometne nesre e s poginulim i ozlije enim biciklistima



Slika 50. Prikaz žarišta prometnih nesre a s poginulim i ozlije enim biciklistima

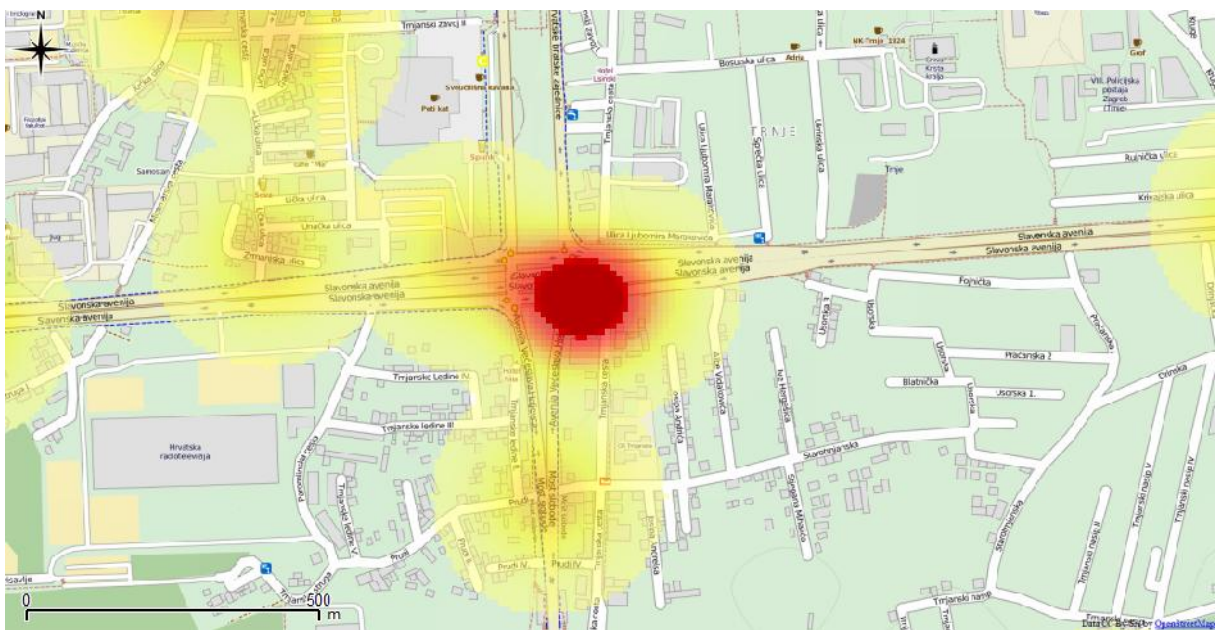
Druga najugroženija skupina sudionika u prometu su biciklisti. U nastavku se analiziraju prometne nesre e se poginulim i ozlije enim biciklistima. Treba napomenuti da je kao pozadina korišten OpenCycleMap OpenLayersa zbog kartiranih biciklisti kih staza (Slika 50). Slika služi za bolju orijentaciju pri analizi žarišta koja su analizirana u nastavku.



Slika 51. Prikaz žarišta prometnih nesre a s poginulim i ozlije enim biciklistima  
(donje lijevo žarište - 5 nesre a; donje desno žarište - 3 nesre e; gornje žarište - 4 nesre e)

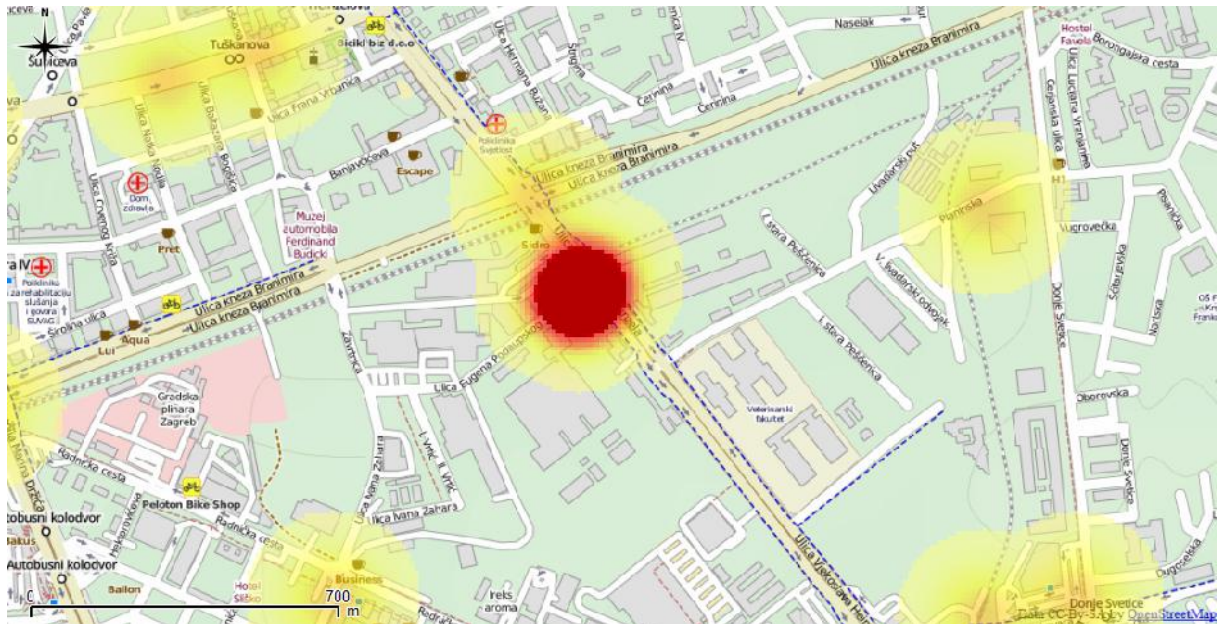


Slika 51 prikazuje tri žarišta. Prvo na križanju Zagrebačke avenije, Ulice Tina Ujevića i Nehajске ulice (dolje lijevo), drugo na križanju Savske ceste i Odranske ulice (dolje desno), a treće na Savskoj cesti kod Uiteljskog fakulteta (gore). Na donjem lijevom žarištu prevladavaju nesreće uzrokovane nepoštivanjem svjetlosnog znaka i nepoštivanjem prednosti prolaska. Većina nesreća se dogodila tijekom dana u stabilnim atmosferskim prilikama. Na donjem desnom žarištu nesreće nemaju sličnih faktora. Na gornjem žarištu prevladavaju nesreće uzrokovane nepropisnim uključivanjem u promet koje su se dogodile tijekom dana i u stabilnim atmosferskim prilikama. Slika je izdvojena iz razloga što prikazuje velika žarišta i time opasna područja za bicikliste.



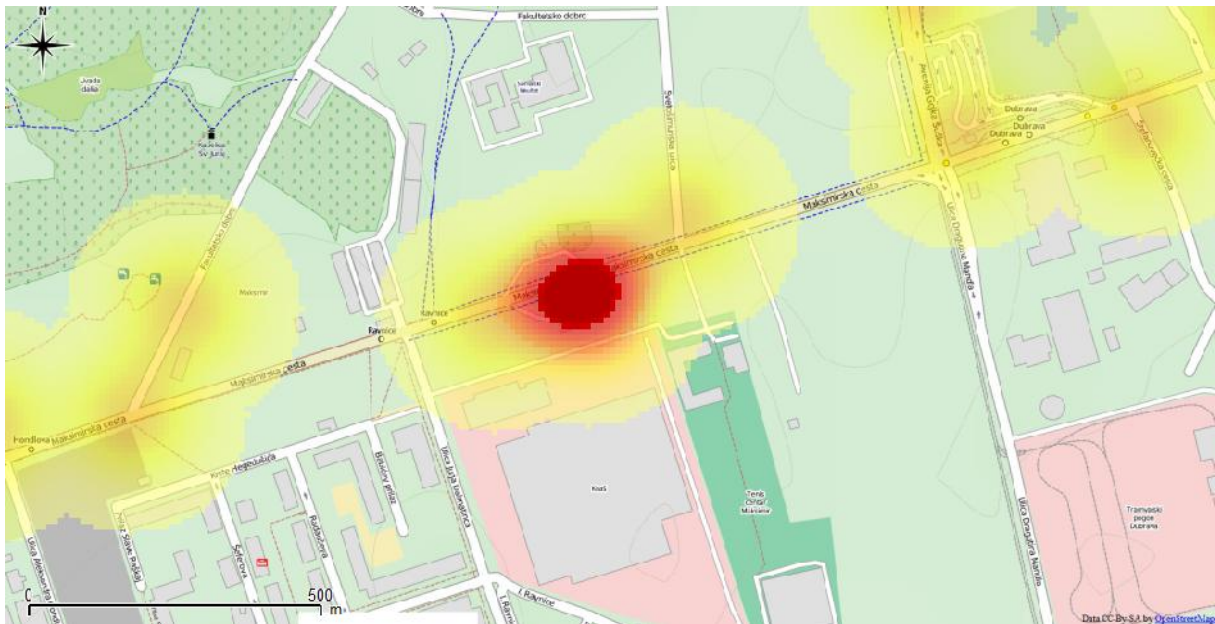
*Slika 52. Prikaz žarišta prometnih nesreća s poginulim i ozlijeđenim biciklistima  
(5 nesreća)*

Slika 52 prikazuje žarište na križanju Slavonske avenije i Avenija Ve eslava Holjevca. Na žarištu su se dogodile nesreće uzrokovane nepoštivanjem svjetlosnog znaka i nepropisnim uključivanjem u promet. Sve su se dogodile tijekom dana i u stabilnim atmosferskim prilikama. Slika je izdvojena zbog velike koncentracije nesreća.



*Slika 53. Prikaz žarišta prometnih nesre a s poginulim i ozlije enim biciklistima  
(6 nesre a)*

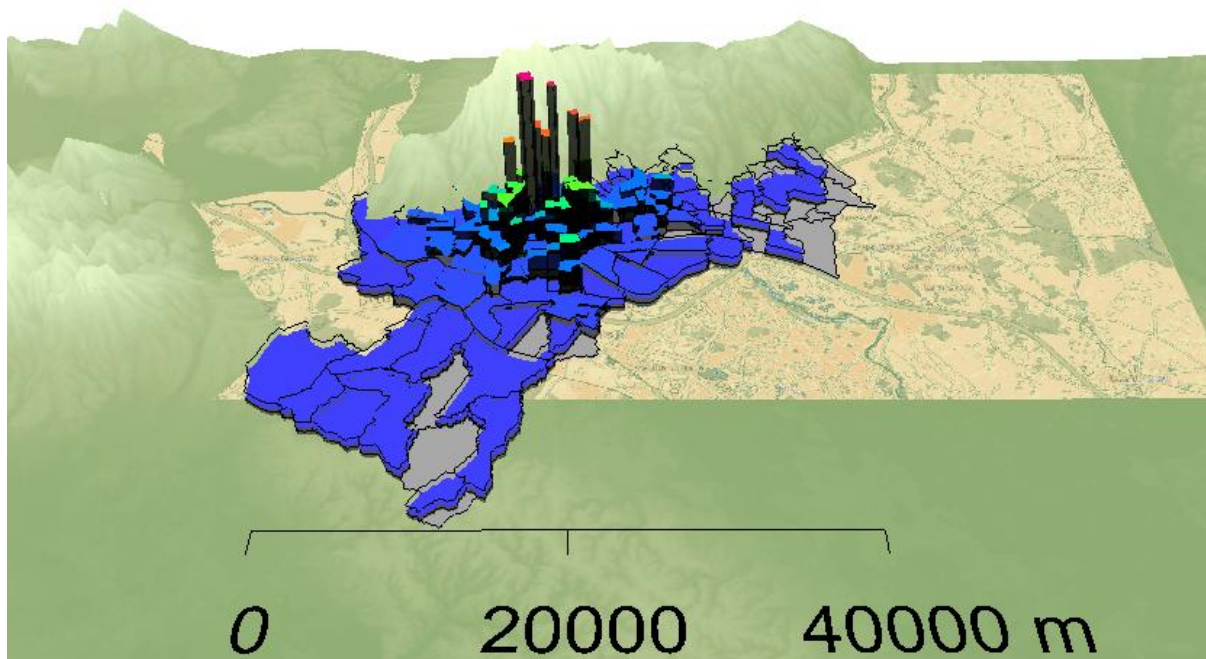
Slika 53 prikazuje žarište na križanju Ulice Vjekoslava Heinzela i Ulice Eugena Podaupskog. Ve ina nesre a je uzrokovana nepropisnim uklju ivanjem u promet. Sve su se dogodile tijekom dana, izme u svibnja i kolovoza te u stabilnim atmosferskim prilikama. Slika je izdvojena zbog velike koncentracije nesre a.



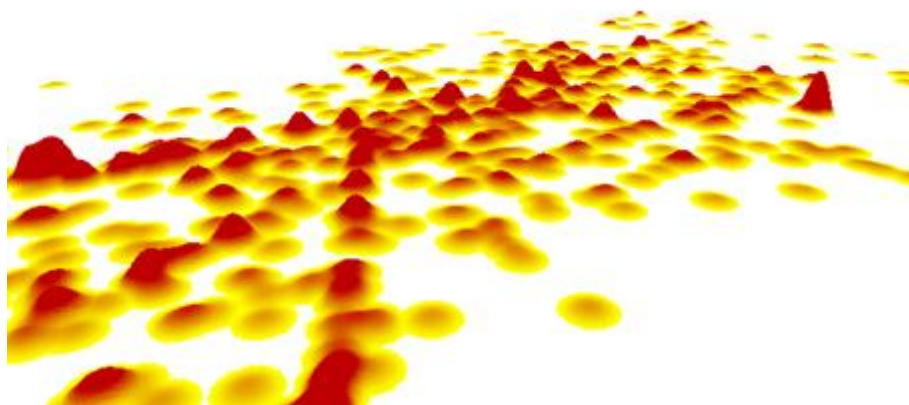
*Slika 54. Prikaz žarišta prometnih nesre a s poginulim i ozlije enim biciklistima  
(5 nesre a)*

Slika 54 prikazuje žarište na Maksimirskoj cesti južno od Šumarskog fakulteta. Sve su nesre e uzrokovane nepropisnim uklju ivanjem u promet. Dogodile su se u stabilnim atmosferskim prilikama, a ve ina se dogodila tijekom dana. Slika je izdvojena zbog velike koncentracije nesre a.

## 4.5. Dodatne mogućnosti vizualizacije rezultata



Slika 55. 3D prikaz distribucije gustoće prometnih nesreća s teško ozlijeđenim sudionicima



Slika 56. 3D prikaz žarišta prometnih nesreća s pješacima

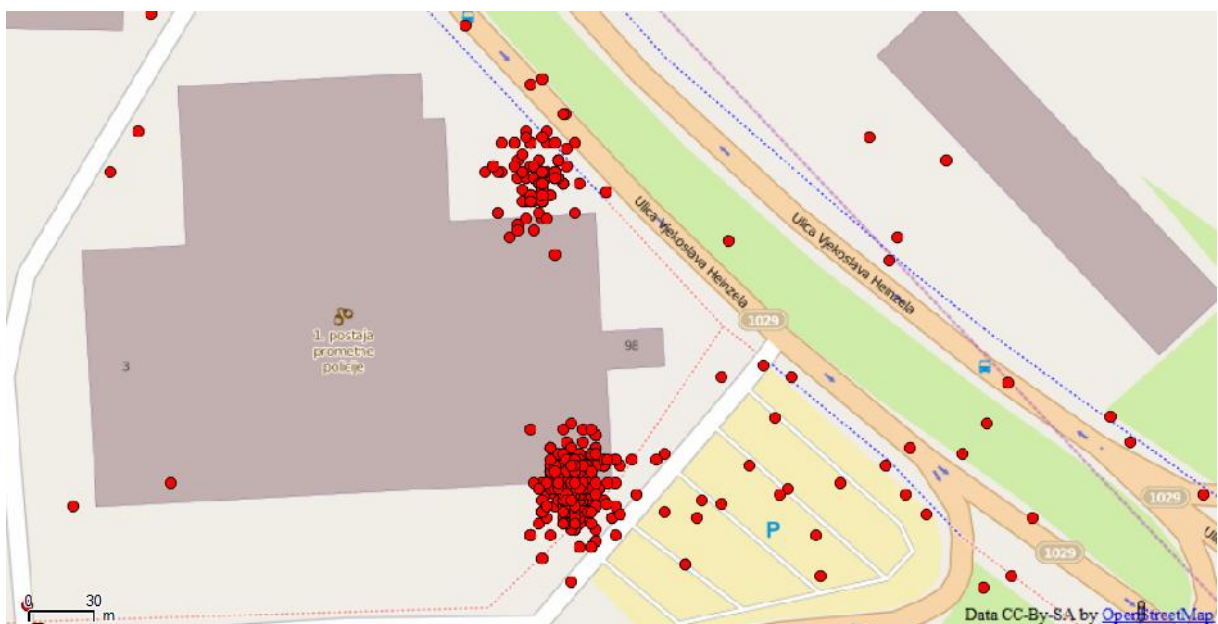
Slika 55 prikazuje 3D prikaz distribucije gustoće prometnih nesreća s teško ozlijeđenim sudionicima. Prikaz je izrađen u ILWIS aplikaciji. Slika 56 prikazuje 3D prikaz žarišta prometnih nesreća s pješacima izrađen u SAGA GIS aplikaciji. Ova dva primjera predstavljaju dodatne mogućnosti vizualizacije koji naglašavaju razlike između klasa i intenziteta žarišta.

## 4.6. Problemi uo eni tijekom obrade i analize podataka



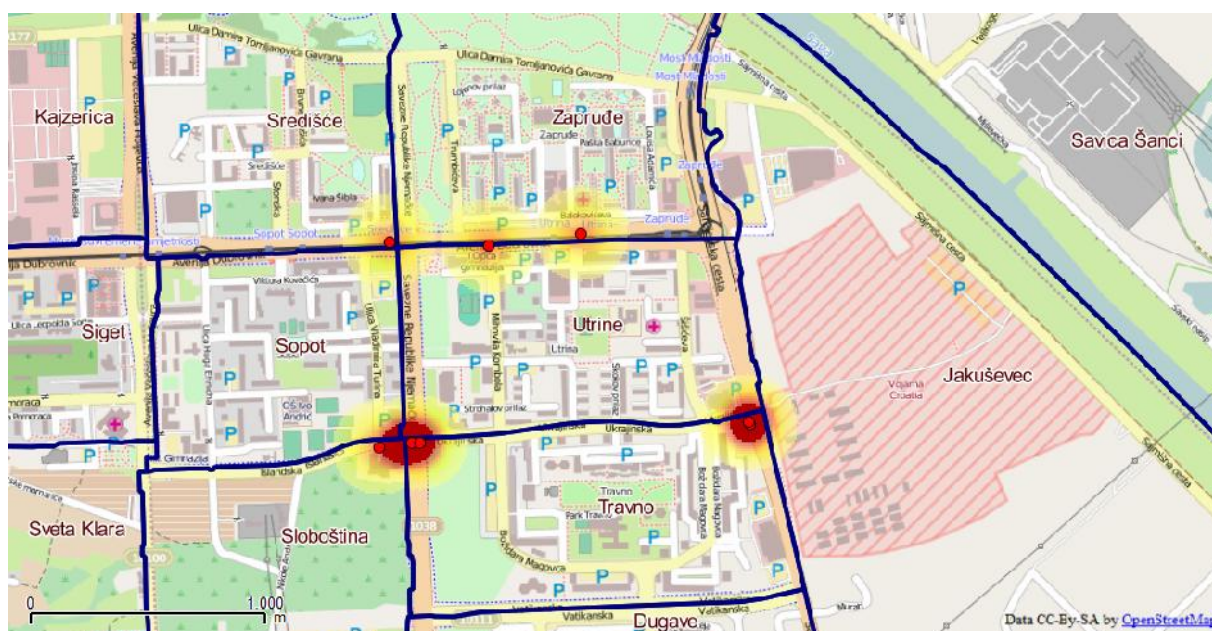
Slika 57. Prikaz loše pozicioniranih prometnih nesreća

Slika 57 predstavlja jedan od nedostataka fiksiranja GNSS prijmnika za policijsko vozilo. Prometne nesreće registrirane na autobusnom stajalištu su se zapravo dogodile na Rotoru. Zbog protoka prometa i sigurnosti ostalih sudionika nužno je isključiti nesrećena vozila iz prometa. U poglavlju *Rasprava* je ovaj problem detaljnije analiziran.

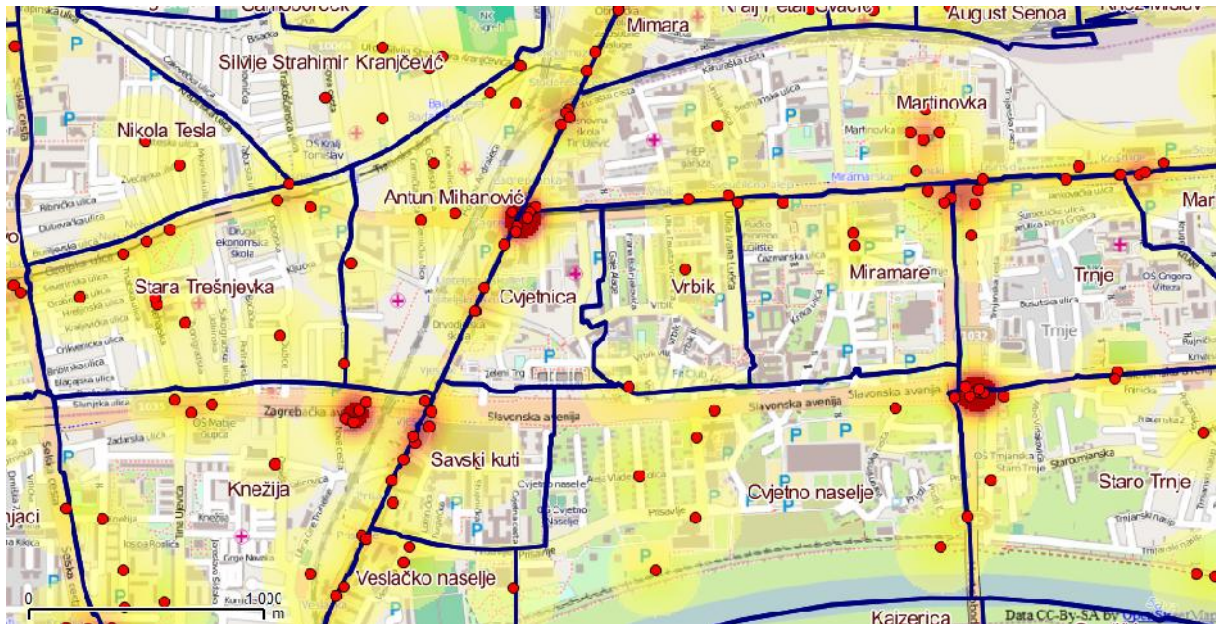


Slika 58. Prikaz pogrešno pozicioniranih prometnih nesreća

Slika 58 prikazuje krivo pozicionirane nesre e na 1. postaji prometne policije. Na toj je lokaciji zabilježeno oko 700 prometnih nesre a. Za ve inu nesre a se lako može odrediti da su krivo pozicionirane na osnovu atributa ulica, ali problem su nesre e koje su se dogodile u Ulici Vjekoslava Heinzela ili na Slavonskoj aveniji. Budu i da je rije o duga kim prometnicama, situacija je još gora ako ne piše ku ni broj. U tom slu aju postoji vjerojatnost da su nesre e koje su se dogodile na ili blizu križanja Ulice Vjekoslava Heinzela i Slavonske avenije dobro pozicionirane, ali se gube u oblaku krivo pozicioniranih to aka. S godinama se trend krivo pozicioniranih to aka na toj lokaciji smanjuje, s oko 350 nesre a u 2010. godini na 50 to aka u 2013. Slika prikazuje napredak MUP-a u popravljanju uo enih nedostataka.



*Slika 59. Prikaz prometnih nesre a s poginulim sudionicima pomo u to kastih oznaka unutar pojedinih mjesnih odbora Grada Zagreba*



*Slika 60. Prikaz prometnih nesreća s teško ozlijeđenim sudionicima pomoću točastih oznaka unutar pojedinih mjesnih odbora Grada Zagreba*

Slika 59 i Slika 60 prikazuju usporedbu metode normalizacije i KDE metode. Obojena mjesnih odbora su zanemarena zbog itljivosti slika. Treba obratiti pozornost na činjenicu da se granice mjesnih odbora poklapaju s prometnicama. Najvažnije je riječ o glavnim prometnicama na kojima se događaju najveći broj nesreća. U poglavlju *Rasprava* je ovaj problem detaljnije obrađen. Slika naglašava prednosti KDE metode nad normalizacijom.

## 5. Rasprava

Poučeni iskustvom istraživanja iz Hong Konga (Loo, 2006), prije obrade podataka je utvrđena kvaliteta i iskoristivost podataka. Kvaliteta zapisnika je daleko od idealne i najveći izazov predstavlja georeferenciranje. Napisan je algoritam u Pythonu koji ispravlja prostorne koordinate pomoću naziva ulica iz zapisnika. Algoritam se pokazao uspješnim u ovom zadatku. Pojavio se manji broj prometnih nesreća opterećen grubim pogreškama koje nisu mogle biti ispravljene. Riječ je o zanemarivom broju koji je isključen iz daljnje analize.

Nužno je otkriti uzroke loših zapisa prostornih koordinata. To ne uključuje samo pregled postojećih baza, nego i način unos podataka. Kao što je ranije navedeno, prometni službenici koriste GNSS prijavnike fiksirane u službenim vozilima. To se pokazalo kao loša praksa. Vozači moraju isključiti unesrećena vozila iz prometa i to ih udaljava od mjesta nesreće. Policijski službenici nisu u mogućnosti uzeti GNSS prijamnik i zabilježiti pravu lokaciju nesreće, već očitavaju položaj svog automobila nakon isključivanja iz prometa. To rezultira grubim pogreškama koje zahtijevaju ručno ispravljanje.

Sljedeći nedostatak je korištenje analognih zapisnika. Vozači moraju unijeti golemu količinu podataka u kratkom vremenu. Unos podataka u bazu podataka najčešće provodi osoba koja nije bila na mjestu nesreće i ne može provjeriti pouzdanost podataka. Ne postoji sustav provjere pouzdanosti podataka. Prostorne koordinate omogućuju vizualizaciju lokacije nesreće, a policijski službenici rade skice na mjestu nesreće. Logično rješenje je vizualizirati unesene podatke u GIS aplikaciji i usporediti ih sa skicom. U radu su navedene tri besplatne GIS aplikacije otvorenog koda koje su više nego dovoljne za ovaj jednostavan postupak.

Moguće rješenje je ukidanje dosadašnjeg načina rada i uspostava automatiziranog sustava registracije podataka. Također treba preispitati smisao atributa i način unos podataka. Nema potrebe opterećivati policijske službenike unosom prostornih koordinata i ostalih atributa koje sustav može u pozadini registrirati umjesto njih. Zapisnik trenutno sadrži ogromnu količinu atributa koji nisu prikladni za provođenje analiza. Važni atributi poput datuma i vremena nesreće su podijeljeni na više polja bez obzira što baza podataka ima posebne tipove podataka namijenjene za upravljanje vremenskim podacima. Sve aplikacije korištene u radu su besplatne i javno dostupne što znači da MUP može tijekom imenja rada započeti na uspostavi smislenog i kvalitetnog sustava.



Prva metoda statističke obrade podataka korištena u radu je hi-kvadrat test homogenosti. Provedena su tri testa s ciljem ispitivanja proporcionalnosti broja prometnih nesreća s različitim posljedicama po određenim atributima. Ti su atributi: karakteristike ceste, dijelovi dana te mjeseci u godini.

Vrlo bitan korak u provedbi testa je postavljanje nul-hipoteze. Kako je riječ o testovima homogenosti, sve su nul-hipoteze postavljene na sličan način. U svatkoj od tri slučaja nul-hipoteza nalaže da su omjeri broja posljedica prometnih nesreća jednaki raspoređeni s obzirom na analizirane attribute (karakteristike ceste, dijelovi dana i mjeseci). Temeljem rezultata svih triju testova, koji su detaljnije izloženi u poglavlju *Rezultati*, navedene nul-hipoteze su odbacene.

Prednost ove metode je mogućnost organiziranja velikog skupa podataka s brojnim atributima u tablice. Na temelju toga, može se dobiti podatak o mogućoj povezanosti atributa, ali ne i kojih.

Metoda ima mnogo nedostataka. Postoje određeni uvjeti, navedeni u poglavlju *Materijal i metode*, koji moraju biti ispunjeni kako bi metoda uopće bila primjenjiva. Metoda nije prikladna za detaljniju analizu prostorne distribucije prometnih nesreća jer ne omogućava vizualizaciju njihovog smještaja.

Iako se metodom ne može detaljnije analizirati prostorna distribucija prometnih nesreća, moguće je za attribute (skupine) pri provedbi testa izabrati varijable koje su povezane s prostorom i vremenom. U radu je upravo to i napravljeno. Na taj je način provedbom hi-kvadrat testova homogenosti ostvaren opći uvid u odnose između podataka.

Metoda normalizacije je uvedena kao primjer često korištene metodologije proučavanja prostorne distribucije prometnih nesreća [(LaScala et al., 2000), (Noland i Quddus, 2004), (Quddus, 2008), (Wang et al., 2009), (Wier et al., 2009)]. Riječ je o podjeli promatranog područja na prostorne jedinice i brojanju opažanja unutar pojedine jedinice. Suma opažanja unutar jedinice može se podijeliti s ukupnim brojem opažanja ili s površinom jedinice na koju se odnosi. Slični se rezultati mogu postići podjelom promatranog područja na pravokutnu mrežu gdje svaka ćelija predstavlja određenu prostornu jedinicu.

Ukoliko su dostupni podaci iz drugih izvora, mogu se uključiti u analizu s ciljem identifikacije prostornih korelacija. Primjerice, podaci iz popisa stanovništva omogućili bi usporedbu određene vrste prometne nesreće s demografskim karakteristikama prostorne

jedinice. U radu su kao prostorne jedinice korišteni mjesni odbori i samo podaci preuzeti od MUP-a.

Kao što je obrađeno u poglavlju *Materijal i metode*, autori Smith, Goodchild i Longley (Smith et al., 2007) su istaknuli određene nedostatke ove metode. Među njima su naveli da treba uzeti u obzir podatke koji nedostaju te da korištenje neprikladnog djelatnika ili podataka neizvjesne vrijednosti može dovesti do nesuvislih rezultata. Ovi su problemi vezani za metode koje vrše podjelu prostora na diskretne jedinice i ovisi o veličini, obliku i smještaju prostorne jedinice. Pri definiciji ovih parametara treba uzeti u obzir gustoću i raspored opažanja.

Rezultati ove metode iz poglavlja *Rezultati* prikazuju da u svakom slučaju postoje mjesni odbori u kojima se nije dogodila nijedna nesreća. U našem slučaju granice mjesnih odbora prate velike prometnice. Ako se većina nesreća događa na tim prometnicama, to nije povoljno za analizu. U tom se slučaju dio nesreća pridružuje jednoj prostornoj jedinici, a ostatak nesreća drugoj. Utjecaj tih nesreća se širi na cijelu prostornu jedinicu bez obzira što je riječ o opažanjima na granici jedinice. To onemogućuje identifikaciju istinski opasnih lokacija. Slika 59 predstavlja primjer ovog slučaja. Velik broj nesreća se dogodio na raskrižju koje se nalazi na granici dva odbora, ali sve su pridružene jednom odboru dok je drugi 'prazan'. Prvi odbor ocjenjuje iznimno opasnim, a drugi kao siguran.

U radu je korištena površina mjesnog odbora kao djelatnik. Također se kao djelatnik mogao koristiti ukupni broj prometnih nesreća, ali to onda drastično mijenja konačne rezultate. Izbor djelatnika je subjektivan i ovisi o osobi koja provodi analizu i o tome što želi istaknuti.

Ukoliko je neka prostorna jedinica ocijenjena kao iznimno opasna, a sadrži velik broj prometnica, nije poznato koje su prometnice zaista opasne. To iziskuje ručno pregledavanje podataka što je dug i mukotrpan posao podložan pogreškama.

Da bi se izbjegli navedeni nedostaci, zadnjih se godina prometne nesreće obrađuju identifikacijom žarišta [(Anderson, 2009), (Depaire et al., 2008), (Erdogan et al., 2008), (Prasannakumar et al., 2011)]. U radu je obrađena KDE metoda koja daje izvrsne rezultate. Postoje određeni nedostaci, ali prednosti su velike.

Od nedostataka treba istaknuti da metoda zbog velikog broja računskih operacija nije pogodna za ručno radovanje, posebno ako se uzme u obzir količina podataka za obradu. Duljina trajanja izračuna raste primjenom složenijih funkcija. Rezultat izračuna je rasterska datoteka. Velikina datoteke ovisi o parametrima definiranim prije izvršavanja operacije. To se odnosi na

veličinu piksela i zadanu širinu područja. Ovisno o skupu podataka koji se obrađuje, veličina izlaznih rastera može zauzeti i do 5 GB memorije na računalu.

Prednosti ove metode su zaista neizmjerne. Ne postoje hipoteze koje se moraju postaviti prije obrade podataka već se samo uzima u obzir distribucija prostornih podataka. Time je metoda neovisna o korisniku i daje objektivnije rezultate. Za razliku od normalizacije, nije potrebna prethodno definirana podjela promatranog područja na prostorne jedinice ili ćelije. Time se izbjegavaju nedostaci povezani s dimenzijama ćelija i njihovim rasporedom.

Kao što je navedeno u poglavlju *Materijal i metode*, metoda obrađuje točku po točku. Samo se analizira prostorna distribucija opažanja, a rezultati najviše ovise o izboru funkcije i definiranoj širini područja od strane korisnika. Metoda naglašava samo područja s velikim koncentracijama opažanja, odnosno žarišta. Teoretski, žarište je mreža ćelija položena preko podataka. Ali za razliku od metoda koje prije obrade definiraju mrežu koja pokriva cijelo područje, ova mreža pokriva samo dio gdje se nalazi žarište. Time se u jednoj datoteci dobije nekoliko mreža koje samo ovise o parametrima definiranim od strane korisnika. Vrijednosti ćelija tih mreža se popunjavaju vrijednostima koje najbolje prikazuju rezultate metode.

Analiza rezultata je jednostavno izvediva u GIS-u. Nesreće se pohranjuju u vektorskom sloju kao točke, a žarišta u rasterskom sloju. Rasterski sloj prikazuje opasne lokacije, a preklapanjem njega s vektorskim slojem vidi se koje nesreće pripadaju određenim žarištima. Označavanjem određenog žarišta označavaju se nesreće koje mu pripadaju. One se mogu izdvojiti za daljnju analizu. Nad atributnim podacima izdvojenih nesreća mogu se provoditi analize koje omogućuju identifikaciju uzroka nesreća na određenim lokacijama.

## 6. Zaključci

Cilj rada je uspostava metodologije obrade i analize prometnih nesreća. Po našim saznanjima, ovo je prva primjena prostorne statistike nad ovim skupom podataka u Republici Hrvatskoj. Korištene su besplatne aplikacije otvorenog koda za obradu podataka i predstavljene su statističke metode koje omogućuju analizu ovog skupa podataka. Istaknuti su nedostaci postojećeg sustava i predložena potencijalna rješenja.

Neophodno je na početku provesti ispitivanje kvalitete dostupnih podataka. To se najviše odnosi na kvalitetu prostornih koordinata prometnih nesreća, ali i na attribute poput naziva ulica iz zapisnika. S ciljem provođenja kvalitetnih analiza, važno je uključiti u obradu samo one podatke koji su to nijim podacima.

Naglasak rada je na primjeni prostorne statistike u analizi distribucije prometnih nesreća. Detaljno su obrađene sve korištene statističke metode tako da korisnici mogu odabrati metodu koja im najbolje odgovara. Ukoliko korisnici ne znaju odakle krenuti s analizom, hi-kvadrat testovi predstavljaju dobru polaznu osnovu, ali izostaje vizualizacija prostornog smještaja prometnih nesreća. Ako je potrebno analizirati podatke na većem području, metoda normalizacije daje zadovoljavajuće rezultate. Daleko najbolje rezultate pruža KDE metoda. U relativno kratkom vremenu mogu se identificirati žarišta vrsta nesreća koja korisnici zahtijevaju. Ne zahtijeva ulazne pretpostavke i hipoteze te ne uključuje podjelu područja od interesa na prostorne jedinice. U potpunosti se oslanja na prostornu distribuciju prometnih nesreća, a korisnici definiraju kvalitetu rezultata. To im je omogućeno kroz izbor funkcija, veličine piksela i definiranje širine područja.

Analize ovog tipa su neopisivo važne za društvo. U konačnici je riječ o ljudskim životima, povećanju sigurnosti u prometu i smanjenju materijalnih troškova koje donose prometne nesreće. Nadamo se da će ovaj rad poslužiti kao polazna osnova za daljnja istraživanja i doprinijeti poboljšanju sigurnosti u prometu.

## Zahvale

Htjeli bismo iskoristiti ovu priliku i zahvaliti se mentoru prof.dr.sc Damiru Medaku, dipl. ing. geod. i asistentu Mariju Mileru, dipl.ing geod., bez kojih izrada ovog rada ne bi bila moguća. Zahvaljujemo im se na prilici i nesebičnoj pomoći koju su nam pružili, na znanju koje su nam prenijeli te na strpljenju koje su nam pokazali tijekom proteklih mjeseci. Njihova otvorenost i predanost odabranoj profesiji je vrijedna svake pohvale i treba poslužiti kao primjer ne samo predavačima već i studentima. Tijekom izrade rada susreli smo se s mnogim preprekama koje su nam u tom trenutku djelovale nepremostivo i za njihovo svladavanje nismo imali dovoljno iskustva. Njihovo je iskustvo neprocjenjivo, ali su nam samo predstavljali potencijalna rješenja koja smo sami morali istražiti. To je u konačnici rezultiralo mnogim pogreškama, ali su imali dovoljno razumijevanja i vjere da ćemo s vremenom savladati sve prepreke.

Nadalje, htjeli bismo se zahvaliti asistentu s Fakulteta prometnih znanosti, mr. sc. Marku Ševroviću, dipl. ing. prom., čija su nam iskustva i preporuke uvelike olakšale obradu podataka i dale uvid u stvarno stanje. Zahvaljujemo mu se na pruženoj prilici i podacima koje nam je proslijedio. Nadamo se da će rezultati ovog rada još više zbližiti naše struke i omogućiti donošenje kvalitetnijih zaključaka po pitanju sigurnosti u prometu. Također se zahvaljujemo asistentu dr.sc. Ivanu Medvedu, dipl.ing.geod., čije je znanje iz domene prostorne statistike uvelike olakšalo analizu danih podataka.

Na kraju bismo se htjeli zahvaliti asistentu Draženu Odobašiću, dipl. ing. geod. te kolegama Franu Glasinoviću, mag. ing. geod. et geoinf., Franu Periću, mag. ing. geod. et geoinf., i Mauriciju Vidulinu, univ. bacc. ing. geod. et geoinf., na tehničkoj podršci tijekom obrade podataka i moralnoj podršci koju su nam pružili tijekom izrade rada.

Želimo naglasiti da nam izrada ovog rada predstavlja jedno neprocjenjivo iskustvo koje nam je otvorilo brojne mogućnosti i nepovratno proširilo vidike. Suraivali smo s divnim ljudima, koji su spremni podijeliti svoje znanje i vrijeme s neiskusnim, ali ambicioznim studentima. Nadamo se da ćemo jednog dana biti u stanju pružiti drugima priliku koja je svojevremeno pružena nama te da ćemo pokazati kvalitete koje smo susreli u našoj okolini.

## Popis literature

- Abdel-Aty, M. A., Radwan, A. E. (2000): Modeling traffic accident occurrence and involvement, *Accident Analysis and Prevention* 32, 633–642.
- Al-Ghamdi, A. S. (2003): Analysis of traffic accidents at urban intersections in Riyadh, *Accident Analysis and Prevention* 35, 717–724.
- Anderson, T. K. (2009): Kernel density estimation and K-means clustering to profile road accident hotspots, *Accident Analysis and Prevention* 41, 359–364.
- Bard, G. V. (2007): Spelling-error tolerant, order-independent pass-phrases via the Damerau-Levenshtein string-edit distance metric, *ACSW '07 Proceedings of the fifth Australasian symposium on ACSW frontiers* 68, 117-124.
- Burt, J. E., Barber, G. M. (1996): *Elementary Statistics for Geographers*, The Guilford Press, New York, New York, United States of America.
- Cheng, W., Washington, S. P. (2005): Experimental evaluation of hotspot identification methods, *Accident Analysis and Prevention* 37, 870–881.
- Depaire, B., Wets, G., Vanhoof, K. (2008): Traffic accident segmentation by means of latent class clustering, *Accident Analysis and Prevention* 40, 1257–1266.
- Erdogan, S., Yilmaz, I., Baybura, T., Gullu, M. (2008): Geographical information systems aided traffic accident analysis system case study: city of Afyonkarahisar, *Accident Analysis and Prevention* 40, 174–181.
- Gibson, H. R., Dillard, B. L. (2012): *Elementary Statistics*, Kendall Hunt Publishing Company, Dubuque, Iowa, United States of America.
- Guidoum, A. C. (2013): Kernel Estimator and Bandwidth Selection for Density and its Derivatives, *The kedd Package* 1, 1-3.
- Jones, A. P., Langford, I. H., Bentham, G. (1996): The application of K-function analysis to the geographical distribution of road traffic accident outcomes in Norfolk, England, *Social Science & Medicine*, 879-885.

- LaScala, E. A., Gerber, D., Gruenewald, P. J. (2000): Demographic and environmental correlates of pedestrian injury collisions: a spatial analysis, *Accident Analysis and Prevention* 32, 651–658.
- Levine, N. (2013): Kernel Density Interpolation, *CrimeStat IV Documentation*, Houston, Texas, United States of America, 5-22.
- Loo, B. P. (2006): Validating crash locations for quantitative spatial analysis: A GIS-based approach, *Accident Analysis and Prevention* 38, 879–886.
- Neis, P., Zipf, A. (2012): Analyzing the Contributor Activity of a Volunteered Geographic Information Project - The Case of OpenStreetMap, *ISPRS International Journal of Geo-Information*, 146-165.
- Noland, R. B., Quddus, M. A. (2004): A spatially disaggregate analysis of road casualties in England, *Accident Analysis and Prevention* 36, 973–984.
- Prasannakumar, V., Vijith, H., Charutha, R., Geetha, N. (2011): Spatio-Temporal Clustering of Road Accidents: GIS Based Analysis and Assessment, *Procedia Social and Behavioral Sciences* 21, 317–325.
- Quddus, M. A. (2008): Modelling area-wide count outcomes with spatial correlation and heterogeneity: An analysis of London crash data, *Accident Analysis and Prevention* 40, 1486–1497.
- Smith, M. J., Goodchild, M. F., Longley, P. A. (2007): *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools*, Matador (Troubador Publishing Ltd.), Leicester, United Kingdom.
- Tegge, R., Ouyang, Y. (2009): Correcting erroneous crash locations in transportation safety analysis, *Accident Analysis and Prevention* 41, 202–209.
- Wang, C., Quddus, M., Ison, S. (2009): The effects of area-wide road speed and curvature on traffic casualties in England, *Journal of Transport Geography*, 385–395.
- Wier, M., Weintraub, J., Humphreys, E. H., Seto, E., Bhatia, R. (2009): An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning, *Accident Analysis and Prevention* 41, 137–145.

URL 1: World Health Organization,

[http://www.who.int/gho/road\\_safety/en/index.html](http://www.who.int/gho/road_safety/en/index.html), (8.1.2014.)

URL 2: OpenStreetMap Wiki,

[http://wiki.openstreetmap.org/wiki/Main\\_Page](http://wiki.openstreetmap.org/wiki/Main_Page), (4.2.2014.)

URL 3: OpenStreetMap: Data Model, Crowdsourcing, and Technology,

<http://www.applied-geoinformatics.org/index.php/agse/agse2010/paper/view/181/115>,  
(16.4.2014.)

URL 4: Python Programming Language,

<http://www.python.org/>, (3.2.2014.)

URL 5: PostgreSQL,

<http://www.postgresql.org/>, (3.2.2014.)

URL 6: BoundlessGeo,

<http://boundlessgeo.com/>, (4.2.2014.)

URL 7: Chi-Square Analysis: The CHI-square Test of Independence,

[http://www.pindling.org/Math/Statistics/Textbook/Chapter11\\_Chi\\_Square/independence.html#contingency%20table](http://www.pindling.org/Math/Statistics/Textbook/Chapter11_Chi_Square/independence.html#contingency%20table) (10.4.2014.)

URL 8: Stat Trek: Chi Square Test of Homogeneity,

<http://stattrek.com/chi-square-test/homogeneity.aspx> (10.4.2014.)

URL 9: Geographic Information Technology Training Alliance,

<http://www.gitta.info/Statistics/en/html/index.html>, (26.2.2014.)

URL 10: iPredict: Kernel Smoothing,

<http://www.ipredict.it/Methods/KernelSmoothing.aspx>, (25.2.2014.)



URL 11: StatsBlogs: Exploratory Data Analysis: Kernel Density Estimation,

<http://www.statsblogs.com/2013/06/09/exploratory-data-analysis-kernel-density-estimation-in-r-on-ozone-pollution-data-in-new-york-and-ozonopolis/>, (25.2.2014.)

URL 12: Data Science Rules: Classification of Hand-written Digits,

<http://datasciencrules.blogspot.com/2012/10/classification-of-hand-written-digits-4.html>, (25.2.2014.)

URL 13: SAS Institute: The KDE Procedure,

[http://support.sas.com/documentation/cdl/en/statug/63347/HTML/default/viewer.htm#kde\\_toc.htm](http://support.sas.com/documentation/cdl/en/statug/63347/HTML/default/viewer.htm#kde_toc.htm), (20.2.2014.)

URL 14: QGIS,

<http://www.qgis.org/en/site/>, (4.2.2014.)

URL 15: OpenLayers: Free Maps for the Web,

<http://openlayers.org/> (15.4.2014.)

URL 16: SAGA: System for Automated Geoscientific Analysis,

<http://www.saga-gis.org/en/index.html> (9.4.2014.)

URL 17: ILWIS, the free user-friendly raster and vector GIS,

<http://www.ilwis.org/index.htm> (9.4.2014.)

## Sažetak

Tomislav Šimunovi , Filip Todi

*Geostatisti ka analiza prostorne distribucije prometnih nesre a na podru ju Grada Zagreba u razdoblju od 2010. do 2013. godine*

Prometne nesre e su jedan od svakodnevnih problema suvremenog društva. Identifikacija žarišta prometnih nesre a je klju na pri provo enju u inkovitih strategija smanjenja podru ja s visokom gusto om prometnih nesre a. U radu su obra ene prometne nesre e na podru ju Grada Zagreba za razdoblje od 2010. do 2013. godine. Prometna mreža je preuzeta sa slobodne karte svijeta OpenStreetMap, a podaci iz zapisnika prometnih nesre a su obra eni u programskom jeziku Python. Posebna pozornost je posve ena provjeri kvalitete dostupnih prostornih podataka te otkrivanju i uklanjanju pogrešaka. Obra eni su podaci uneseni u sustav za upravljanje objektno-relacijskim bazama podataka PostgreSQL koji uklju uje proširenje za geoprostorne podatke i operacije - PostGIS. Nakon provjere to nosti dostupnih podataka testirane su tri naj eš e korištene statisti ke metode za ovu vrstu opažanja: hi-kvadrat test, normalizacija i Kernel Density Estimation. U radu se detaljno obra uju prednosti i nedostaci svake od metoda i predlaže model geostatisti ke analize prometnih nesre a. Za podru je Grada Zagreba precizno su odre ena kriti na mjesta na kojima je u predmetnom razdoblju naj eš e dolazilo do nesre a s teškim posljedicama, s naglaskom na posebno ugrožene sudionike u prometu – pješake i bicikliste. Model e poslužiti kao temelj za daljnja istraživanja i kao primjer Ministarstvu unutarnjih poslova Republike Hrvatske za identifikaciju žarišta prometnih nesre a s ciljem provo enja kvalitetnijih protumjera.

*Klju ne rije i: žarišta prometnih nesre a, OpenStreetMap, PostgreSQL, geostatistika*

## Summary

Tomislav Šimunovi , Filip Todi

*Geostatistical analysis of traffic accident spatial distribution in the city of Zagreb from 2010 to 2013*

Traffic accidents are an important problem of contemporary society. The identification of traffic accident hotspots is crucial when it comes to administering effective strategies pertaining to areas with a high density of traffic accidents. The data set of traffic accidents that occurred in the city of Zagreb from 2010 to 2013 was processed in this study. Roadway data was obtained from the OpenStreetMap project and the traffic accidents data set was obtained from the Ministry of Interior of the Republic of Croatia. Both data sets were processed in the Python programming language. The emphasis is put on the available spatial data quality check, error detection and removal processes. The processed data was imported into the PostgreSQL object-relational database management system that incorporates the PostGIS spatial extension. After reviewing the accuracy of the available data set, three common statistical methods were tested: Chi-square test, Normalization and Kernel Density Estimation. Pros and cons are established for each method and a geostatistical model for processing traffic accidents is proposed. Traffic accidents hotspots in the city of Zagreb are precisely determined, especially for the vulnerable traffic participants like pedestrians and cyclists. This model will serve as a viable basis for future research as well as an example to the Ministry of Interior of the Republic of Croatia for the identification of traffic accidents hotspots in hopes of administering effective countermeasures.

*Key words: traffic accident hotspots, OpenStreetMap, PostgreSQL, geostatistics*