

SVEU ILIŠTE U ZAGREBU

GEODETSKI FAKULTET

Tomislav Šimunovi i Filip Todi

Geostatisti ka analiza prostorne distribucije prometnih nesre a na
podru ju 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.	Opći i specifični 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 pomoću programskog jezika Python	11
3.3.	Unos i obrada podataka pomoću PostgreSQL-a i PostGIS-a	16
3.4.	Analiza podataka pomoću hi-kvadrat testa.....	18
3.5.	Analiza podataka metodom normalizacije	20
3.6.	Analiza podataka statističkom metodom Kernel Density Estimation	24
3.6.1.	Kernel Density Estimation (KDE)	26
3.6.2.	Određivanje 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 statističkom metodom normalizacije	56
4.4.	Rezultati obrade statističkom metodom KDE	61
4.4.1.	Prometne nesreće s poginulim sudionicima	61
4.4.2.	Prometne nesreće s poginulim i ozlijednim sudionicima	64
4.4.3.	Prometne nesreće s poginulim i ozlijednim pješacima	72
4.4.4.	Prometne nesreće s poginulim i ozlijednim biciklistima	76
4.5.	Dodatne mogunosti 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 stopu smrtnosti. Na svjetskim cestama 50% poginulih su posebno ranjivi sudionici u prometu (23% motociklisti, 22% pješaci, 5% biciklisti). Ozljede dobivene u prometnim nesrečama su osmi vodeći uzrok smrти i prvi uzrok smrти među osobama od 15. do 29. godine, što predstavlja važan globalni zdravstveni problem (URL 1).

Od 194 zemalja koje su sadržane u statistici, 64% ih ima ograničenje brzine 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 nesreča je u porastu na globalnoj razini iz razloga što prometna infrastruktura ne prati razvoj ostalih sektora poput industrije i razvoja stambenih naselja. Prometne nesreče su među glavnim uzrocima smrти na globalnoj razini, a većina nesreča je rezultat ljudskih pogrešaka, najčešće nemara vozača ili pješaka. Vjerovatnost prometnih nesreča i njihova težina mogu biti smanjeni sustavnom analizom scenarija nesreča te korištenjem odgovarajućih rješenja koja uključuju primjenu potrebnih uređaja za kontrolu prometa, prikladnim dizajnom prometnika i u inkovitim intervencijama prometne policije. Donošenje u inkovitim rješenjima iziskuje analizu prostornih i vremenskih obrazaca na mjestima prometnih nesreča što se postiže primjenom geoprostornih tehnologija (Prasannakumar i dr. 2011).

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

popravljanja istih. Problem nastaje kada nije mogu e pouzdano razlikovati mjesta koja su zaista visokorizi na od mjesta koja slu ajno imaju zabilježen ve i broj prometnih nesre a u promatranom razdoblju. Zbog oscilacije broja prometnih nesre a iz godine u godinu, metode koje se oslanjaju na jednostavno pra enje u estalosti nesre a pove avaju rizik od velikog broja tzv. lažnih pozitivnih nesre a'. To vodi do mogu eg poboljšanja sigurnosnih uvjeta na relativno sigurnim lokacijama, a otvara mogu nost previda zaista opasnih lokacija što u kona nici rezultira lošim ulaganjima u unapre enje prometne infrastrukture (Cheng, Washington 2005).

Iz tog je razloga klju na identifikacija žarišta prometnih nesre a. Cilj je razvoj i provo enje u inkovitih sigurnosnih strategija na podru jima s velikom gusto om prometnih nesre a. U tom su procesu važni atributni podaci pridruženi prometnim nesre ama. Oni omogu uju bolje razumijevanje procesa koji se odvijaju na žarištima što vodi kvalitetnijoj raspodjeli resursa namijenjenih poboljšanju sigurnosti u prometu. Literatura o prometnim nesre ama ne pruža univerzalno prihva enu definiciju žarišta prometnih nesre a, što prisiljava stru njake da ih rangiraju na osnovi težine ili u estalosti nesre a. Analiza žarišta tradicionalno se koncentrira na cestovne segmente ili na specifi na križanja te zahtijeva opsežno razumijevanje stupnja uklju enosti vozila u proces, težinu ozljeda i okolinu. GIS (*eng. Geographic information system*) aplikacije su prakti ne za ovu vrstu problema jer omogu uju u inkovitu manipulaciju, analize i vizualizaciju prostornih podataka. Identifikacijom žarišta prometnih nesre a u GIS-u može se ste i potpunije razumijevanje uzroka nesre a. Žarišta prometnih nesre a otkrivaju koncentracije istovrsnih nesre a i ukazuju na prostornu ovisnost izme u individualnih slu ajeva koji mogu imati zajedni ke uzroke (Anderson 2009).

GIS tehnologija je popularan alat za vizualizaciju podataka prometnih nesre a i analizu žarišta te važan upravlja ki alat za sigurnost u prometu. Omogu ava manipulaciju velike koli ine podataka koja se može pohraniti, dijeliti i obra ivati. Pruža platformu za analizu i vizualizaciju podataka s ciljem istraživanja odnosa izme u njih te može pružiti grafi ke i negrafi ke rezultate. Važnost analize prostornih podataka leži u atributima i njihovim vezama, uzimaju i pritom u obzir prostornu lokaciju promatrane pojave. Informacijski sustavi prometnih nesre a zasnovani na GIS-u mogu uo iti veze izme u prostornih pojava koje su skoro neodredive s neprostornom bazom podataka. Prostorni podaci dobiveni GIS-om i prostornim analizama su jedni od najvažnijih izvora informacija pri analizi prometnih nesre a jer pružaju stru njacima mnoštvo informacija o opasnim lokacijama, žarištima, itd. Stru njaci pomo u GIS-a mogu spojiti i geokodirati podatke nesre a i prometnica, izra unati u estalost

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

Identifikacija opasnih lokacija prostornim analizama pomoći u GIS-u doprinosi smanjenju broja prometnih nesreća, ali uspjeh tih analiza uvelike ovisi o točnosti, pouzdanosti i opsežnosti zapisnika prometne nesreće, posebno pri unosu i obradi podataka. Analiza prometnih nesreća jako je složena tema zbog brojnih faktora koji prostorno utječu na nesreće i stoga zapisnici moraju biti detaljni i prikladno strukturirani za prostornu i statistiku analizu. Umjesto klasičnog analognog zapisnika, GNSS (*eng. Global Navigation Satellite System*) i GIS moraju se integrirati u novi sustav zapisnika čime se uklanjuju mnoge sistematske i slučajne pogreške te u isto vrijeme povećava prostorna točnost. Uključivanje multimedijskih mogućnosti (digitalne slike i video zapisi) dodatno poboljšava učinkovitost sustava (Erdogan i dr. 2008).

S prostornog gledišta, u analizi prometne sigurnosti cestovna mreža se dijeli na individualna mjesta poput raskrižja, homogenih segmenata ili koridora. Statistički modeli se razvijaju kako bi se odredila frekvencija prometnih nesreća na određenim mjestima. Rezultati modela mogu ujuti prometnim institucijama lociranje mesta s učestalim prometnim nesrećama u cestovnoj mreži i uočavanje značajnih faktora koji doprinose nesrećama. Cilj tog procesa je poboljšanje prometne sigurnosti pomoći u inženjerskih i zakonskih mjerama te dodatnog obrazovanja. Razvoj modela zahtjeva više tipova podataka o prometnoj nesreći. Najčešće je riječ o vrsti nesreće i lokaciji na kojoj se dogodila. Različite vrste sirovih podataka moraju se vremenski i prostorno povezati sa svakim mjestom kako bi se dobili ulazni podaci za ovisne varijable, primjerice broj prometnih nesreća određenog tipa na nekoj lokaciji u određenom vremenskom razdoblju te opisne varijable, primjerice duljina segmenta, gustoća prometa ili vremenske prilike. Točnost prostornih informacija o nesrećama i točnost prometne mreže ključna je za proces povezivanja podataka, stoga pogreške i nekonzistentnosti u zapisnicima prometnih nesreća predstavljaju realan problem koji estično umanjuje točnost rezultata u obliku sigurnosnog modela (Tegge, Ouyang 2009).

Podatke vezane za prometne nesreće rijetko kad prikuplja jedna ustanova, već su različiti tipovi podataka prikupljani i održavani od strane različitih agencija i ustanova. Primjerice, informacije o prometnim nesrećama 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 slične ustanove.

Pogreške i nekonzistentnosti u prostoru, odnosno lokacijama, mogu se dogoditi iz više razloga. S jedne strane, podaci o lokaciji prometne nesreće nerijetko se izvlače iz analognih policijskih zapisnika. To most te metode uvelike ovisi o iskustvu policijskog službenika i osoblja za pripremu podataka. Otežavaju a okolnost je što podaci o prometnim nesrećama i o prometnoj mreži mogu biti razvijani na različitim softverskim platformama koje koriste različite prostorne koordinatne sustave. Takođe može doći 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 nesreće locirane na granici dvaju ili više segmenta, ne iz razloga što su se dogodile točno na toj lokaciji, već zato što te lokacije policijski službenici lakše identificiraju. Ukoliko su položaj prometne nesreće ili prometnice podložni pogreškama, potreban je postupak ispravljanja prostornih koordinata. Pritom treba imati na umu da ne postoji procedura za sustavno i uinkovito ispravljanje mogu ih prostornih pogrešaka u podacima prometnih nesreća i prometnica (Tegge, Ouyang 2009).

1.1. Prethodna istraživanja

Istraživanje provedeno u Hong Kongu (Loo 2006) predlaže metodologiju potvrđivanja i identifikacije preciznih lokacija prometnih nesreća s cestovnom mrežom bez korištenja zone obuhvata (eng. *buffer*). Autorica tvrdi da vrijednosti rezultata modela ovise o dostupnosti i kvaliteti podataka prometnih nesreća. Prije analize podataka, potrebno je ispraviti krivo georeferencirane varijable. U istraživanju su analizirane prometne nesreće Hong Konga iz razdoblja od 1993. do 2004. godine. Podaci prometnih nesreća preuzeti su iz policijske baze podataka prometnih nesreća (*eng. Traffic Accident Database System - TRADS*), a digitalna cestovna mreža je preuzeta od Odjela za katastarske poslove Hong Konga. Budući 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 područjima s velikom gustošćom prometnica. Sa sustavom validacije prostornih podataka zasnovanom na GIS-u 65-80% policijskih zapisnika prometnih nesreća su imali točne attribute ulica i okruga.

U Riyadhu, glavnem gradu Saudijske Arabije, provedeno je istraživanje (Al-Ghamdi 2003) karakteristika prometnih nesreća na križanjima i ravnim cestovnim potezima. Korišteni su policijski zapisnici prometnih nesreća kako bi se odredile generalne karakteristike prometnih nesreća na određenim lokacijama. Skup podataka je podijeljen na teške nesreće i lige nesreće s materijalnom štetom. Autor zaključuje da je najčešći uzrok teških nesreća na ravnim

cestovnim segmentima udarac u fiksni objekt, a 50% teških nesreća uključuje pješake. Neprimjereno ponašanje vozača je uzrok većine prometnih nesreća. To uključuje preveliku brzinu, oduzimanje prednosti i prolazak kroz crveno svjetlo.

U Ujedinjenom Kraljevstvu provedena su tri istraživanja (Noland, Quddus 2004, Wang i dr. 2009, Quddus 2008) koja analiziraju utjecaj prometne infrastrukture i demografskih karakteristika prostornih jedinica na prometne nesreće. Podaci su preuzeti iz različitih institucija. Prostorni podaci su preuzeti iz EDINA UKBORDERS baze podataka i Ordnance Surveya, demografski podaci iz popisa stanovništva Ujedinjenog Kraljevstva, a podaci o prometnim nesrećama iz STATS19 baze podataka prometnih nesreća. Jedan od ključnih elemenata istraživanja je integracija više baza podataka u opsežan GIS što je omogućeno geokodiranjem lokacija prometnih nesreća. Rezultati prvog istraživanja koje su proveli Noland i Quddus (Noland, Quddus 2004) pokazali su da urbanizirana i gusto naseljena područja imaju manji broj prometnih nesreća. Autori prepostavljaju da velika gusto prometa rezultira manjim brzinama. Područja s visokom gusto zaposlenosti imaju veći broj prometnih nesreća, vjerojatno zbog visoke aktivnosti na ulicama. Područja s visokom stopom siromaštva imaju više nesreća, a karakteristike prometnika (broj obilaznica i križanja) ne utječu na nesreće. Rezultati drugog istraživanja koje su proveli Wang, Quddus i Ison (Wang i dr. 2009) potvrđili su pretpostavku da je prosječna brzina statistički značajna i povezana s težinom nesreće zbog gusto prometa. Zakrivljenost ceste je statistički značajna s negativnim koeficijentom. To ukazuje da su vozači i svjesniji stanja na cesti na zakrivljenim cestama i da opreznije voze kroz naseljena i komercijalna područja gdje ima više zavojova. Zakrivljene ceste u prostornoj jedinici mogu se smatrati zaštitnim faktorom u pogledu sigurnosti na cestama. Rezultati trećeg istraživanja kojeg je proveo Quddus (Quddus 2008) pokazali su da gusto prometa povezana sa žrtvama u prometu. Varijable poput kvaliteta vozila i zaposlenosti su statistički značajne u svim modelima, a prosječna brzina prometa u popisnim krugovima nema utjecaja na težinu ozljeda.

U Sjedinjenim Američkim Državama provedena su dva istraživanja na području San Francisca (LaScala i dr. 2000, Wier i dr. 2009). Analizirani su utjecaji prometne infrastrukture i demografskih karakteristika prostornih jedinica na stopu prometnih nesreća a motornih vozila i pješaka. Podaci su preuzeti iz različitih institucija. Podaci o prometnim nesrećama su preuzeti su 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 i dr. 2000) ukazuju na značajnu geografsku korelaciju između okolišnih i demografskih karakteristika u nesrećama vozila i pješaka. Najveća vjerojatnost nesreća u vozila i pješaka postoji u području s većim gustoćom stanovništva i većom nezaposlenošću. Detaljniji uvid pokazuje da je riječ o području 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 fizikalnim poslovima koji izlazu radnike opasnostima u prometu, a nezaposlenost može biti povezana s većim brojem aktivnosti na otvorenom i izlaganjem rizicima u prometu. Nesreće s pješacima češće su u dijelovima grada s većom gustoćom prometa i većom koncentracijom barova. Rezultati drugog istraživanja koje su proveli Wier, Weintraub, Humphreys, Seto i Bhatia (Wier i dr. 2009) pokazali su da je gustoća prometa primarni okolišni uzrok nesreća u vozila i pješaka na promatranih područjima.

Cilj istraživanja provedenog u središnjoj Floridi (Abdel-Aty, Radwan 2000) je uspostava matematičkog modela koji objašnjava razliku između frekvencije prometnih nesreća i karakteristika autocesta. Podaci o prometnim nesrećama preuzeti su 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 duga ka 227 km koja povezuje istočnu i zapadnu obalu središnje Floride. Autocesta prolazi kroz centar grada Orlanda. Rezultati su pokazali da se povećanjem prosječne godišnje dnevne prometne (eng. *Annual Average Daily Traffic - AADT*) znatno povećava frekvencija prometnih nesreća. Širina i broj prometnih traka su među znatnjim geometrijskim parametrima. Uske trake i veliki broj traka povećavaju frekvenciju prometnih nesreća, a urbani cestovni segmenti su potencijalno opasniji od ruralnih. Žene, mlađi i stariji vozači su podložniji prometnim nesrećama u većim gustoćama prometa, na užim trakama i većem broju traka. Muškarci i mlađi vozači su češće uključeni u prometne nesreće povezane s visokim brzinama.

Istraživanje provedeno u Ujedinjenom Kraljevstvu na području Norfolka (Jones i dr. 1996) nastoji utvrditi važnost vremena odaziva hitne pomoći i 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 i napravljene su pomoći u 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 nesrečama 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 je predviđalo vjerovatnosc ishoda.

U istraživanju provedenom u Belgiji (Depaire i dr. 2008) provedena je analiza prometnih nesreča na području Bruxellesa s ciljem proučavanja inkovitosti tehnikе grupiranja (*eng. clustering*). Podaci prometnih nesreča 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 nesreča. Grupiranje po klasama kao primarna analiza može otkriti skrivene odnose i pomoći stručnjacima u prometnoj sigurnosti pri segmentiranju prometnih nesreča.

Istraživanje provedeno u južnoindijskom gradu Thiruvananthapuramu (Prasannakumar i dr. 2011) provedeno je s ciljem identifikacije i analize žarišta prometnih nesreča uzrokovanih neodgovarajućim razvojem prometne mreže. Podaci o prometnih nesrečama su preuzeti iz sjedišta prometne policije grada Thiruvananthapurama za godinu 2008. Granice popisnih krugova su digitalizirane i korištene za izdvajanje cestovne mreže s 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 nesreče ovise o vremenskim i prostornim varijablama.

U istraživanju provedenom u Afyonkarahisaru, Turska (Erdogan i dr. 2008), identificirana su žarišta na autocestama unutar spomenutog administrativnog područja. Podaci o prometnih nesrečama su preuzeti od područnog ureda prometne policije (*eng. District Traffic Agency Officers*) za razdoblje od 1996. do 2006. godine. Autoceste spomenutog područja su digitalizirane sa mjerila 1:1000000. Rezultati pokazuju da je gusto prometnih nesreča najveće u kolovozu i prosincu. Frekvencija prometnih nesreča na dnevnoj razini je najviša petkom i vikendom, što ukazuje na utjecaj turističkih putovanja. Većina prometnih nesreča dogodila se oko podneva, ali najteže prometne nesreče su najčešće oko poноći. Njihov najčešći uzrok je neprimjereno ponašanje vozača.

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

Tako je obrazno pridruživanje atributnih podataka spomenutim zonama i identifikacija zona na temelju grupiranja atributa oko najbliže srednje 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 pomoći u K-means grupiranja određena su područja s velikom gustošćom prometnih nesreća. Na taj je način uspostavljen model koji omogućuje usporedbu i grupiranje žarišta sa sličnim karakteristikama.

2. Opći 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, to 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. Uzveši u obzir prethodno navedena inozemna istraživanja, obraćene 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 omoguće prostornu vizualizaciju podataka, ali je dobra polazna osnova jer omoguće generalni pregled stanja podataka. Normalizacija je obaraćena kao primjer metode koja uključuje podjelu na prostorne zone. KDE je relativno nova statistička metoda koja omoguće identifikaciju žarišta prometnih nesreća.

Sigurnost u prometu dobiva sve veću značaj, a statističke analize su neophodne u istraživanju uzroka i provođenju u inkovitim protumjera. Cilj rada je uspostaviti metodologiju obrade i analize prometnih nesreća metodama prostorne statistike. Rezultati bi mogli ili 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

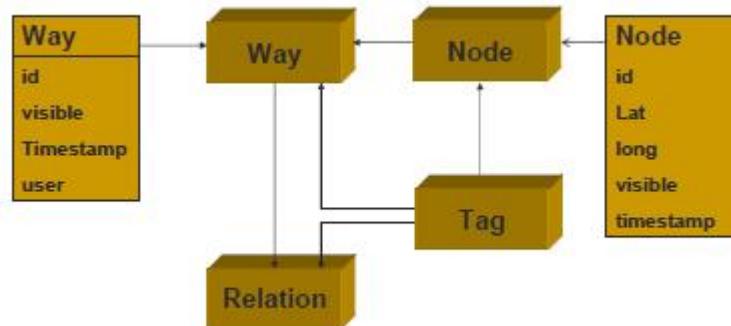
Podaci o prometnim nesreama preuzeti su od Ministarstva unutarnjih poslova Republike Hrvatske (u dalnjem tekstu: MUP) preko Fakulteta prometnih znanosti Sveučilišta u Zagrebu. Rijeđe je o podacima za 2010., 2011., 2012. i 2013. godinu. Podaci o prometnoj mreži preuzeti su sa slobodne karte svijeta OpenStreetMap. 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 primjenjeno je prostorno proširenje baze podataka PostgreSQL poznato pod nazivom PostGIS. Kao što je ranije navedeno, primjenjene 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 mognosti vizualizacije rezultata posljednje dvije metode.

3.1. Preuzimanje prometne mreže Grada Zagreba s karte svijeta OpenStreetMap

OpenStreetMap (u dalnjem tekstu: OSM), projekt slobodne karte svijeta, omogućuje korisnicima suradnju u prikupljanju i dijeljenju geografskih podataka, prvenstveno prometnica u i između 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 ljudе da ih koriste na kreativne i produktivne načine (URL 2). Projekt je započet 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, Zipf 2012).

Geografske informacije u OSM bazi podataka pohranjene su koriste i tri tipa objekata: vor, put i relacija (*eng. node, way, relation*) (Slika 1). Vorovi u bazi podataka sadrže informacije o lokaciji. Rijeđe je o točkama s pridruženom geografskom širinom i dužinom. Linije poput cesta i poligoni se spremaju kao putevi i relacije, koji definiraju logike 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 zaključujuće 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, Zipf 2012).



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

Provedena je analiza prometnih nesreća 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 nesreća.

3.2. Obrada podataka iz policijskih zapisnika pomoći u programskog jezika Python

Podaci o prometnim nesrećama su preuzeti u CSV (*eng. Comma-Separated Values*) formatu. Svako vozilo prometne policije ima GPS prijemnik, što znači da svaka prometna nesreća ima prostorne koordinate (geografska širina i dužina). Otežavajuće je okolnost nepoznat kutni format zapisa. Primjerice, geografska širina ima vrijednost *4512345*, a geografska dužina *1812345*. Takav način zapisivanja 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 nesreća može imati najviše dva atributa ulica, *ulica1* i *ulica2*. Koristeći mrežu ulica preuzetu s OSM-a, moguće je napraviti kontrolu konverzije prostornih koordinata. Problem nastaje kad u zapisniku ne piše naziv ulice na kojoj se dogodila nesreća ili kad naziv ulice ne odgovara nazivu ulice iz OSM-a.

U prvom slučaju ne postoji mogućnost kontrole konverzije prostornih koordinata. Tada se koordinate konvertiraju u decimalne minute. Taj je format odabran iz razloga što GNSS uređaji policijskih službenika ispisuju koordinate u decimalnim minutama (Slika 2, Slika 3).



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



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

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

- *Pozitivnost:* $d(s_1, s_2) \geq 0$;
- *Nedegeneracija:* $d(s_1, s_2) = 0$ ako i samo ako $s_1 = s_2$;
- *Simetričnost:* $d(s_1, s_2) = d(s_2, s_1)$;
- *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 te četiri operacije koje mogu konvertirati znak s_1 u s_2 , ali duljina najkratke 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 osnovi 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, akademski i u komercijalne svrhe. Razvoj je započeo 1989. godine, a objavljen je po etkom 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 praktične 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, Java za Jython i .NET-u za IronPython. Među najpoznatijim korisnicima Pythona su Google i YouTube (URL 4).

Napisana je skripta u Pythonu koja učitava 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 tog dijela postupka je eliminacija toaka koje se nalaze izvan područja Grada Zagreba ili države, odnosno toaka opterećenih grubim pogreškama.

Prometne nesreće koje me u atributima imaju barem jedan naziv ulice konvertiraju se u sva tri kutna formata. Drugim riječima, nastaju tri točke koje predstavljaju istu prometnu nesreću, ali se nalaze na različitim lokacijama (Tablica 1). Oko svake točke kreira se zona obuhvata (objekt tipa poligon) s radijusom od 1000 m. Stvara se unija tri zone obuhvata što u veličini ajeva rezultira objektom tipa multipolygon. Multipolygon se preklapa s mrežom ulica. Cilj tog postupka je smanjenje obujma ravnatelja. To se postiže izdvajanjem ulica koje se nalaze unutar multipolygona ili ga sijeku. Naziv ulice iz zapisnika usporeduje se s nazivima izdvojenih ulica iz OSM-a pomoću Damerau-Levenshteinovog algoritma udaljenosti. Prethodnim testiranjem utvrđeno je da prihvatljiva granična 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 ravnatelja. To se postiže eliminacijom ulica iz multipolygona i njihovi nazivi nimalo ne odgovaraju nazivu ulice iz zapisnika. Sljedeći korak je ravnatelj najmanje udaljenosti između točki i preostalih linija, odnosno prometne nesreće i potencijalnih ulica. Uzvešći u obzir da postoje tri točke s okolnim linijama, uzima se u obzir par koji postiže najmanju udaljenost. Format te točke uzima se kao ispravan, a ako je udaljenost manja od 150 m, to kaže se ocjenjuje pouzdanom i ide u daljnju obradu. U suprotnom se zanemaruje. Tablica 1 prikazuje primjer tog postupka. Sve tri točke predstavljaju istu prometnu nesreću, samo su formati koordinata, a time i lokacije, druga i treća. Izračunata je udaljenost svake točke do najbliže linije. Treba napomenuti da ne mora biti udaljenost do iste linije jer se točke mogu nalaziti na udaljenosti i do nekoliko kilometara. Budući da najmanja udaljenost iznosi 3 metra, prometnoj nesreći se pridružuju koordinate DMS formata. Druge dvije točke se zanemaruju.

Tablica 1. Primjer odabira formata koordinata

Točka br.	Format koordinata	Udaljenost do najbliže linije [m]
1	DMS	3
2	Decimalne minute	5
3	Decimalni stupnjevi	7

Uzveši u obzir današnju to nost i dostupnost GNSS tehnologije, vrijednosti od 100 m, odnosno 150 m za najmanju udaljenost djeluju velikodušno. Treba uzeti u obzir brojne okolišne faktore i stru nu spremu osobe koja koristi prijemnik. Ve ina nesre a doga a se u urbanim dijelovima grada gdje multipath, odnosno refleksija GNSS signala, igra presudnu ulogu. GNSS prijamnici fiksno su postavljeni u policijsko vozilo (Slika 2, Slika 3) što otežava situaciju ako se nesre a dogodila na prometnoj ulici. Automobili se moraju isklju iti iz prometa što ih nerijetko udaljava od mjesta nesre e i kvari podatke (Slika 57). Prometne nesre e su este i stresne situacije za sudionike i policijske službenike koji rade uvi aj. Jedan od najvažnijih faktora je stru na spremu policijskog službenika koji koristi GNSS prijamnik. Ako im nije dobro objašnjena važnost i prednost spomenutog sustava, može do i do raznih propusta. Ukoliko se nesre a dogodila na sporednoj ulici ili prilazu ve oj ulici, policijski službenik e unijeti naziv ve e ulice. Uzveši u obzir koli inu atributa koju moraju unijeti u kratkom vremenu, previdi su normalna i opravdana pojava. Ukoliko zaborave unijeti koordinate iz GNSS prijamnika na mjestu nesre e, postoji mogu nost da kasnije pokušavaju do i do njih na druga ije na ine. Primjerice, vra anjem na mjesto nesre e ili korištenjem aplikacija poput Google Eartha koje koriste druga ije kutne formate prostornih koordinata.

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

<i>Ulica1/Ulica2 (policijski zapisnik)</i>	<i>OpenStreetMap</i>
Palmoti eva ulica	Ulica Junija Palmoti a
Boškovi eva ulica	Ulica Josipa Jurija Boškovi a
or i eva ulica	Ulica Ignjata or i a
Vlaška	Vlaška ulica
Ka i eva ulica	Ulica Fra Andrije Ka i a Mioši a
Klai eva ulica	Vjekoslava Klai a
Mihanovi eva ulica	Antuna Mihanovi a
Heinzelova	Ulica Vjekoslava Heinzela
Podaupskog	Ulica Eugena Podaupskog
Jarunska ulica	Jarunska

Tablica 2 prikazuje razlike u nazivima ulica izme u zapisnika prometne policije i podataka OSM-a. Policijski zapisnici vode se analogno. Uzveši u obzir koli inu atributa koje policijski

službenici trebaju unijeti, propusti su esta pojava. Primije en je velik broj nesre a 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 pomo u PostgreSQL-a i PostGIS-a

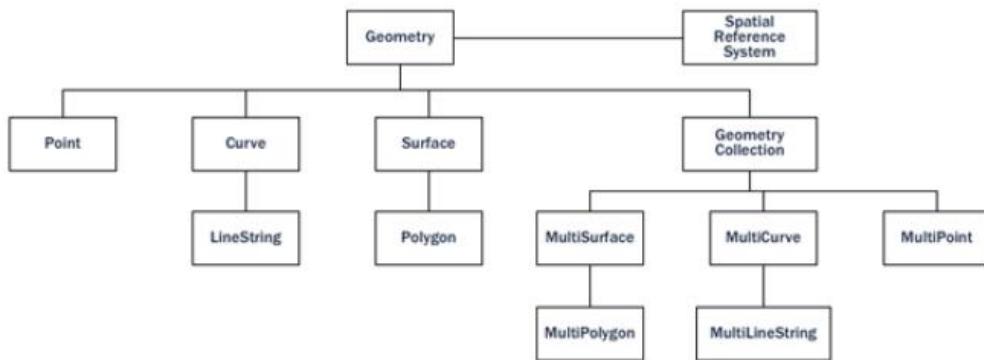
Rezultati obrade podataka zapisnika spremjeni su u CSV datoteku i uneseni u bazu podataka. Korištena je baza podataka PostgreSQL s prostornim proširenjem PostGIS. PostgreSQL je moderni sustav za upravljanje objektno-relacijskim bazama podataka (*eng. Object-Relational Database Management System - ORDBMS*) 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 ve 15 godina i poznat je po pouzdanosti, integritetu podataka i tonosti. Radi na svim većim operativnim sustavima, uključujući 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 veću 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 OpenFTS-a i PostGIS-a koji koriste PostgreSQL. OpenFTS (*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 Oracleve 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 uinkovito 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 obuhvaaju prostorne strukture poput granice i dimenzije. Tipovi prostornih podataka su organizirani hijerarhijski. Svaki podtip nasljeđuje strukturu i ponašanje svog nadtipa (URL 6).



Slika 4. Hjerarhija geometrijskih tipova (URL 6)

Baze podataka pružaju metode pristupa, poznate kao indeksi, koje omogućuju brz pristup podacima. Indeksiranje standardnih tipova se vrši izvodi pomoći u B-tree indeksu, 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 računanje s poligonima zahtjevno, ali tako brzo 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*. Ključne prometne nesreće (oznaka iz zapisnika: kljuc_pn) u tablici *nesrece* je primarni ključ na osnovi kojeg se povezuju zapisi iz tablica *sudionici* i *vozila*. Svakoj prometnoj nesreći su 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 posti i u itavanjem 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 u itavanje 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, 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 (vrijednosti u 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 (vrijednosti u stupcima)	Stupac 1	Stupac 2	Stupac c	n (ukupan 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 me usobno isključuju. Prepostavimo 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 jednaku proporciju 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):

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

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

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

$$E_{r,c} = \frac{n_r * n_c}{n} \quad (2)$$

gdje je $E_{r,c}$ o ekivana 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):

$$X^2 = \sum \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}} \quad (3)$$

prijeđenu je $O_{r,c}$ opažana frekvencija, a $E_{r,c}$ o ekivana frekvencija određena 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 uspore uveć 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 zahtjeva ispunjenje sljedećih uvjeta (URL 8):

- Varijable u tablici kontingencije moraju biti brojano odredive;
- Suma o ekivanih frekvencija mora biti jednaka sumi opažanih frekvencija;
- Ako tablica kontingencije ima više od 2 elije, manje od 20% o ekivanih 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 opštine koje prikazuju populaciju stanovništva. Obito se jedinstvena vrijednost predstavlja određenom bojom na karti ne može primijeniti na cijelu promatrani zonu. To nije i smisleniji način jer je podjela područja na zone i prikaz prostorno intenzivnih atributa (*eng. spatially intensive attribute*). Primjerice, gusto a populacije kroz zone (Smith i dr. 2007).

Ako su podaci organizirani po područjima, esto 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 tom kontekstu, proces se odnosi na dijeljenje broja ili težine podataka na osnovi atributne vrijednosti koja će pretvoriti informaciju u mjeru intenziteta za promatranoj zonu. Taj proces uklanja posljedice povezane s podacima koji su povremeni iz zona različitih površina ili veličina populacije i omogućava usporedbu podataka temeljenih na zonama (*eng. zone based data*) (Smith i dr. 2007).

Esto su korištena tri tipa normalizacije: prosjeci, postoci i gustoća. 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 dijeli se s brojem kućanstava u istoj zoni. U slučaju postotaka ili pretvaranja u raspon [0, 1], odabrana vrijednost atributa dijeli se s maksimalnom vrijednošću ili ukupnim brojem u setu podataka. Primjerice, broj nezaposlenih osoba u zoni dijeli se s ukupnim brojem radno sposobnih ljudi (npr. 18-65). Kod gustoće, djelitelj je površina zone gdje je pronađen odredeni, odnosno traženi atribut. Rezultat je skup vrijednosti koje prikazuju populaciju po jedinici površine, odnosno mjera gustoće. Primjerice, broj stanova/km², tona pšenice/jutro i sl. (Smith i dr. 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. Taj oblik reprezentacije odgovara rasteru ili mrežnom GIS-u. Zbog toga rasterske reprezentacije općenito ne trebaju normalizaciju na osnovi površina (*eng. area-based normalisation*) jer se već nalaze u prikladnom formatu. Proses generiranja mrežne reprezentacije može uključiti izračun vrijednosti za svaku eliju iz vektorskih podataka, primjerice korištenjem uniformnog modela (Smith i dr. 2007).

Spomenuta opažanja naglašavaju nekoliko nedostataka (Smith i dr. 2007):

- Proses dijeljenja mora imati način rješavanja slučaja gdje je djelitelj 0;
- Treba uzeti u obzir podatke koji nedostaju;
- Ako su atributi 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 zbumuju ih rezultata zbog nestabilnosti varijance (*eng. variance instability*). Do nje dolazi kada se odre uju pojave na osnovi 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 razdoblju. 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 djelitelja ili izvornih podataka neizvjesne vrijednosti ili važnosti može dovesti do neprihvatljivih rezultata;
- Odreivanje prikladnog djelitelja može biti problematično. Primjerice, ako se traže obrasci pojedine bolesti u odre enoj godini, podaci popisa stanovništva nisu prikladan izvor informacija s obzirom na veliku ugroženu populaciju.

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 takvih 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 sofisticirane metode poput raznih vrsta analiza grupiranja (Smith i dr. 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 esto su grupirane u klase koje pojednostavljaju izgled 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 računa se 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 te metode je jednostavno raunanje klasa i svaka je klasa približno jednako zastupljena na kona noj karti. Nedostatak te 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 gusto a kao tip normalizacije. Primjerice, broj prometnih nesre a odreene vrste unutar mjesnog odbora je podijeljen s površinom odbora. Obzirom da je korištena metoda kvantila, svaka karta ima razlike vrijednosti granica klasa zbog razliitog broja prometnih nesre a odreene vrste. Na kartama su prometnice prikazane zelenom bojom zbog boljeg kontrasta i orijentacije. Mjesni odbori u kojima nije bilo nesre a s odreennim atributom prikazani su sivom bojom. Ostali mjesni odbori prikazani su razliitim nijansama crvene boje, od svijetlih nijansi (mala gusto a) do tamnih nijansi (velika gusto a).

3.6. Analiza podataka statisti kom metodom Kernel Density Estimation

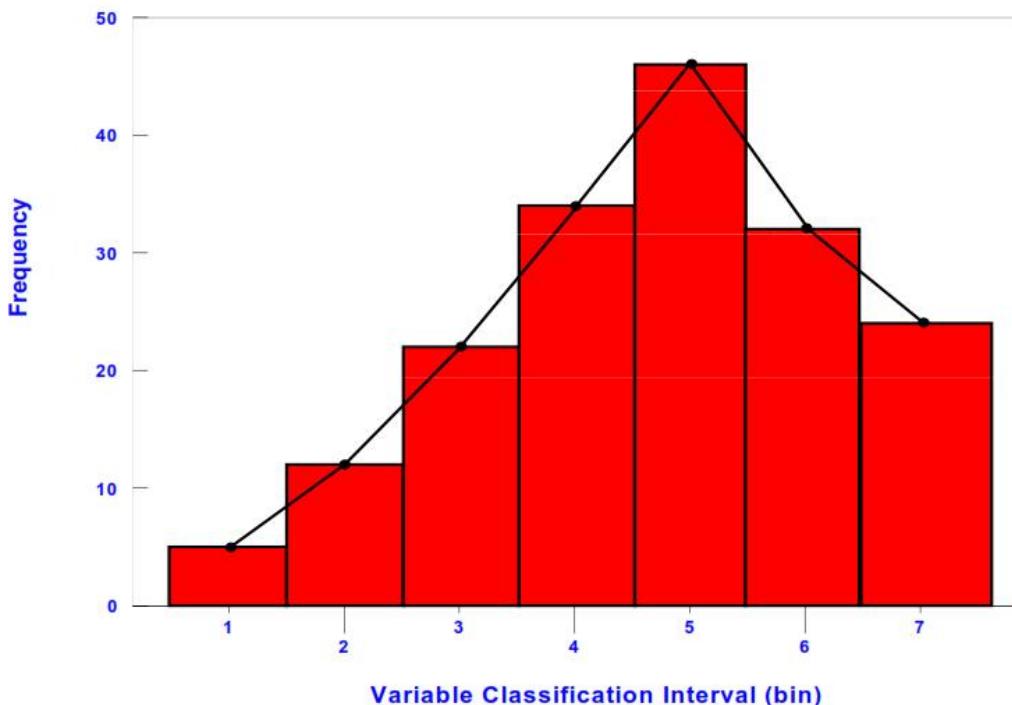
Prostorno opsežne varijable ne bi se smjele u normalnim uvjetima distribuirati direktno po zonama ve standardizirati. Za varijable koje su podskup ukupnog skupa varijabli esto je poželjno izra unati i kartirati te podatke kao razmjer ukupnog broja, ali je za same ukupne skupove potrebno izra unati mjeru gusto e. Primjerice, dijeljenje ukupne populacije N s površinom zone A . To daje vrijednost gusto e N/A za svaku zonu, ali prepostavlja 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 i dr. 2007).

Ako skup podataka na raspolaganju nije temeljen na zonama ve 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 ukupan broj opažanja podijeljen s površinom zone na koju se odnose opažanja. Nedostatak tog 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 i dr. 2007).

Alternativni pristup ra unanju gusto e dvodimenzionalnih skupova to kastih podataka temelji se na tehnikama zasnovanim na jednoparametarskoj (*eng. univariate*) statisti koj analizi (Smith i dr. 2007). Obrada podataka zasnovana je 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, taj postupak rezultira sljede im problemima (Levine 2013):

- Informacije su odbene 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 tom pristupu jer pretjerano smanjenje širine intervala može dovesti do nepouzdanih rezultata;
- Tehnika je ovisna o proizvoljno definiranoj velini 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 osnovi dostupnog skupa podataka (*eng. sample*). Pretpostavimo da je nasumičan skup podataka (x_1, x_2, \dots, x_n) dobiven na osnovi nepoznate kontinuirane distribucije vjerojatnosti $f(x)$. Cilj je napraviti procjenu ove distribucije vjerojatnosti, koja se obilježava s $\hat{f}(x)$. Postoje brojni načini primjene takvih 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 je 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 komplikirane. Bez obzira je li primjena grafička ili numerička, potreban je način procjene f za svaku vrijednost od X (Burt, 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 to samo parametri od $f(x)$, ovo je jednako procjeni samog $f(x)$. Ako se koriste \bar{x} i s kao procjene, dobije se (Burt, Barber 1996):

$$\hat{f}(x) = \frac{1}{s\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x - \bar{x}}{s}\right)^2\right] \quad (4)$$

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

Kernel estimation je jednostavan pristup koji izbjegava prethodno spomenute nedostatke (Burt, Barber 1996). Kernel je, u doslovnom smislu rije i, jezgra ili središnji dio predmeta. Esto se u raunalnim znanostima odnosi na središnji dio operativnog sustava dok se u geoprostornim analizama odnosi na metode koje ukljuuju raunanja pomo u dobro definiranog lokalnog susjedstva. Primjerice, modeliranje gusto e ili analiza lokalne mreže pomo u bloku elija ili radijalno simetri ne funkcije (Smith i dr. 2007).

Kernel Density Estimation (u dalnjem tekstu: KDE) je tehnika generalizacije lokacija opažanja na cijelo podruje. Dok prostorna distribucija i statistika žarišta pružaju statisti ki pregled samih podataka opažanja, tehnike interpolacije generaliziraju podatke opažanja na itavo podruje. Tehnike interpolacije pružaju procjene gusto e za sve dijelove promatranog podruja. Procjena gusto e je varijabla intenziteta (*eng. intensity variable*), z - vrijednost, koja je procijenjena na odre enoj lokaciji. Može se prikazati kao karta površina ili reljefna karta koja prikazuje intenzitet na svim lokacijama. KDE je tehnika interpolacije prikladna za pojedina ne to kaste lokacije (Levine 2013).

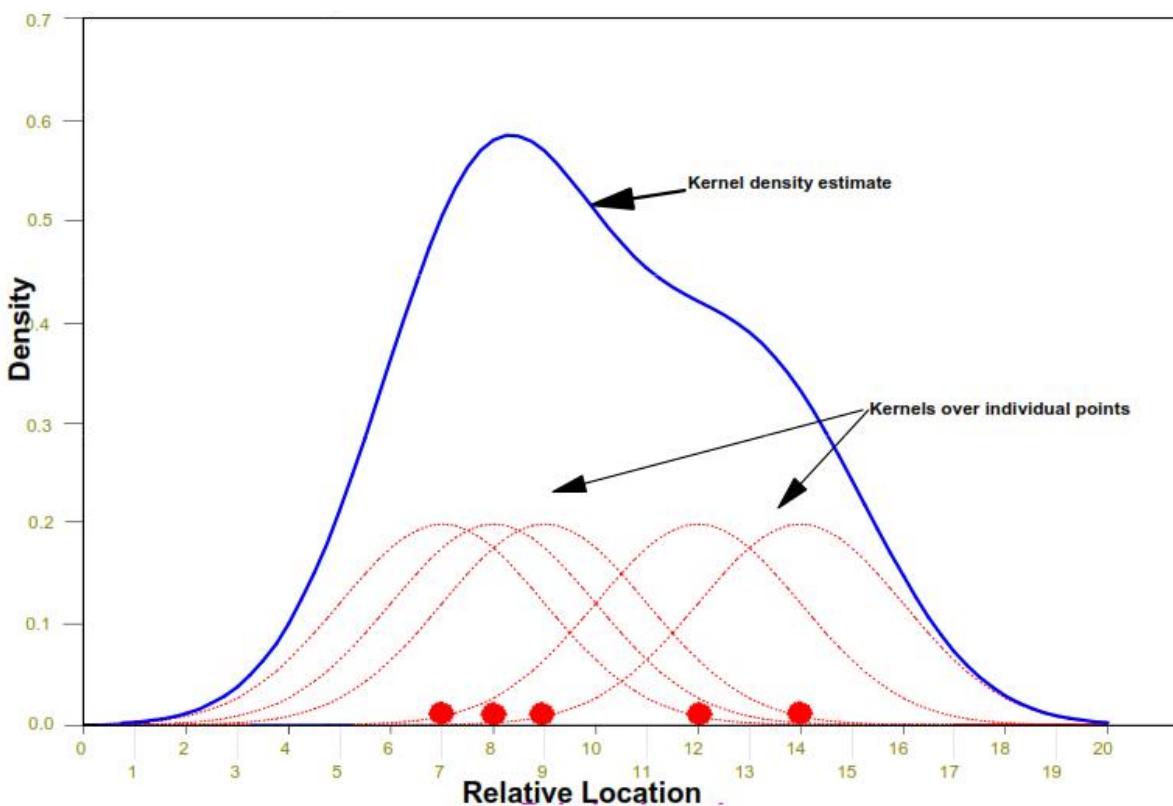
Metoda kernel estimation pruža glatke procjene bez potrebe za eksplicitnim pretpostavkama o formi $f(x)$. U sluaju histograma, svako opažanje se tretira kao to ka. Ako to ka leži u intervalu vjerojatnosti, cijela masa to ke $1/n$ se dodaje vjerojatnosti. S druge strane, kernel metode raspore uju masu svakog opažanja oko opažane vrijednosti. Koli ina širenja je odre ena funkcijom pod nazivom kernel (Burt, Barber 1996).

Kernel funkcija je esto korištена kao normalna distribucija, odnosno kao zvonolika krivulja, koja u svakom smjeru teži u beskonaost, ali ima konačnu (jediničnu) površinu ispod krivulje (Smith i dr. 2007). U veini sluajeva kernel funkcija je simetri na tako da je to ka jednako proširena prema višim i nižim vrijednostima (Burt, Barber 1996). Oblik kernel funkcije nema velik utjecaj na skup vrijednosti gusto e koji je dodijeljen linearnom segmentu ili površini u 2D aplikacijama. Puno značajni utjecaj ima izbor parametra širenja, odnosno širine pojasa (*eng. bandwidth*) (Smith i dr. 2007).

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

funkcijom podrazumijeva se funkcija koja opada jednakom stopom u svim smjerovima proporcionalno udaljenosti od svake točke. U ovom je slučaju reč 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 tog 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 ozнакама, izraz za KDE glasi (Burt, Barber 1996):

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad (5)$$

Gledajući prvo argument K u izrazu (5), vidi se da je udaljenost između točke x i svake vrijednosti skupa podataka izražena kao višekratnik od h . Budući da je h iste mjerne jedinice

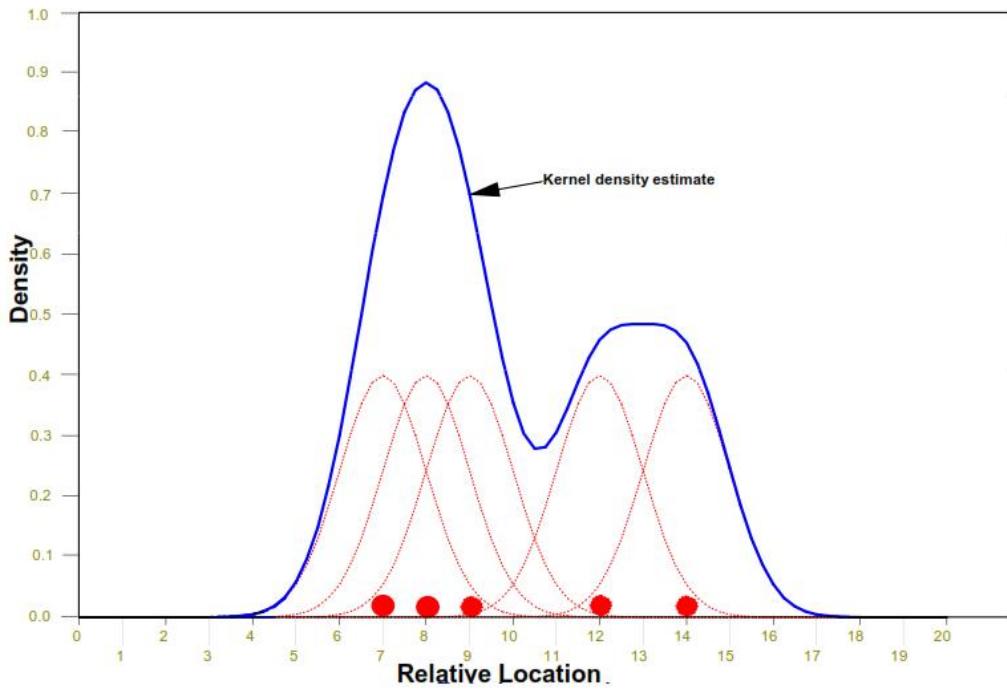
kao i x , omjer je bezdimenzionalan. Ta se skalirana udaljenost koristi u kernel funkciji kako bi se pronašao doprinos svakog opažanja. S K kao opadajućom funkcijom udaljenosti, doprinos točke koja se nalazi na velikoj udaljenosti od x će biti mali, dok će bliže točke imati veći doprinos. Zbroj svih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

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

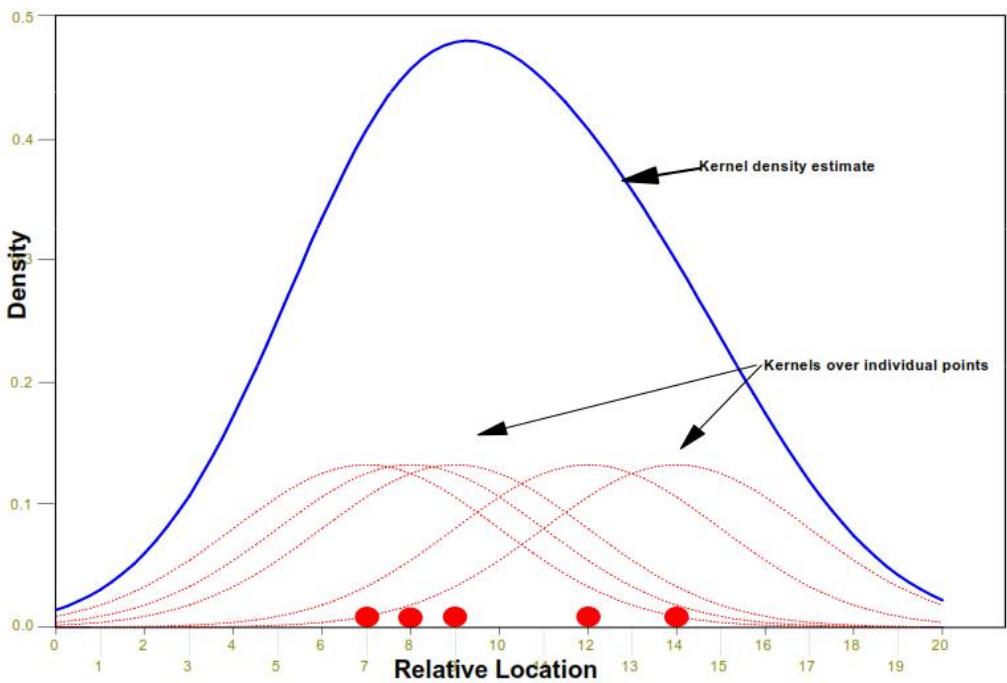
- $K(x) \geq 0$ i $\int_{\mathbb{R}} K(x)dx = 1$;
- Simetrična oko ishodišta $\int_{\mathbb{R}} xK(x)dx = 0$;
- Konačan drugi moment $\mu_2(K) = \int_{\mathbb{R}} x^2 K(x)dx < \infty$ s označom $R(K) = \int_{\mathbb{R}} (K(x))^2 dx$.

Može se donijeti nekoliko općenitih zaključaka o tom pristupu. S jedne strane, $\hat{f}(x)$ nasljeđuje mnoga svojstva od kernel funkcije. Ako je K funkcija vjerojatnosti, $\hat{f}(x)$ će takođe biti funkcija vjerojatnosti (u potpunosti jednaka ili veća od nule, s ukupnom površinom jednakoj 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, Barber 1996).

Širina kernela je prilagodljiva. Uski kerneli tjesno koncentriraju masu oko središnjih (opažanih) vrijednosti, a široki kerneli rezultiraju većim širenjem (Burt, Barber 1996). Manja vrijednost širine područja vodi do gušće mreže procjene gustoće s mnoštvom 'vrhova i dolina' (Slika 7). S druge strane, velika vrijednost širine područja vodi do glatke distribucije i manje varijabilnosti između područja (Slika 8). Dok manje vrijednosti širine područja prikazuju veće razlike između područja, treba uzeti u obzir statistiku preciznost procjene. Ako veličina uzorka nije prevelika, manja širina područja vodi do nepreciznijih procjena, a 'vrhovi i doline' mogu prikazivati samo slujajne varijacije. Ako je na raspolaganju veći 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 gusto e s malom širinom pojasa (Levine 2013)



Slika 8. Funkcija gusto e s velikom širinom pojasa (Levine 2013)

Simetrija je atribut koji se ne naslje uje iz kernela. Procjena gusto e zahtijeva n procjena od K . Uvezši u obzir da velik broj procjena može biti potreban pri stvaranju grafa ili izra una vjerojatnosti, ta metoda o ito nije prikladna za ru no ra unanje (Burt, Barber 1996).

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

$$K(z) = \begin{cases} 1 - |z|, & |z| \leq 1 \\ 0, & |z| > 1 \end{cases} \quad (6)$$

gdje je

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

Taj je kernel samo trokut visok jednu jedinicu, širok dvije jedinice i ima područje jedinstva (*eng. area of unity*). Njegova glavna prednost leži u jednostavnosti izrauna, samo su potrebne operacije oduzimanja i dijeljenja. Nema potrebe za razmatranjem točaka 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 prekide 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 izbjegiće se diskontinuitet središnjeg nagiba (Burt, Barber 1996):

$$K(z) = \begin{cases} \frac{3}{4}(1 - z^2), & |z| \leq 1 \\ 0, & |z| > 1 \end{cases} \quad (7)$$

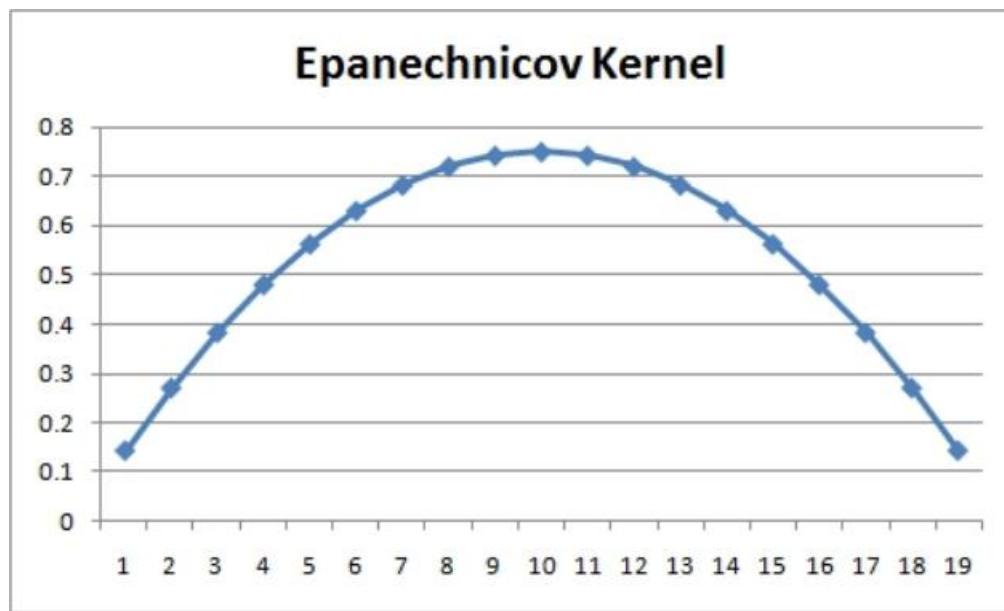
Sada K varira glatko preko središnje točke i \hat{f} je gladki. Još jednom K postaje udaljenost koja teži nuli daleko od opažanja, tako da se samo obližnja opažanja trebaju razmatrati. Međutim, postoje diskontinuiteti nagiba na rubovima tog kernela. Još gledaće procjene stvaraju se korištenjem normalnog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

$$K(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) \quad (8)$$

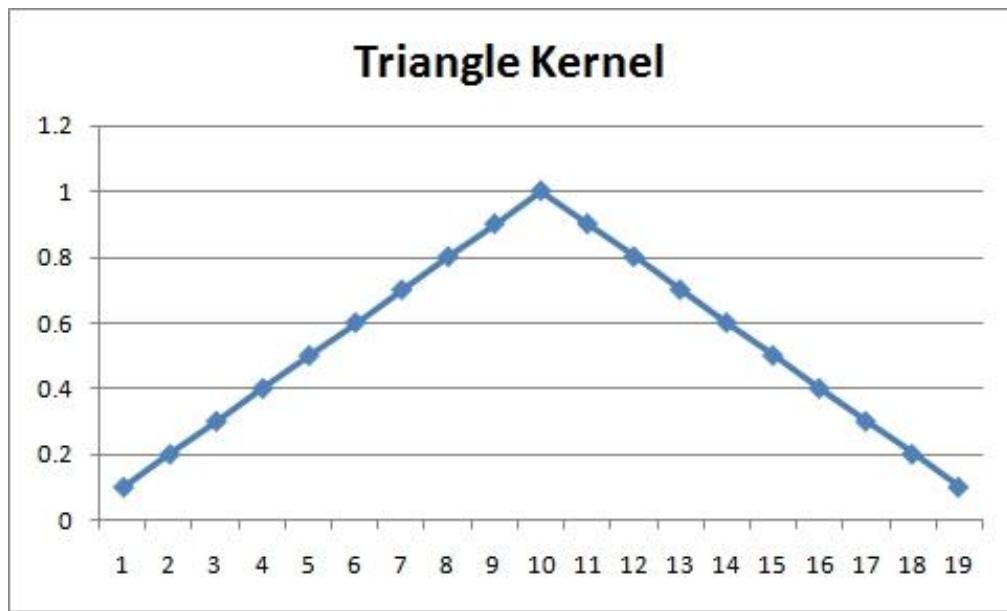
To rezultira vrlo glatkim funkcijom gustoće, ali na trošak više aritmetike. Tu je dodan teret eksponencijala i K je tehnički svugdje ne-nul vrijednost. Budući da se normalna krivulja brzo približava nuli, opažanje udaljeno više od nekoliko višekratnika od h ne može puno pridonijeti \hat{f} . Stoga se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

*Tablica 4. Kernel funkcije primjenjene u obradi prometnih nesreća a KDE metodom
(r je najveća derivacija kernela) (Guidoum 2013)*

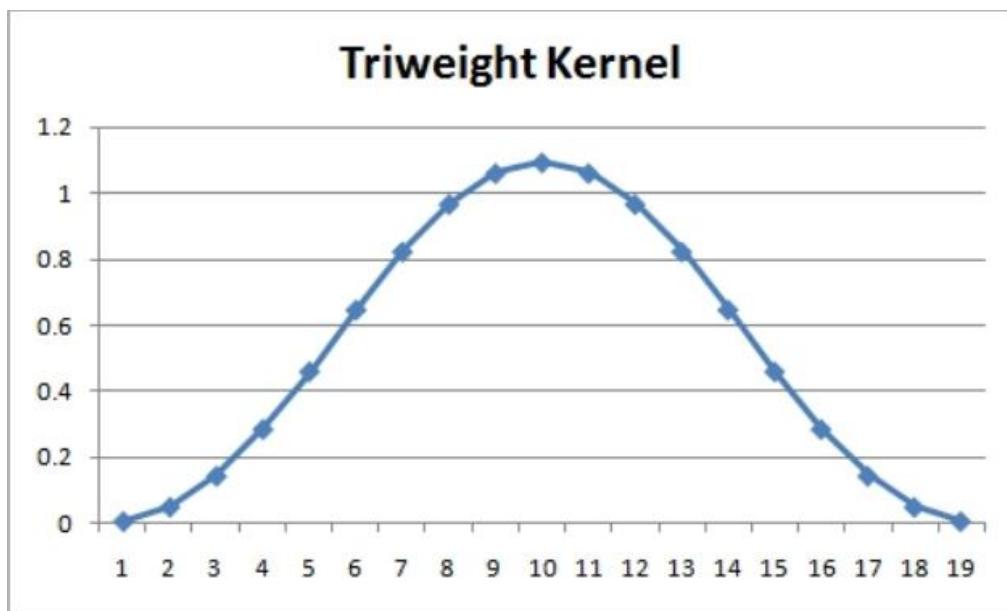
Kernel	$K(z; r)$	$R(K)$	$\mu_2(K)$
Epanechnikov (Quadratic)	$K(z; 2) = \frac{3}{4}(1 - z^2)1_{(z \leq 1)}$	3/5	1/5
Uniform	$K(z; 0) = \frac{1}{2}1_{(z \leq 1)}$	1/2	1/3
Triangular	$K(z; 1) = (1 - z)1_{(z \leq 1)}$	2/3	1/6
Triweight	$K(z; 6) = \frac{35}{32}(1 - z^2)^31_{(z \leq 1)}$	350/429	1/9
Biweight (Quartic)	$K(z; 4) = \frac{15}{16}(1 - z^2)^21_{(z \leq 1)}$	5/7	1/7



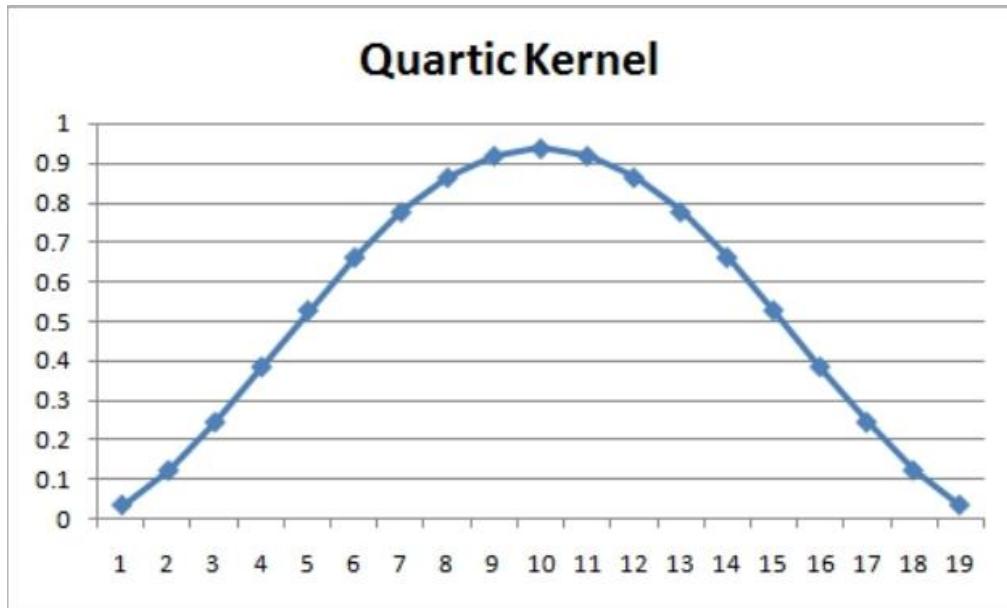
Slika 9. Graf Epanechnikov (Quadratic) kernel funkcije (URL 10)



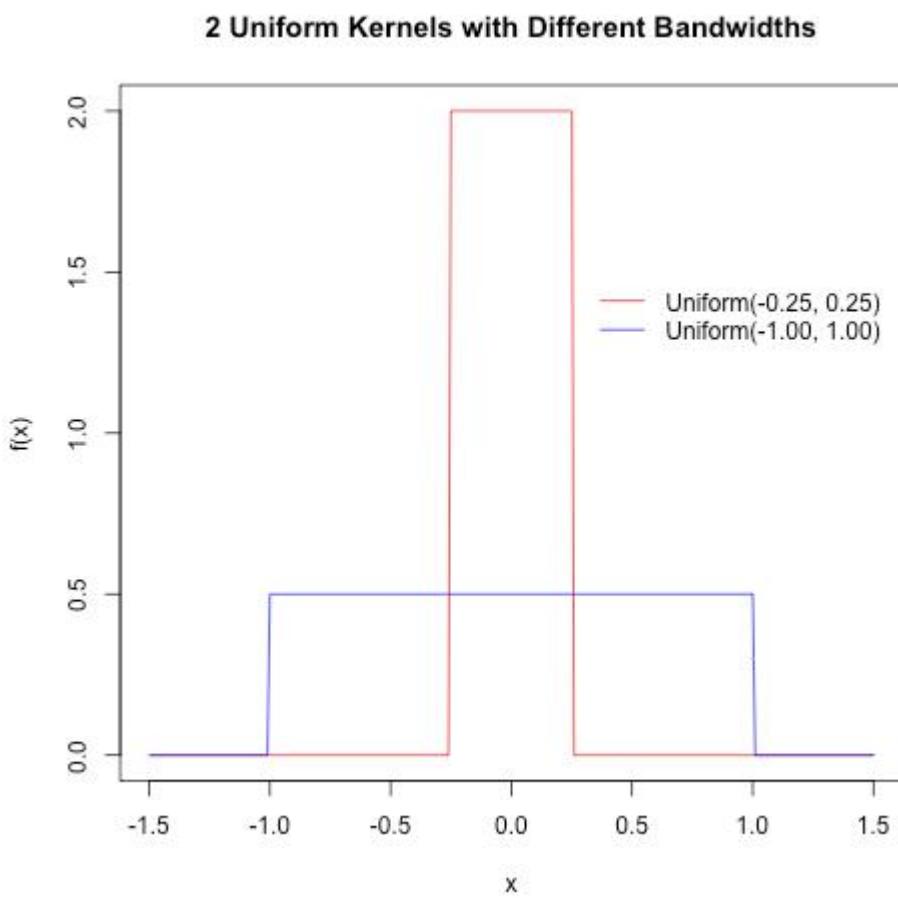
Slika 10. Graf Triangular kernel funkcije (URL 10)



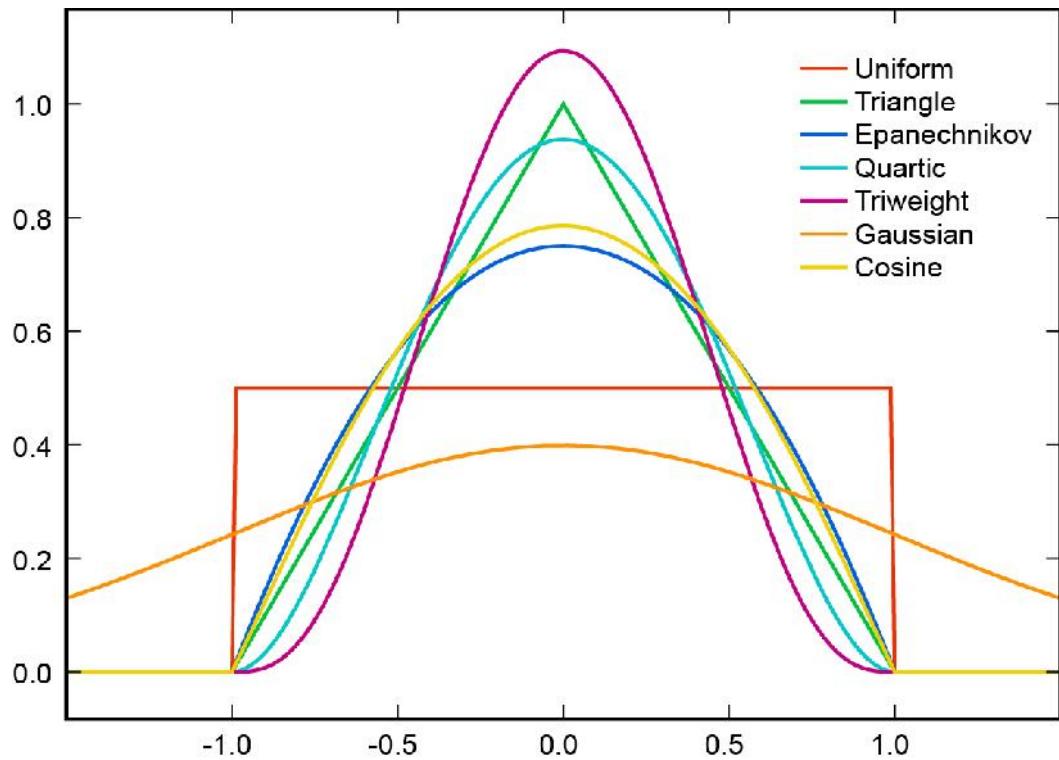
Slika 11. Graf Triweight kernel funkcije (URL 10)



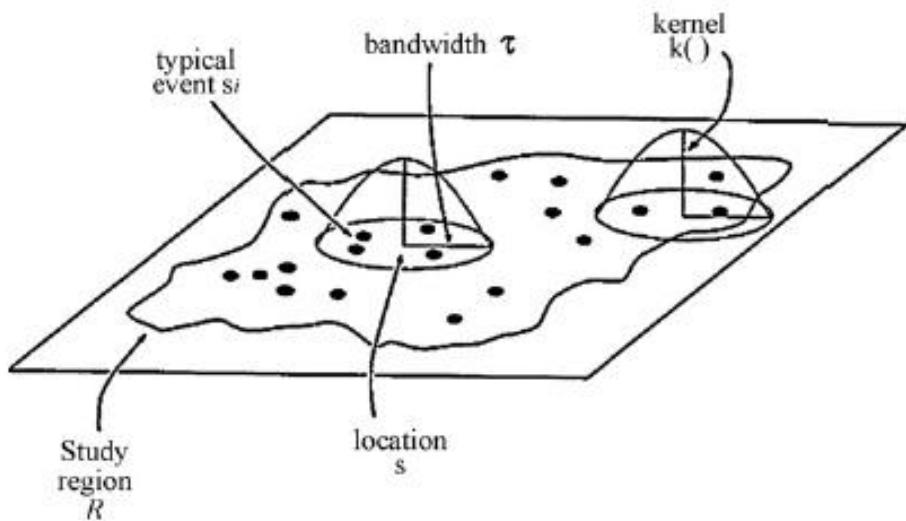
Slika 12. Graf Biweight (Quartic) kernel funkcije (URL 10)



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



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



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

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

sluaju 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 svjetlih 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 još 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 sprjeavlja 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, 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 mjeri koliko se informacija može izvesti 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, Barber 1996).

Kod fiksne širine područja, korisnik određuje interval koji će se koristiti i mjerne jedinice. Ovisno o tipu korištene kernel funkcije, interval ima različito značenje. Kod funkcije normalnog kernela širina područja je standardna devijacija normalne distribucije. Kod funkcije quartic i triangular kernela širina područja je radius područja pretraživanja u kojem se provodi 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 gustošćama. 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 radius tog kruga kao širinu područja oko te točke (Smith i dr. 2007). Nauštrb više ravnatelja, mogu a su dva poboljšanja koja povećavaju iskoristivost informacija iz uzorka (Burt, Barber 1996).

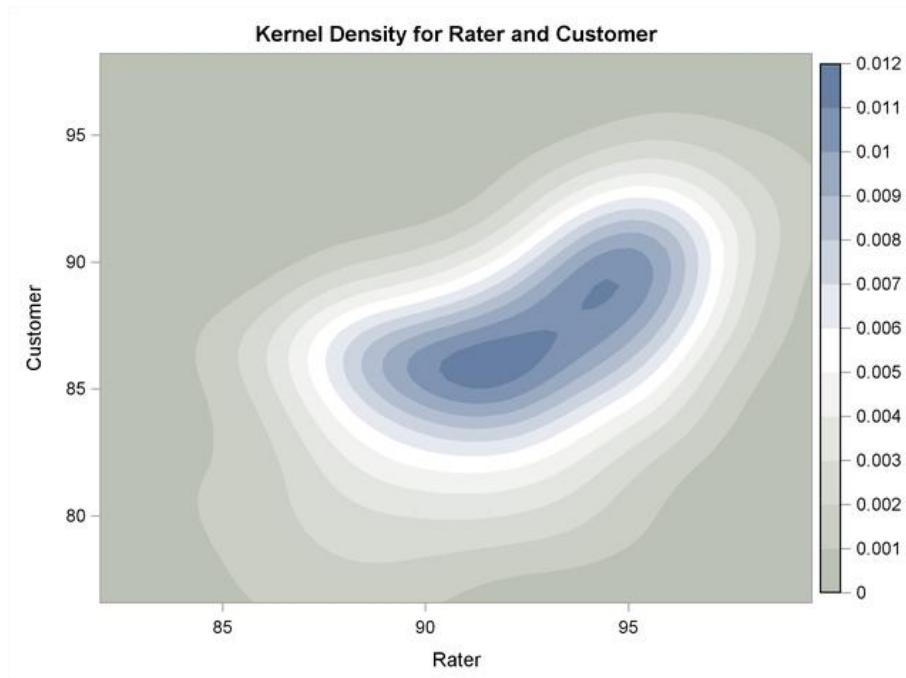
U prvom slučaju, moguće je optimirati izbor širine pojasa odabiranjem vrijednosti koja umanjuje pretpostavljenu pogrešku u \hat{f} . Tehnika se zove *cross-validation* i koristi samo podatke u identifikaciji optimalnog h . Osnovna ideja je da je dobar izbor za h dati dobru procjenu za \hat{f} . Cross-validation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

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

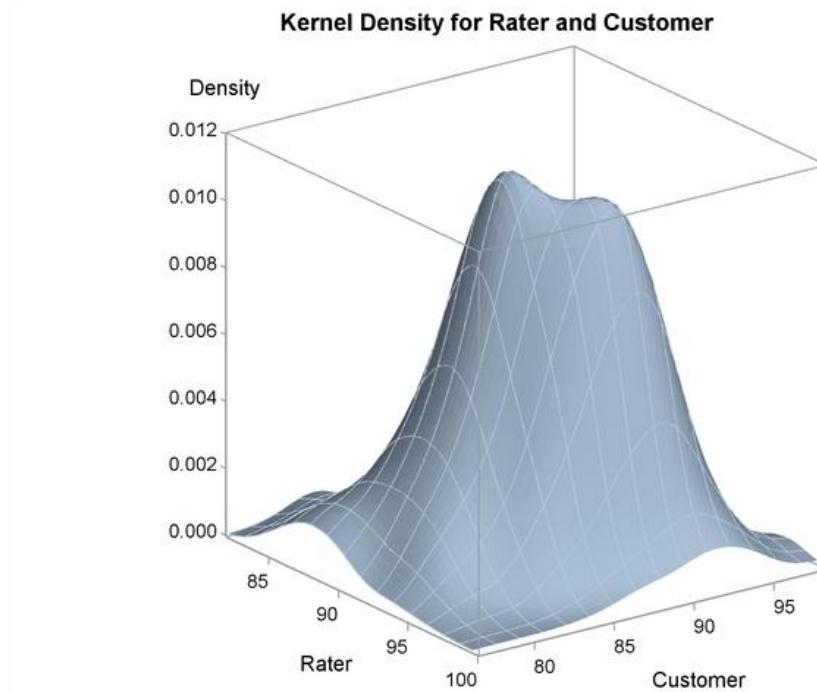
3.6.3. KDE u 2D i 3D

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

dodavanje druge dimenzije, odnosno rotacija funkcije oko svake točke. Primjerice, ako bi se koristila normalna distribucija kao funkcija izglađivanja, 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 pridonijela tačka 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 prostore pogreške, primjenjena zaokruživanja i stvaranje dozvole za isključenjem određenih podataka iz ravnateljstva. Vrijednosti svih točaka na presjeku svake mreže ili za svaku pojedinu mrežu 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 i dr. 2007).

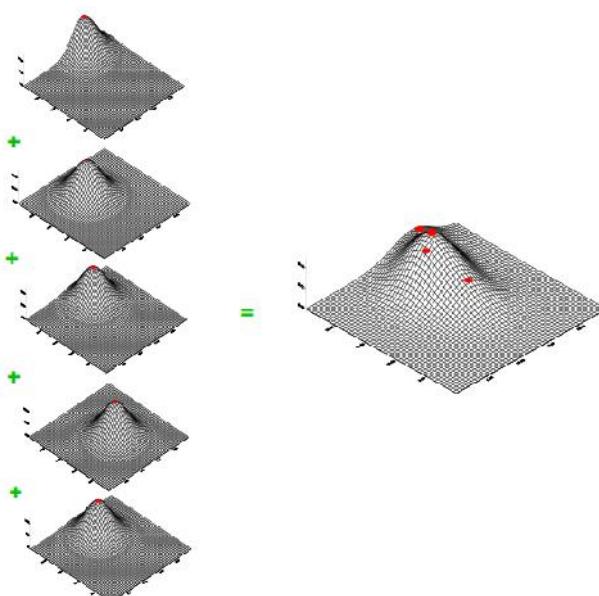


Slika 16. Vizualizacija 2D plohe KDE-a (URL 13)



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 gustoće dobivenu kao zbroj svih pet individualnih površina. Ta 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. Zbrajanje normalnih kernel površina za pet to aka (Levine 2013)

Vrijednosti rezultiraju e mreže mogu biti dobivene kao (Smith i dr. 2007):

- *Relativne gusto e* koje daju vrijednosti opažanja po jedinici površine. One se prilago avaju veli ini mreže (broj opažanja/m² 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. Zbroj vrijednosti svih elija mora biti jednak broju opažanja u analizi;
- *Vjerojatnosti* koje daju vrijednosti broja opažanja podijeljenog s ukupnim brojem opažanja.

Budu i da su ta 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 iskorištена za procjenu apsolutnih gusto a, relativnih gusto a ili vjerojatnosti. Za usporedbu razli itih vrsta opažanja s razli itim vremenskim razdobljima 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 i dr. 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 njene primjene 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, 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 privla ne su 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, Barber 1996):

$$d_i = \sqrt{(x - x_i)^2 + (y - y_i)^2} \quad (9)$$

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

$$K\left(\frac{d_i}{h}\right) = \frac{1}{2\pi} \exp\left[-\frac{1}{2}\left(\frac{d_i}{h}\right)^2\right] \quad (10)$$

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

$$\hat{g}(x, y) = \frac{1}{h^2} \sum_{i=1}^n K\left(\frac{d_i}{h}\right) \quad (11)$$

Treba primijetiti dijeljenje s h^2 u izrazu (11). Budući da je h izražen u metričkim jedinicama, procjena gustoće bit će 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, Barber 1996).

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

Jedan od mogućih problema je procjena dvoparametarske funkcije gustoće u 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 u vjerojatnosti, a volumen ispod površine je vjerojatnost. Da bi se gustoća u točka pretvorila u gustoću u vjerojatnosti, potrebno je samo dijeljenje s brojem točaka skupa podataka (Burt, Barber 1996):

$$\hat{f}(x, y) = \frac{1}{n} \hat{g}(x, y) = \frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{d_i}{h}\right) \quad (12)$$

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 izračun udaljenosti imao smisla. Prirodan izbor je izražavanje obje vrijednosti kao z -vrijednost kod ravnatelja d . Na taj se način h interpretira kao širina kernela u jedinicama standardnog odstupanja umjesto u mjerljivim jedinicama poput kilometara i milja. U svim ostalim slučajevima dvoparametarska metoda je direktno proširenje jednodimenzionalne tehnike (Burt, 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^2 , što je približno jednako protezanju od $30 \times 20 \text{ km}$. Na taj način vrijednost od 200 metara iznosi oko 1% vrijednosti linearног protezanja analiziranog područja. Uvezši 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 osnovi prostorne zavisnosti. Korištenjem te metode definira se proizvoljna prostorna jedinica za analizu koja je homogena za cijelo područje. To omogućuje usporedbu i taksonomiju (Anderson 2009).

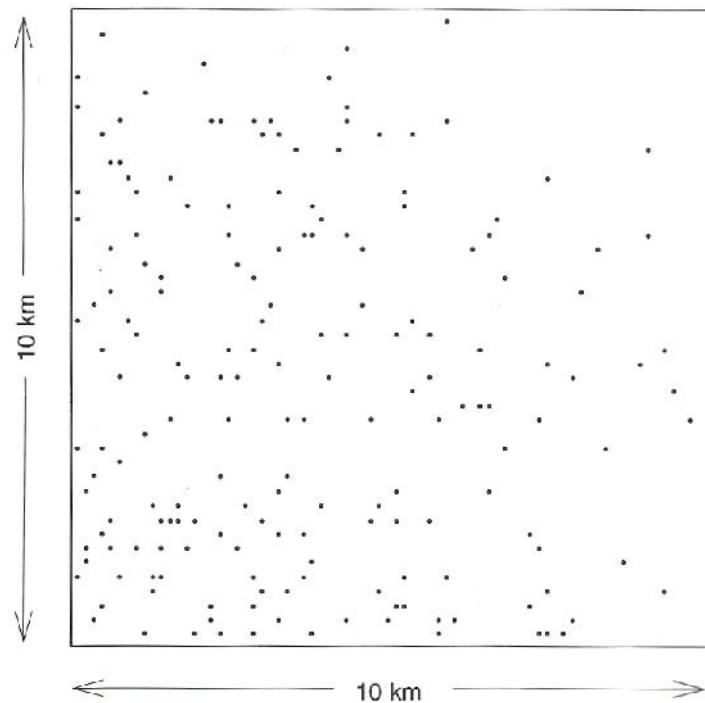
KDE metode imaju razne primjene (Smith i dr. 2007):

- Izglađivanje skupa točkih podataka;
- Kreiranje kontinuiranih površina iz skupa točkih 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.

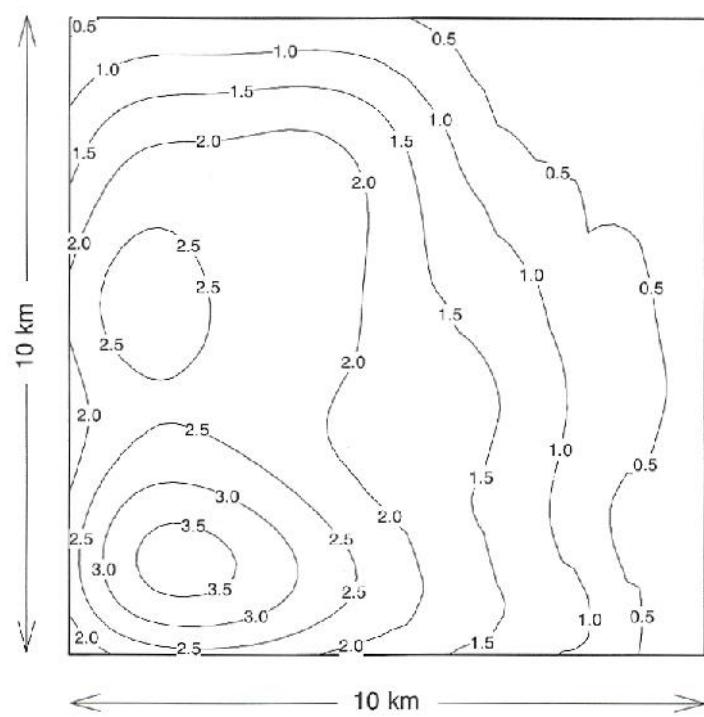
KDE se također može primijeniti nad više od jednog skupa točkastih podataka. Rezultat takve dvostrukе analize skupova podataka esto je omjer primarnog i sekundarnog skupa. Cilj je obrada primarnih obrazaca s uklonjenim ili smanjenim pozadinskim efektima. Pritom treba paziti da rezultati nisu pod utjecajem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

U nastavku je opisan način primjene KDE metode koji su opisali autori Burt i Barber (Burt, 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^2 . 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 regiju može podijeliti na elije i brojati točke unutar svake elije, taj pristup uključuje sve nedostatke povezane s histogramima. Traži se metoda koja daje izgledenu procjenu gustoće i stoga se odabire dvoparametarska 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 zbrajaju na svim lokacijama. Slika 20 predstavlja reljefnu kartu (*eng. contour map*) koja prikazuje rezultat iscrtanih vrijednosti. Najveća gustoća je 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^2 , srednja gustoća je $2 \text{ točke}/\text{km}^2$. Na karti je vidljivo da su ispodprosječne gustoće na isti nivou u najsjevernijim dijelovima promatranog područja. Može se uočiti nagli pad gustoće na južnim i zapadnim rubovima karte. Riječ je o rubnom efektu koji je posljedica toga što nisu uzeti uzorci izvan granica karte. Da su uzeti uzorci izvan karte, procjene blizu rubova bile bi veće. Budući da normalna krivulja naglo pada, jako je malen doprinos točaka koje su na većoj udaljenosti od $2h$. Bilo bi poželjno imati točke u pojusu širine $2h$ oko promatranog područja, ali ne više od toga. Također se rubni efekti ne šire više od $2h$ u području promatranja. Ne vidi se rubni efekt duž iste granice zato što su gustoće toliko niske da točke koje nedostaju nemaju veliki doprinos. Vidi se da kontura od 0.5 ima male izbočinu manje od širine pojasa od 0.1 km. Te su izbočine rezultat metode, odnosno indikacija da je h premala vrijednost za regije male gustoće. Može se koristiti veća širina pojasa, ali u

tom se služuju gube detalji u područjima visoke gustoće. Bolji izbor je adaptivni kernel koji se automatski prilagođava varijacijama gustoće.



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



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

3.7. Vizualizacija rezultata

Rezultati analiza vizualizirani su u tri GIS aplikacije. Rezultati normalizacije i KDE metode vizualizirani su 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 započeo 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 pomoćno u 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 pristupa mnogim grafičkim tehnologijama i mnoge mogunosti 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 aeroizmjeru i geoznanosti. Kombinira operacije nad rasterskim, vektorskim i tematskim podacima u jednu opsežnu aplikaciju namijenjenu desktop korisnicima. Pristupa 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).

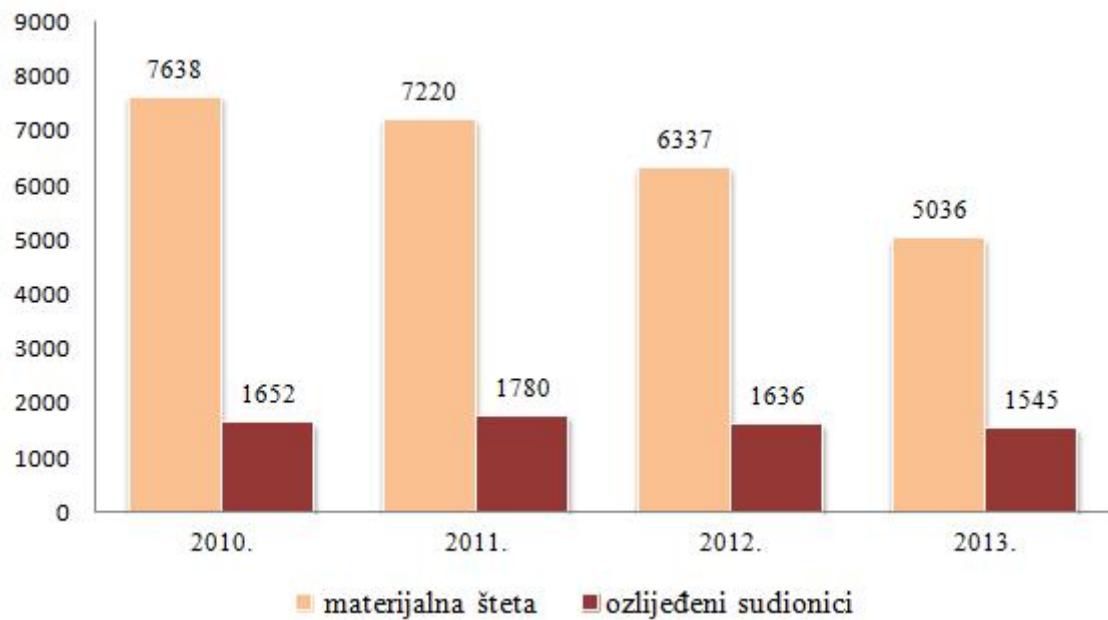
4. Rezultati

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

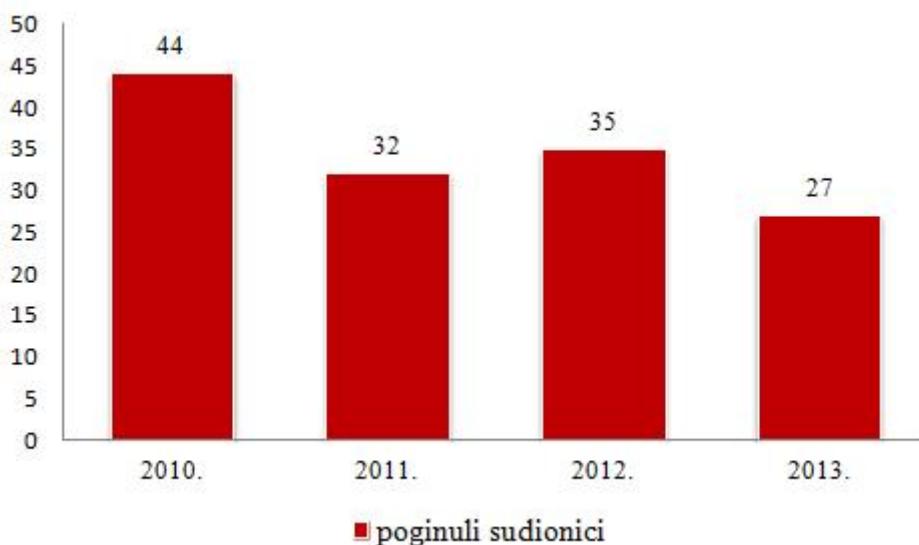
4.1. Osnovni podaci za obradu

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. Grafička 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 ame 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									
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
Ukupno	10060	3347	47	10783	2432	75	1231	457	6	22074	6236	128

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

Vrsta prometne nesreće	Križanje			Ravni cestovni potez			Zavoj			Ukupno		
	Materijalna šteta	Ozlijede sudionici	Poginuli sudionici									
Međusobni 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
Ukupno	10060	3347	47	10783	2432	75	1231	457	6	22074	6236	128

Tablica 6 prikazuje odnos između okolnosti prometnih nesreća i karakteristika cesta na kojima su se dogodile. Okolnosti prometne nesreće uključuju neprimjerena ponašanja sudionika u prometu. Karakteristike cesta uključuju križanja, ravne cestovne poteze, zavoje i ostale oblike. Ostali oblici nisu analizirani jer uključuju oblike poput mostova, pješaka kih staza i prijelaza preko pruge. Riječ je o mjestima na kojima se dogodio mali broj nesreća. Najveći broj nesreća s poginulim sudionicima dogodio se na ravnim cestovnim potezima, najviše zbog brzine neprimjerene uvjetima i nepropisnih radnji voza. Najveći broj nesreća s ozlijedima enim sudionicima dogodio se na križanjima zbog nepoštivanja prednosti prolaska i nepropisnih radnji voza. Nesreća s materijalnom štetom podjednako su zastupljene na križanjima i ravnim cestovnim potezima. Kod križanja je riječ o nepoštivanju prednosti prolaska, a kod ravnih cestovnih poteza o nepropisnim radnjama voza. Na zavojima se dogodilo malo nesreća, oko 5% ukupnog broja nesreća. Najveći 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 nesreća i karakteristika cesta na kojima su se dogodile. Vrsta prometne nesreće uključuje međusobni sudar vozila, nalet na pješaka, biciklista i slično. Najveći broj nesreća s poginulim sudionicima dogodio se na ravnim cestovnim potezima. Na križanjima se dogodio najveći broj nesreća s ozlijedima enim sudionicima. U oba slučaja najviše su vrste prometnih nesreća nalet na pješaka i međusobni sudar vozila. Tablica daje pregled koliko i u kojim pojedinim vrstama prometnih nesreća na različitim dijelovima cesta može pomoći pri donošenju odluka o detaljnjoj analizi karakteristika lokacija.

4.2. Rezultati obrade hi-kvadrat testovima

Tablica 8. Distribucija prometnih nesreća po posljedicama i karakteristikama cesta

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

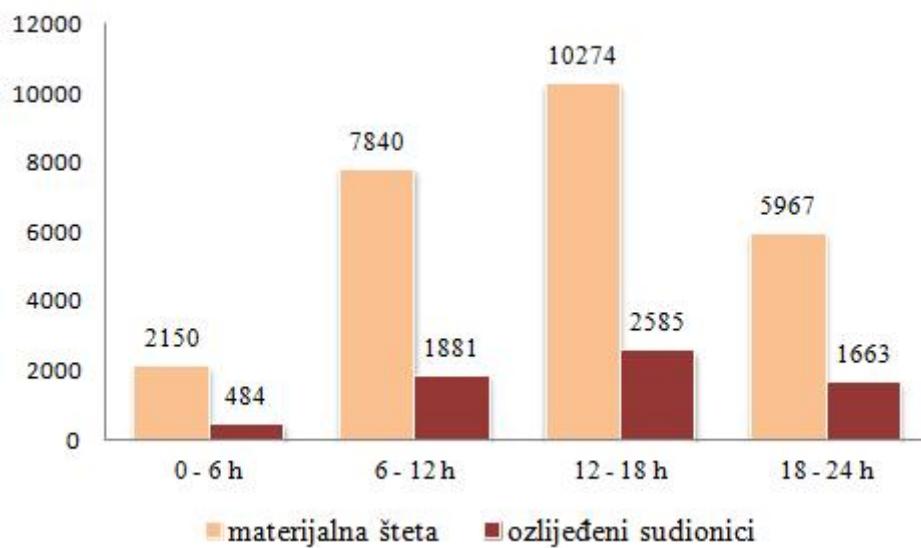
Prometne nesreće	Križanje (%)	Ravni cestovni potez (%)	Zavoj (%)	Ukupno (%)
S materijalnom štetom	10060 (74.77)	10783 (81.14)	1231 (72.67)	22074 (77.62)
S ozljeđenim sudionicima	3347 (24.88)	2432 (18.30)	457 (26.98)	6236 (21.93)
S poginulim sudionicima	47 (0.35)	75 (0.56)	6 (0.35)	128 (0.45)
UKUPNO	13454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

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

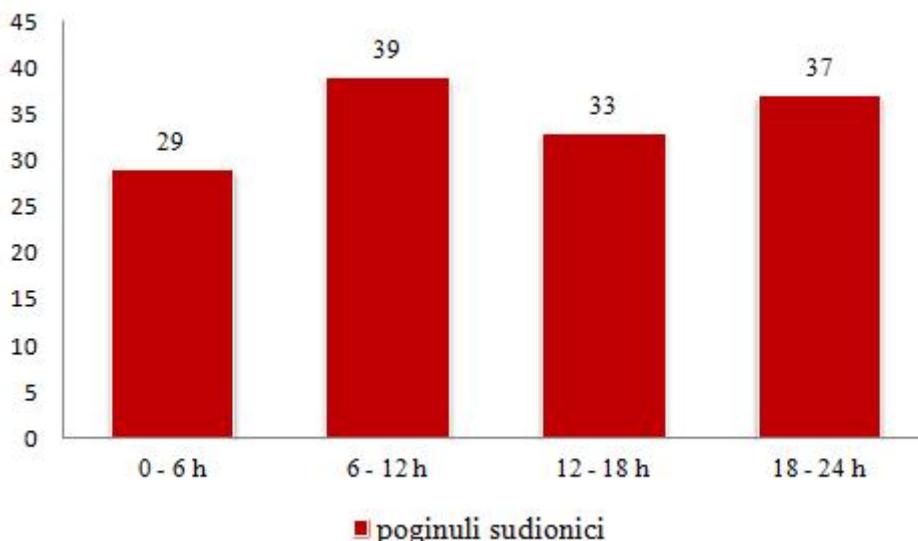
Tablica 9. Distribucija prometnih nesreća po posljedicama i dijelovima dana

($P=0.000$, 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 (%)
<i>S materijalnom štetom</i>	2150 (80.74)	7840 (80.33)	10274 (79.69)	5967 (77.83)	26231 (79.53)
<i>S ozljeđenim sudionicima</i>	484 (18.17)	1881 (19.27)	2585 (20.05)	1663 (21.69)	6613 (20.05)
<i>S poginulim sudionicima</i>	29 (1.09)	39 (0.40)	33 (0.26)	37 (0.48)	138 (0.42)
UKUPNO	2663	9760	12892	7667	32982



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



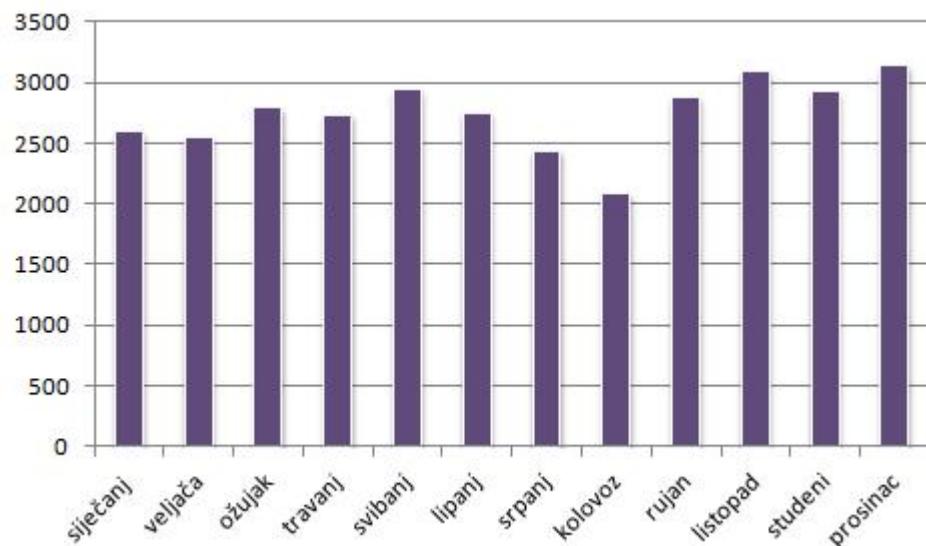
Slika 24. Grafi ki prikaz broja prometnih nesre a s poginulim sudionicima po dijelovima dana

Tablica 9 prikazuje tablicu kontingencije za provedbu hi-kvadrat testa homogenosti za provjeru postojanja omjera u raspodjeli prometnih nesre a s razliitim 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 taj na in i vizualno provjeriti. Kod nesre a s materijalnom štetom i ozlje 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 takvih prometnih nesre a zabilježeno je od 6 h do 12 h, a znajno je manji njihov udio izme u 12 h i 18 h).

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

 $(P = 0.000, \text{hi-kvadrat} = 144.07, \text{broj stupnjeva slobode} = 22, \text{razina značajnosti} = 5\%)$

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

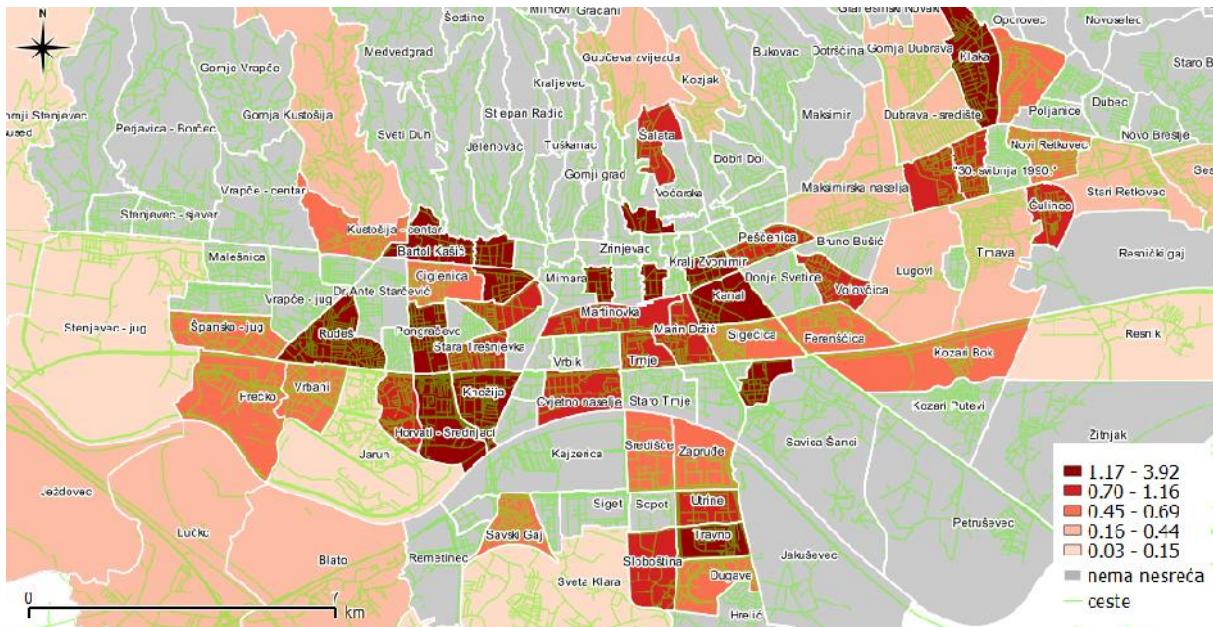


Slika 25. Broj prometnih nesreća po mjesecima u godini (2010.-2013.)

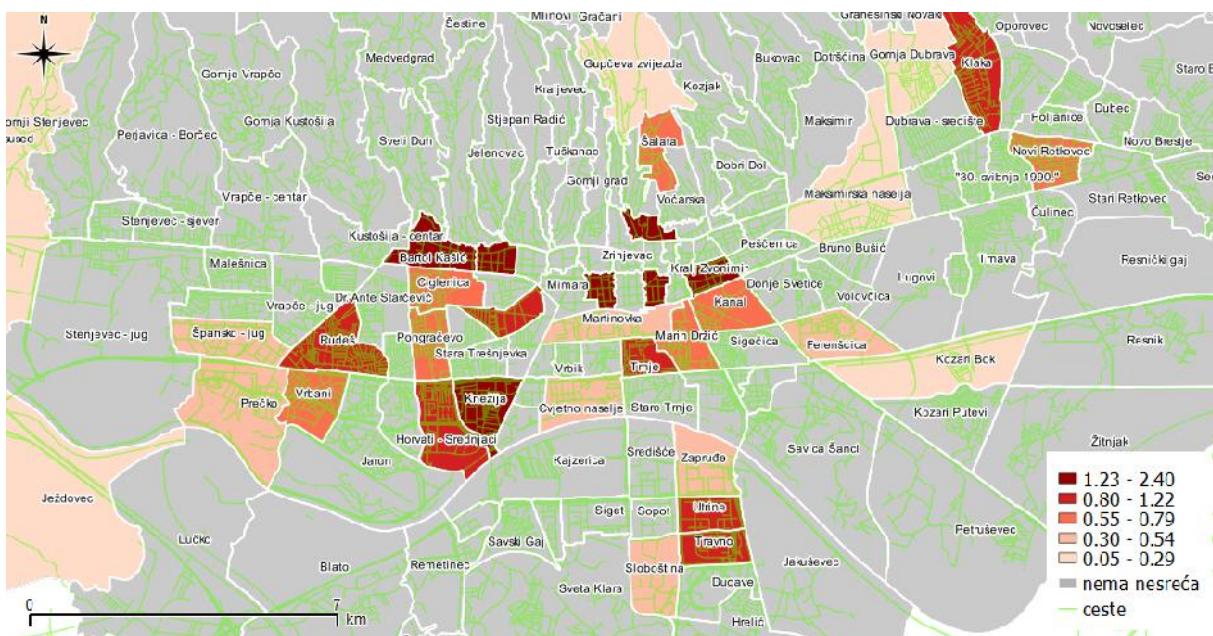
Tablica 10 prikazuje distribuciju prometnih nesreća po mjesecima tijekom etiri godine. Poslužila je kao tablica kontingencije u hi-kvadrat testu homogenosti. Postavljena je nul-hipoteza da ne postoji razlika u omjerima broja prometnih nesreća različitim posljedica po mjesecima u godini. Na temelju dobivenog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

Slika 25 prikazuje ukupni broj prometnih nesreća po mjesecima tijekom etiri 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 isti će razdoblja svestra omestalošću nesreća.

4.3. Rezultati obrade statističkih metodom normalizacije



Slika 26. Broj prometnih nesreća s poginulim osobama po km^2

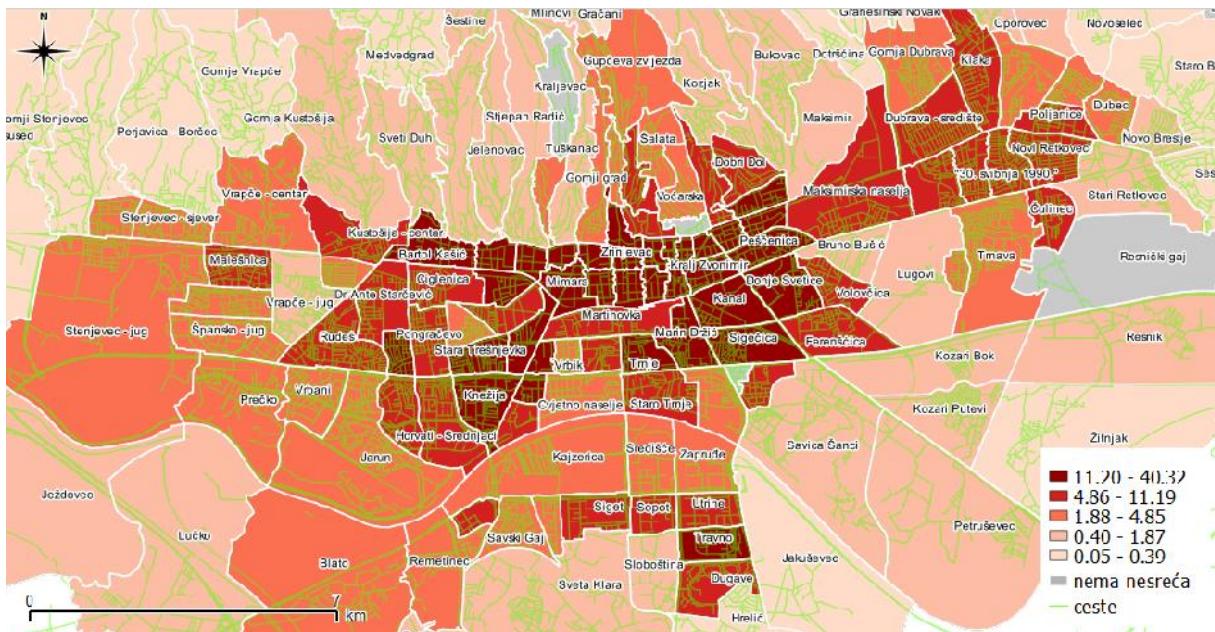


Slika 27. Broj prometnih nesreća s poginulim pješacima po km^2

Vizualni prikaz rezultata normalizacije prometnih nesreća s poginulim sudionicicima nije najprikladniji za analizu zbog malog broja nesreća (Slika 26, Slika 27). Većina mjesnih odbora s najvećom gustoćom je zastupljena u obje slike što ukazuje da se veći dio nesreća s poginulim sudionicicima odnosi na pješake. Postoji velik broj odbora bez i jedne registrirane

Rezultati

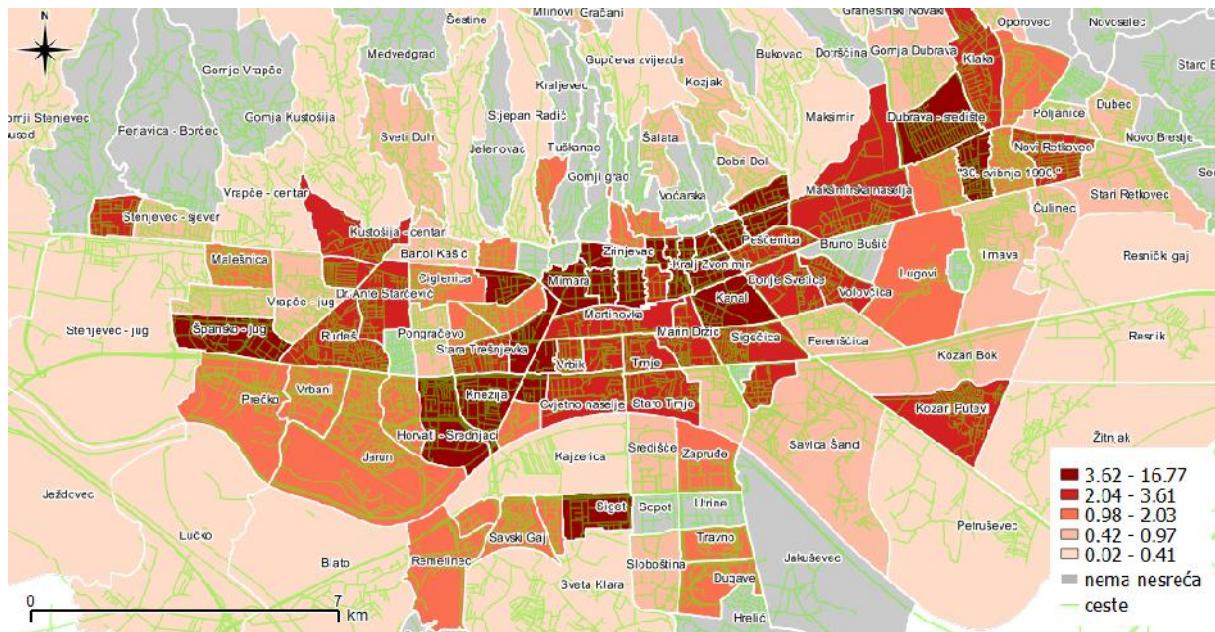
nesreće što onemogu uje daljnju analizu ovog skupa podataka. Slike prikazuju utjecaj malog broja opažanja na metodu normalizacije.



Slika 28. Broj prometnih nesreća s teško ozlijedjenim sudionicima po km^2

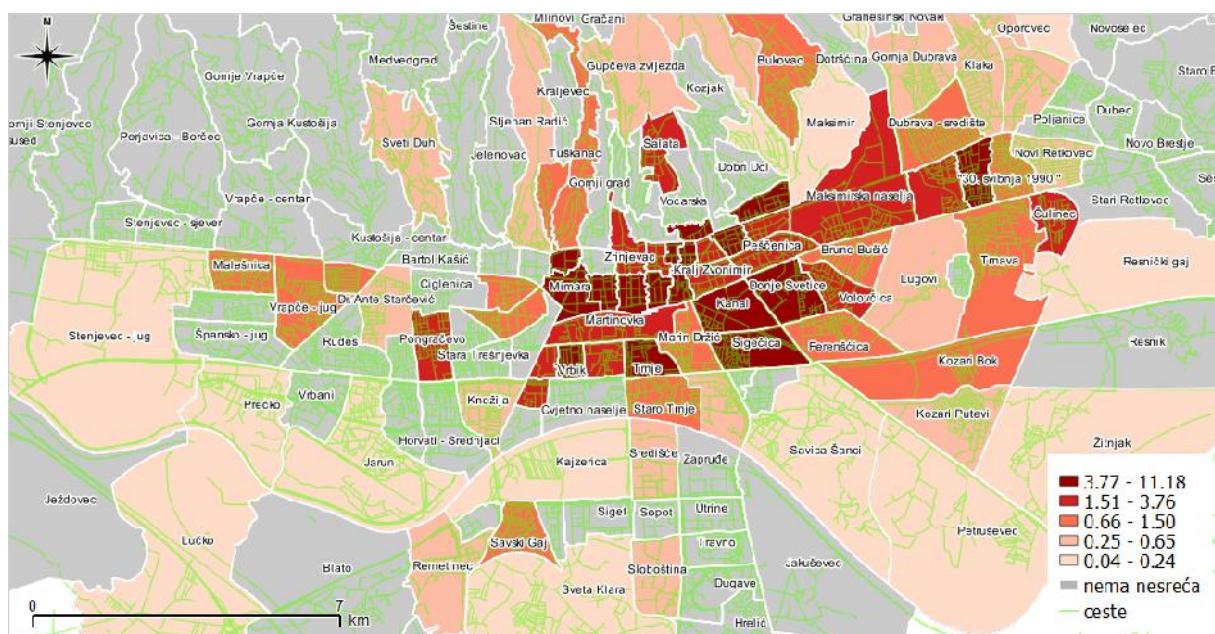
Slika 28 prikazuje distribuciju gustoće prometnih nesreća s teško ozlijedjenim sudionicima. Postoji grupiranje mjesnih odbora s najvećom gustoćom u centru Zagreba i duž glavnih prometnica (Ilica, Savska cesta i Ulica Vjekoslava Heinzena). Ostali mjesni odbori posjeduju znatno manju gustoću. Slikom je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

Rezultati



Slika 29. Broj prometnih nesreća a s poginulim i ozlijedjenim biciklistima po km^2

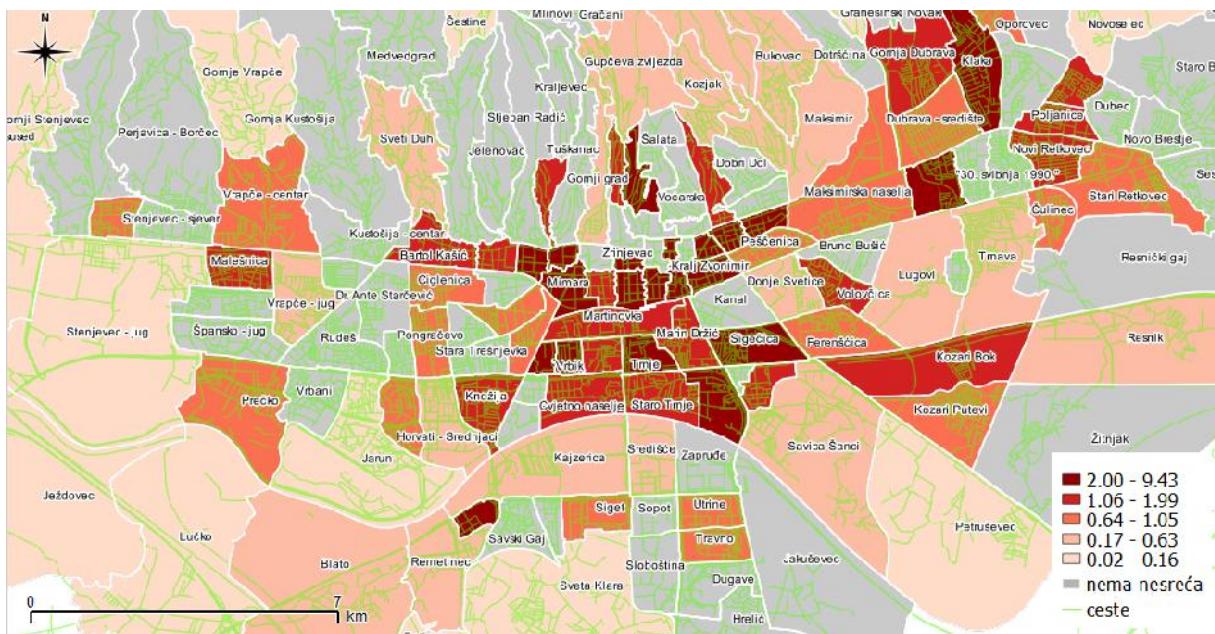
Slika 29 prikazuje distribuciju gustoće prometnih nesreća a s poginulim i ozlijedjenim biciklistima. Uočeno je grupiranje mjesnih odbora s najvećom gustoćom duž Savske ceste, Vlaške ulice, Maksimirske ceste i Ulice Vjekoslava Heinzela. Slika ističe opasne zone za bicikliste.



Slika 30. Broj prometnih nesreća a s poginulim i ozlijedjenim motociklistima po km^2

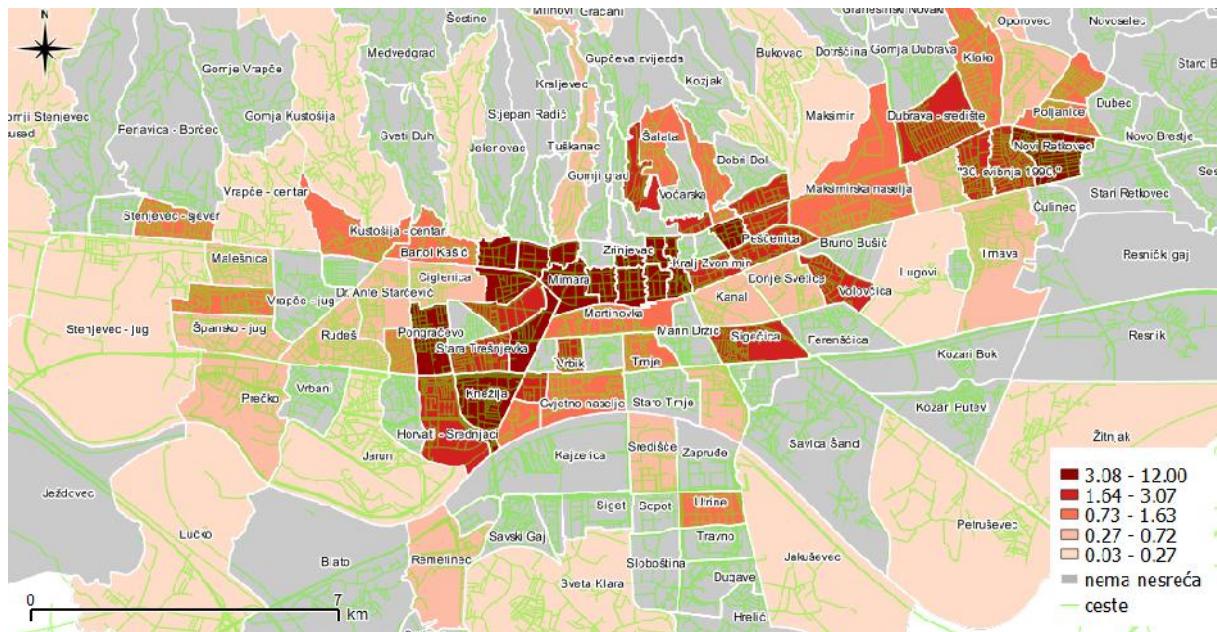
Slika 30 prikazuje distribuciju gustoće prometnih nesreća a s poginulim i ozlijedjenim motociklistima. Na slici su uočena dva grupiranja s najvećom gustoćom. Prvo se odnosi na

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



Slika 31. Broj prometnih nesre a po km² zbog nepropisne brzine

Slika 31 prikazuje distribuciju gusto e prometnih nesre a uzrokovanih nepropisnom brzinom. Uklju ene su sve posljedice prometnih nesre a. Uo ena su dva grupiranja mjesnih odbora s najve om gusto om. Prvo se odnosi na mjesne odbore izme u Ulice kralja Zvonimira i Ulice kneza Branimira te duž Maksimirske ceste. Drugo grupiranje ine mjesni odbori izme u Illice i Ulice Valentina Vodnika. Slika naglašava u kojim zonama treba smanjiti najve u dozvoljenu brzinu.



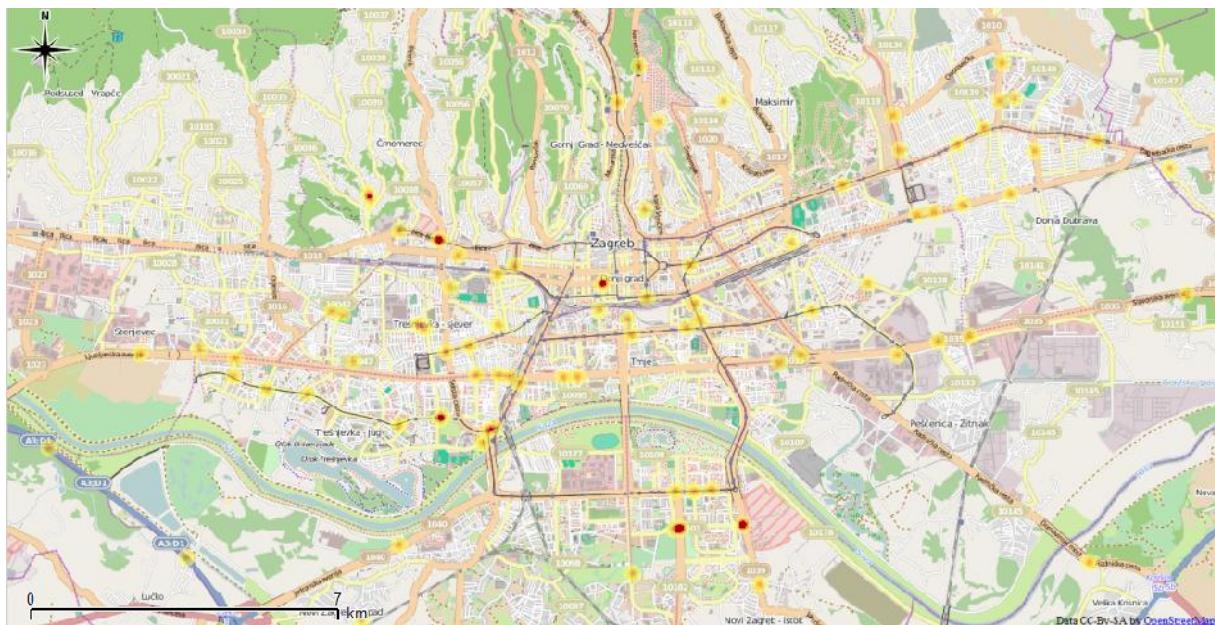
Slika 32. Broj prometnih nesreća po km^2 koje su se dogodile zbog nekorištenja obilježenih pješaka kih prijelaza

Slika 32 prikazuje distribuciju gustoće prometnih nesreća uzrokovanih nekorištenjem obilježenih pješaka kih prijelaza. Uključene su sve posljedice prometnih nesreća. Uočeno je grupiranje mjesnih odbora s najvećim gustoćama 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

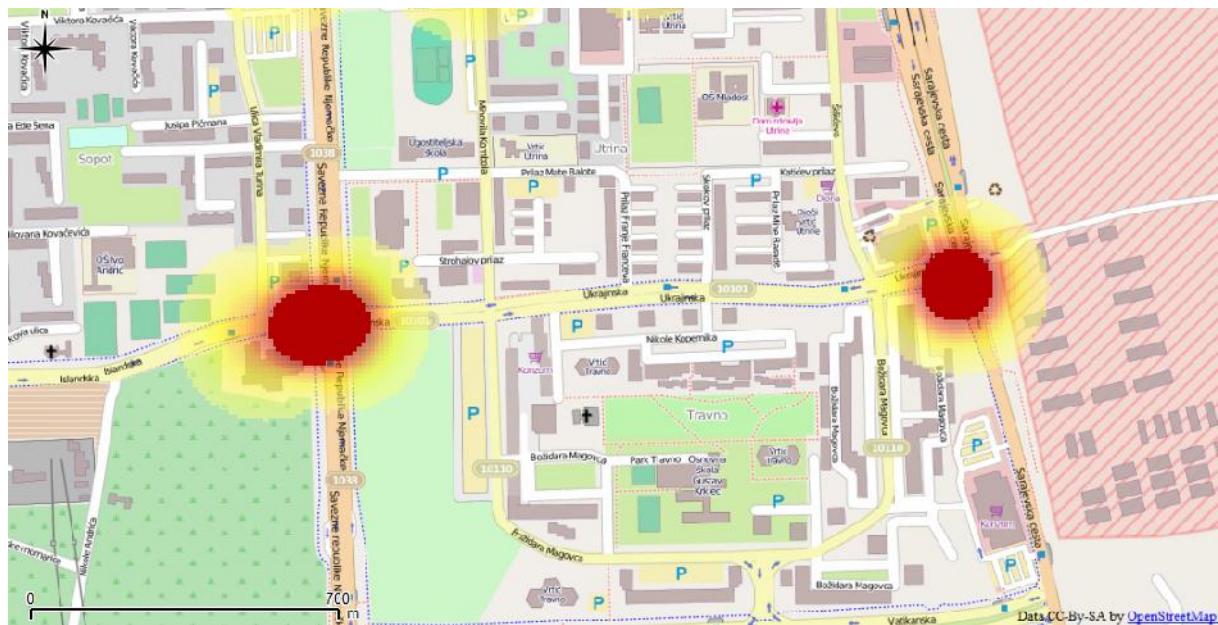
U nastavku su prikazani rezultati KDE metode. Odabrane su vrste prometnih nesre a koje su od najve eg zna aja MUP-u. Prva slika potpoglavlja daje pregled svih žarišta na promatranom podru ju, a zatim slijedi obrada pojedina nih žarišta.

4.4.1. Prometne nesre e s poginulim sudionicima



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

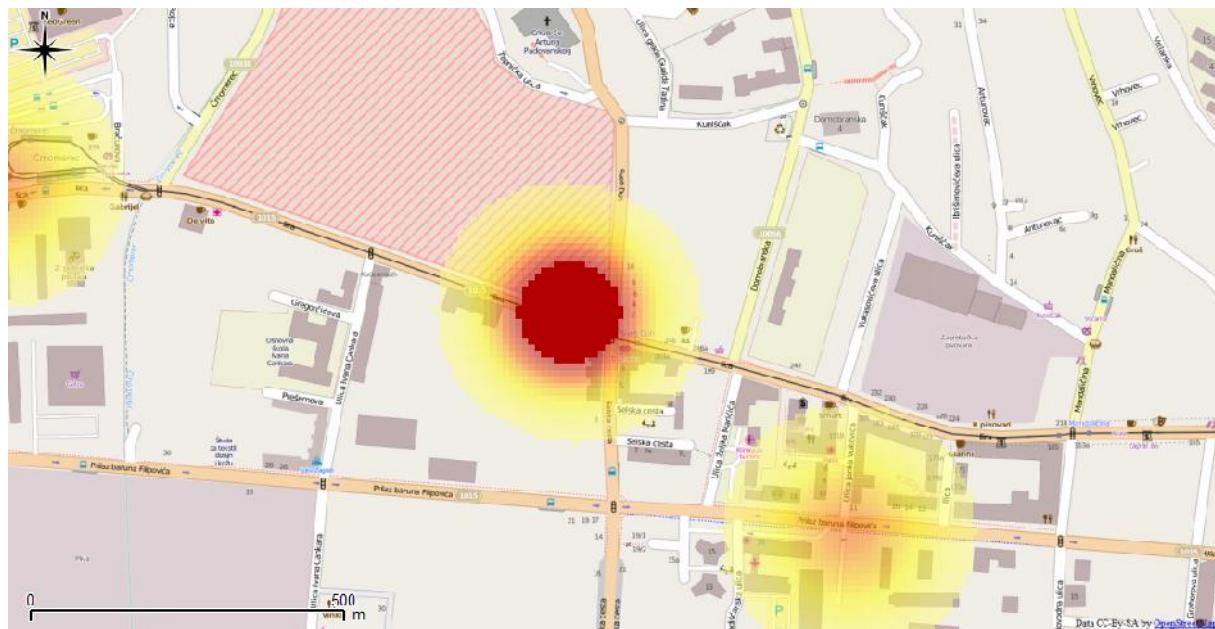
Slika 33 prikazuje žarišta najtežih prometnih nesre a, odnosno prometnih nesre a s poginulim sudionicima. Rije je o malom broju prometnih nesre a, ali su ipak uo ena odre ena žarišta koja su u nastavku detaljnije analizirana. Najteže prometne nesre e su jako važne MUP. Iz tog je razloga ta slika bitna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



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

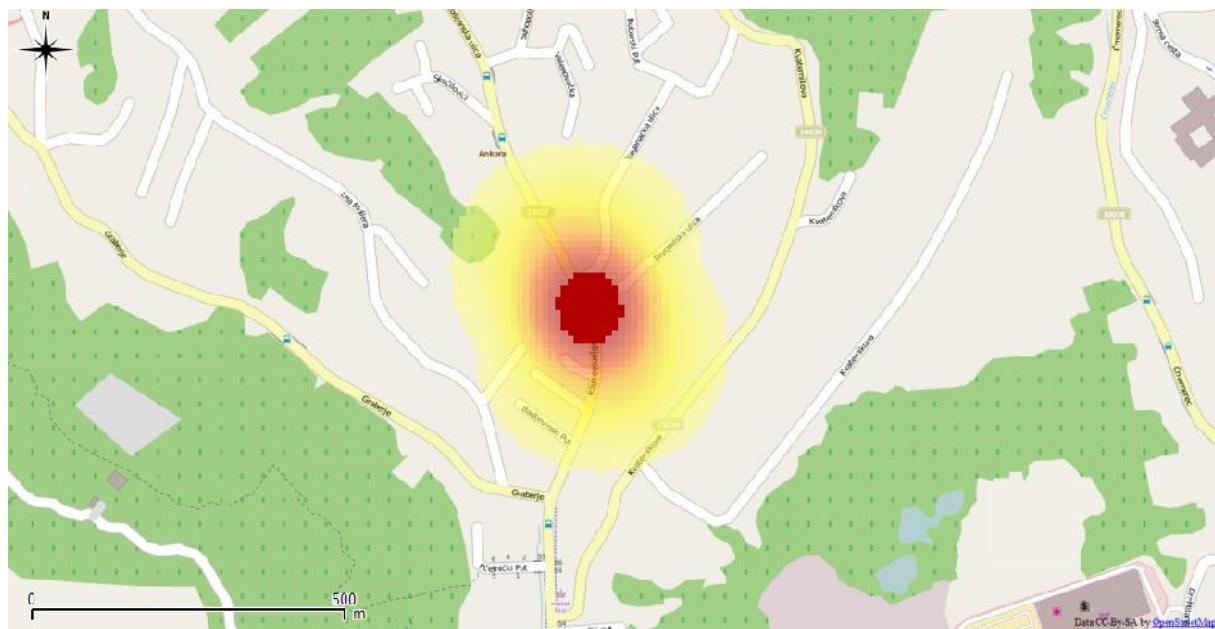
Slika 34 prikazuje dva žarišta. Prvo se nalazi na križanju Ulice Savezne Republike Njemačke i Islandske ulice, odnosno Ukrajinske ulice (lijevo); a drugo na križanju Ukrajinske ulice i Sarajevske ceste (desno). Na lijevom su se žarištu dogodile tri nesreće ovog tipa. Dvije uključuju nalet na pješaka, a treći je slijetanje vozila s ceste. Okolnosti su različite u svim nesrećama. Dogodile su se u popodnevnim satima u stabilnim atmosferskim prilikama. Na desnom su se žarištu dogodila dva bočna sudara unutar mjeseca dana. U oba slučaja je riječ o nepoštivanju prometnih pravila. Nesreće su se dogodile tijekom dana, ali prije različitim vremenskim uvjetima. Ta su žarišta važna zato što isti u mjestu na kojima se najteže nesreće ponavljaju.

Rezultati



Slika 35. Prikaz žarišta najtežih prometnih nesreća (2 nesreće)

Slika 35 prikazuje žarište na križanju Illice i Svetog duha. Riječ je o dva naleta na pješaka zbog neprimjerene brzine. Dogodile su se u ranijim popodnevnim satima i u stabilnim atmosferskim prilikama. Potrebno je istaknuti to žarište zbog sličnosti prometnih nesreća a koje su se dogodile na istom mjestu.

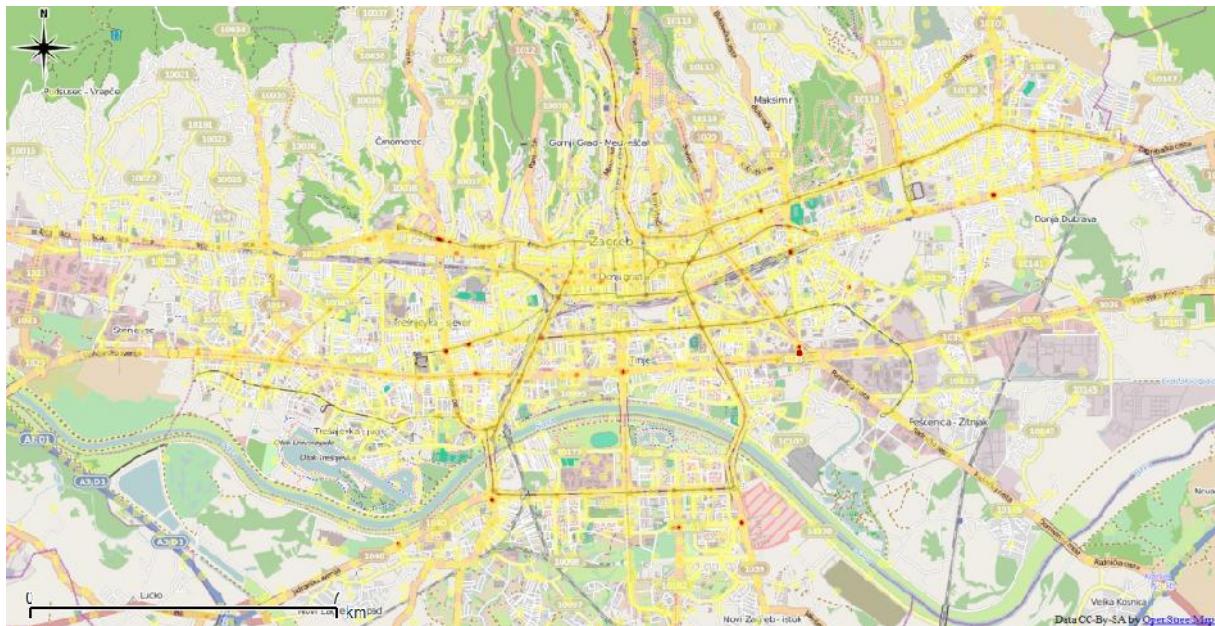


Slika 36. Prikaz žarišta najtežih prometnih nesreća (2 nesreće)

Slika 36 prikazuje žarište na križanju Kustosijanske ulice i Vagana ke ulice. Riječ je o dvije prometne nesreće uzrokovane pogreškama vozača. Dogodile su se tijekom noći i u stabilnim

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

4.4.2. Prometne nesreće s poginulim i ozlijeđenim sudionicima



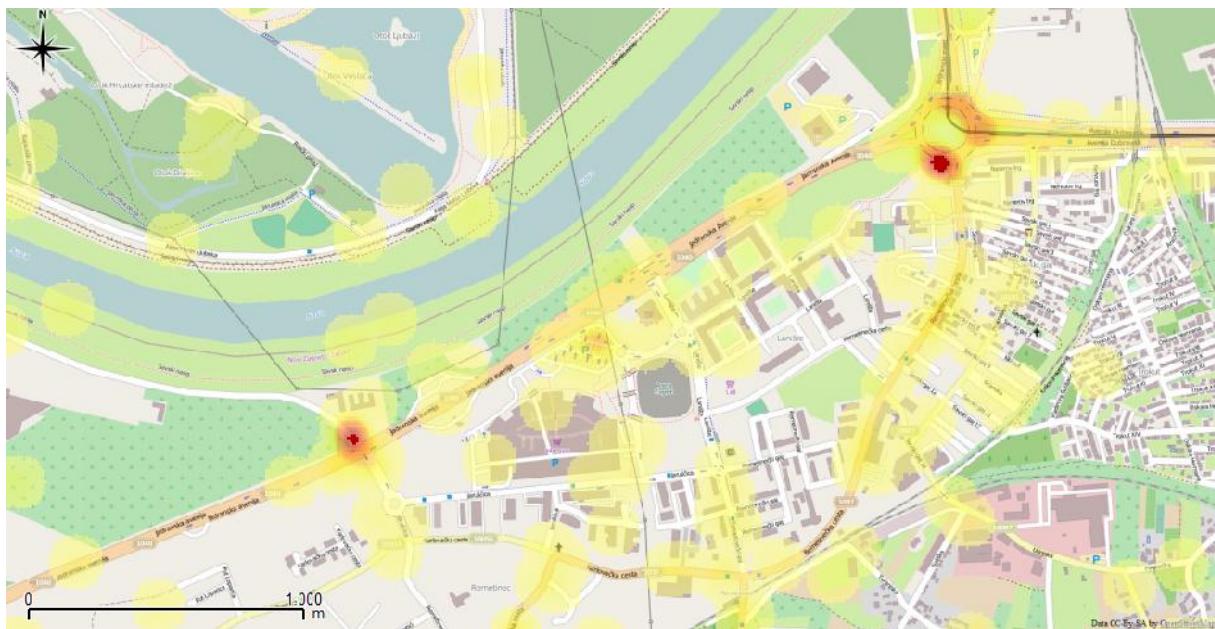
Slika 37. Prikaz žarišta prometnih nesreća s poginulim i ozlijeđenim sudionicima

Uzveši u obzir mali broj prometnih nesreća s poginulim sudionicima, skup podataka je proširen na prometne nesreće s poginulim i ozlijeđenim sudionicima. Cilj je povećanje veličine uzorka radi otkrivanja potencijalnih obrazaca (Slika 37). Slika prikazuje raspored žarišta koja se u nastavku detaljnije obraćaju.



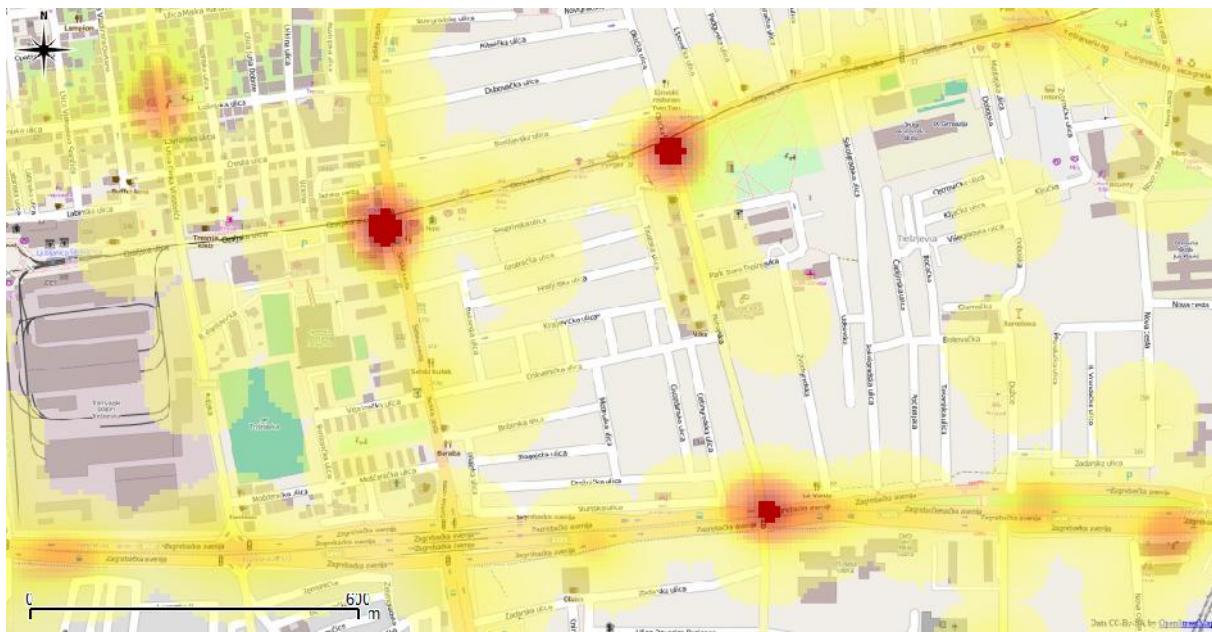
*Slika 38. Prikaz žarišta prometnih nesreća s poginulim i ozlijeđenim sudionicima
(lijevo žarište - 10 nesreća; desno žarište - 11 nesreća)*

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 žarištu prevladavaju nesreće uzrokovane nepoštivanjem prednosti prolaska, nepropisnim uključivanjem u promet i pogreškama pješaka. Najčešći rezultiraju bočnim sudarima, sudarima iz suprotnih smjerova i naletima na pješaka ili biciklista. Veći dio nesreća dogodio se tijekom dana u povoljnim atmosferskim prilikama. Na desnom žarištu prevladavaju prometne nesreće uzrokovane nepoštivanjem prednosti prolaska. One rezultiraju bočnim sudarima i sudarima iz suprotnih smjerova. Događaju se u stabilnim atmosferskim uvjetima tijekom cijelog dana. Nesreće su na toj lokaciji sve češći. Iako nije riječ o velikim žarištima, sliku je potrebno istaknuti zato što su te lokacije očito problematice.



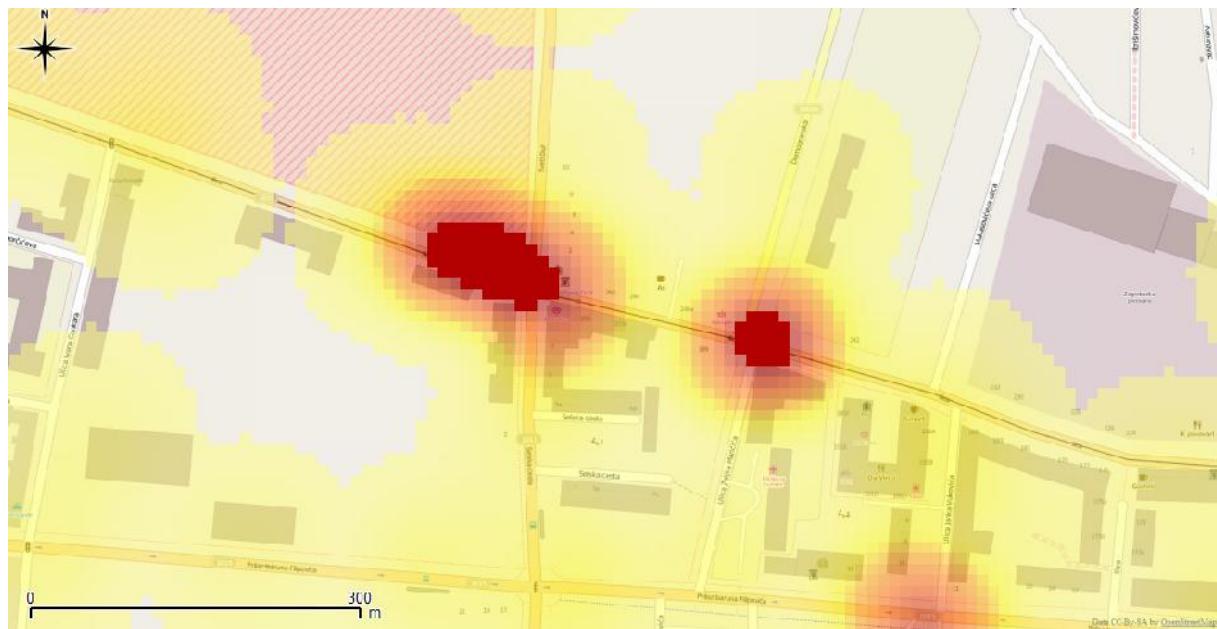
*Slika 39. Prikaz žarišta prometnih nesreća s poginulim i ozlige enim sudionicima
(lijevo žarište - 9 nesreća; desno žarište - 14 nesreća)*

Slika 39 prikazuje dva žarišta. Prvo je na križanju Jadranske avenije i Blata (lijevo), a drugo na južnom dijelu Rotora (desno). Na lijevom žarištu prevladavaju bojni sudari uzrokovani nepoštivanjem svjetlosnog znaka. Druga vrsta nesreća su sudari vozila u slijedu uzrokovani neprimjerenom brzinom. Većina nesreća dogodila se tijekom ljetnih mjeseci u ranijim poslijepodnevnim 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 nesreća su bojni sudari zbog nepoštivanja prednosti prolaska. Većina se nesreća dogodila tijekom kasnih poslijepodnevnih sati, u radnim danima i pri stabilnim atmosferskim prilikama. Žarišta su površinski mala, ali je o tome riječ o velikoj gustoći nesreća i potrebi istražiti.



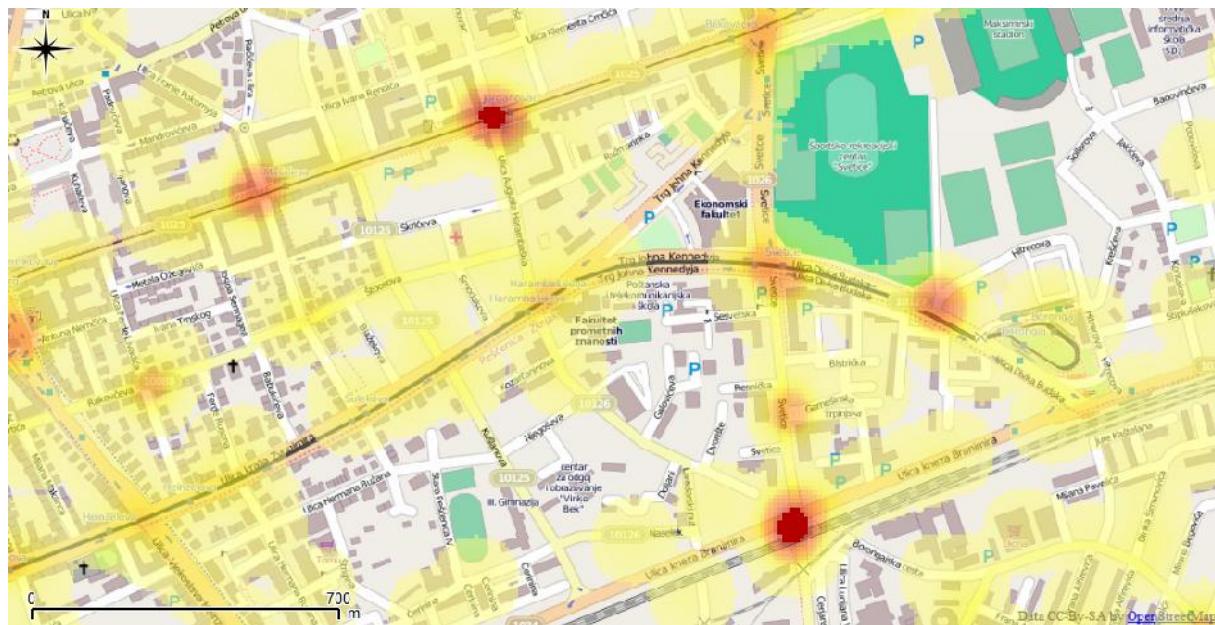
*Slika 40. Prikaz žarišta prometnih nesreća s poginulim i ozlijeđenim sudionicima
(gornje lijevo žarište - 12 nesreća; gornje desno žarište - 9 nesreća; donje žarište - 8 nesreća)*

Slika 40 prikazuje tri žarišta. Prvo je na križanjima Selske ceste i Ozaljske ulice (gore lijevo), drugo na križanju Ozaljske ulice, Okićeve ulice i Nehajske (gore desno), a treće na križanju Nehajske i Zagrebačke avenije (dolje). Na gornjem lijevom žarištu prevladavaju prometne nesreće uzrokovane nepoštivanjem svjetlosnog znaka. Najčešći rezultiraju bočnim sudarima i naletima na motocikliste. Naleti na pješake su većinom uzrokovani pogreškama od strane pješaka. Većina nesreća dogodila se u kasnijim poslijepodnevnim i večernjim satima te u stabilnim atmosferskim prilikama. Na gornjem desnom žarištu prevladavaju nesreće uzrokovane nepoštivanjem svjetlosnog znaka koje rezultiraju bočnim sudarima i sudarima iz suprotnih smjerova. Nesreće su se dogodile pri stabilnim atmosferskim prilikama u prijepodnevnim 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 češći. Kao i u prethodnom slučaju, riječ je o koncentraciji velikog broja nesreća na malim područjima. Tako je, bitno je istaknuti da je većina nesreća na tim lokacijama uzrokovana nepoštivanjem svjetlosnog znaka.



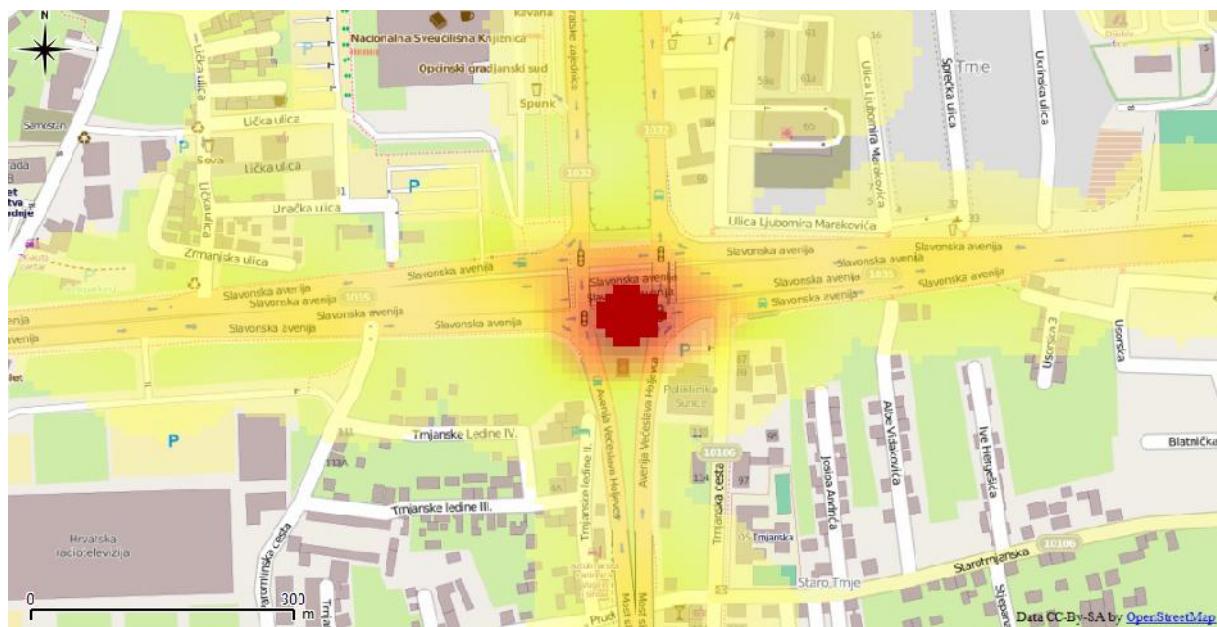
*Slika 41. Prikaz žarišta prometnih nesreća s poginulim i ozlijeđenim sudionicima
(lijevo žarište - 12 nesreća; desno žarište - 15 nesreća)*

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 najčešće prometne nesreće uzrokovane brzinom neprimjerenom uvjetima. Najčešće rezultiraju naletom na pješaka i sudarom vozila u slijedu. Većina nesreća dogodila se 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 posljepodnevnim satima i pri stabilnim atmosferskim prilikama. Na toj lokaciji nema semafora. Potrebno je istaknuti žarišta u kojima većinom stradaju pješaci.



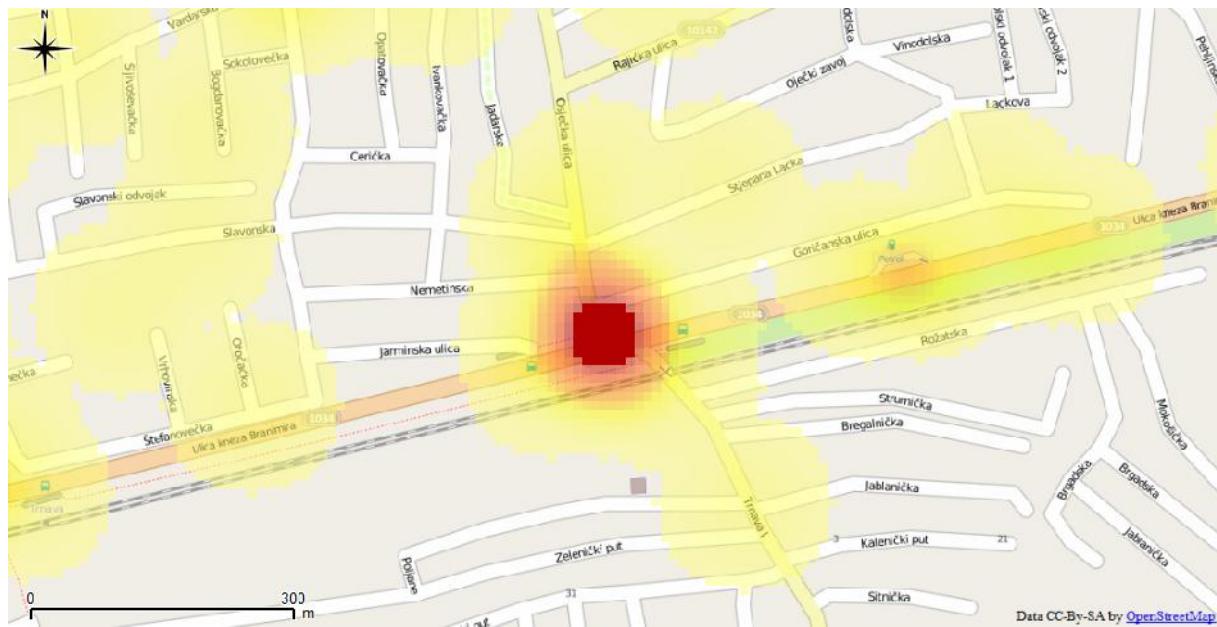
*Slika 42. Prikaz žarišta prometnih nesreća s poginulim i ozlige enim sudionicima
(gornje žarište - 12 nesreća; donje žarište - 18 nesreća)*

Slika 42 prikazuje dva žarišta. Prvo je na križanju Maksimirske ceste s Jordanovcem i Ulicom Augusta Harambaši a (gore), a drugo na križanju Ulice kneza Branimira sa Sveticama i Donjim Sveticama (dolje). Na gornjem žarištu prevladavaju bojni sudari uzrokovani nepoštivanjem svjetlosnog znaka i naleti na pješake uzrokovani nepoštivanjem svjetlosnog znaka, uglavnom od strane pješaka. Većina nesreća dogodila se u poslijepodnevnim satima pri stabilnim atmosferskim prilikama. Na donjem žarištu prevladavaju nesreće uzrokovane nepoštivanjem prednosti prolaska i nepoštivanjem svjetlosnog znaka. Najčešći su rezultirajući bojni sudarima i sudarima iz suprotnih smjerova. Događaju se u poslijepodnevnim satima i stabilnim atmosferskim prilikama. Uočeno je znatno smanjenje broja navedenih vrsta nesreća, ali se povećao broj naleta na pješake koji ne poštuju svjetlosne znakove. Slika jasno prikazuje da je uzrok većine prometnih nesreća a na tom području nepoštivanje svjetlosnog znaka.



Slika 43. Prikaz žarišta prometnih nesreća s poginulim i ozlige enim sudionicima
(24 nesreće)

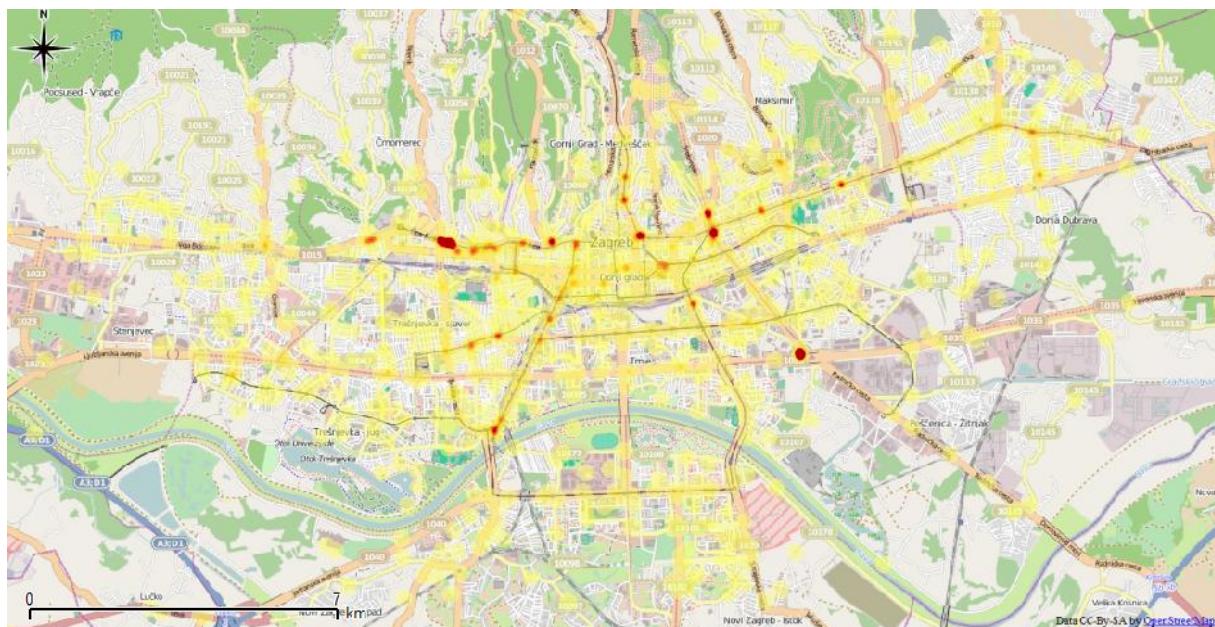
Slika 43 prikazuje žarište na križanju Ulice Hrvatske bratske zajednice i Slavonske avenije. Prevladavaju nesreće uzrokovane nepoštivanjem svjetlosnog znaka i nepropisnim uključivanjem u promet. Najčešće rezultiraju bočnim sudarima i naletima na bicikliste. Nesreće se događaju tijekom cijelog dana i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



*Slika 44. Prikaz žarišta prometnih nesreća s poginulim i ozlige enim sudionicima
(19 nesreća)*

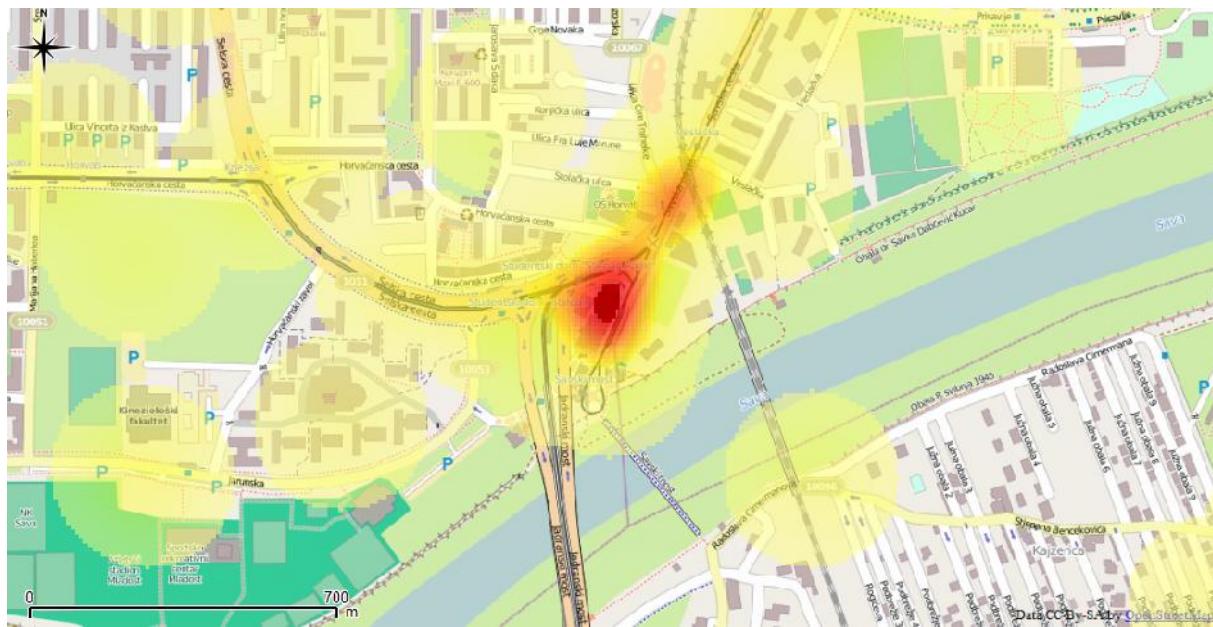
Slika 44 prikazuje žarište na križanju Osjeke ulice, Ulice kneza Branimira i Trnava I. Tu prevladavaju nesreće uzrokovane nepoštivanjem svjetlosnog znaka, nepoštivanjem prednosti prolaska i brzine neprimjerene uvjetima. Najčešći su rezultiraju sudarom vozila u slijedu. Većina nesreća se dogodila tijekom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

4.4.3. Prometne nesreće s poginulim i ozlijeđenim pješacima



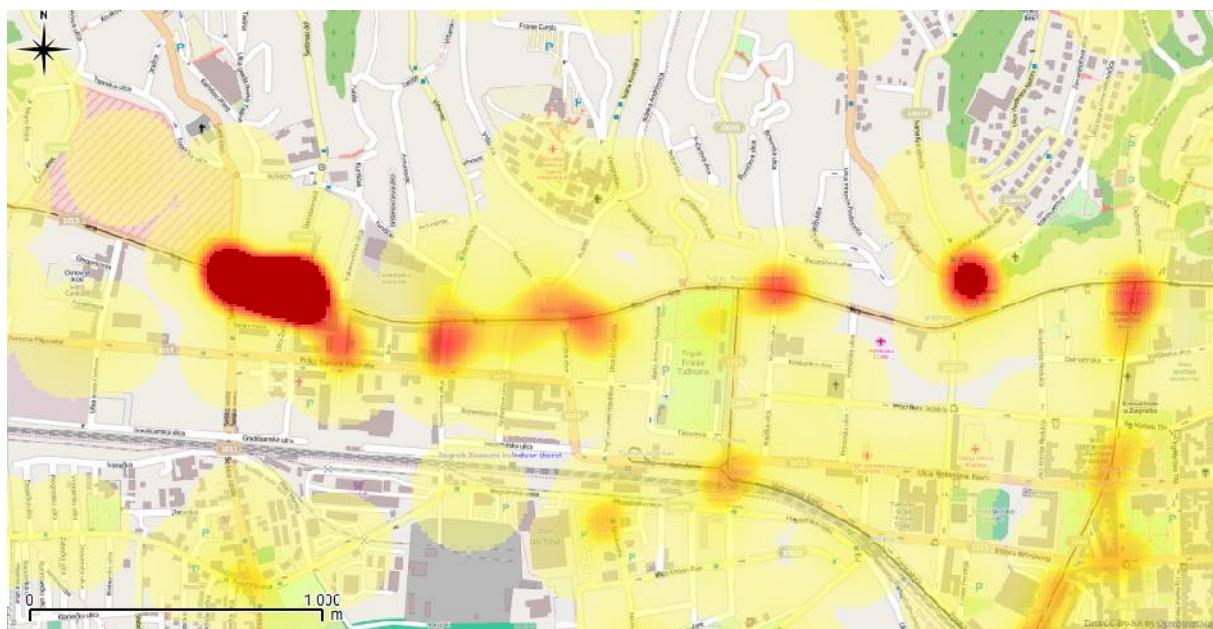
Slika 45. Prikaz žarišta prometnih nesreća s poginulim i ozlijeđenim pješacima

U prethodnoj je analizi uočeno da su naleti na pješaka najčešći uzrokovani brzinom neprimjerenom uvjetima i nepoštivanjem svjetlosnog znaka od pješaka. Budući da su pješaci jedna od najugroženijih skupina sudionika u prometu, sljedeći je korak obrada prometnih nesreća s poginulim i ozlijeđenim pješacima (Slika 45). Slika prikazuje koja su područja najopasnija za pješake.



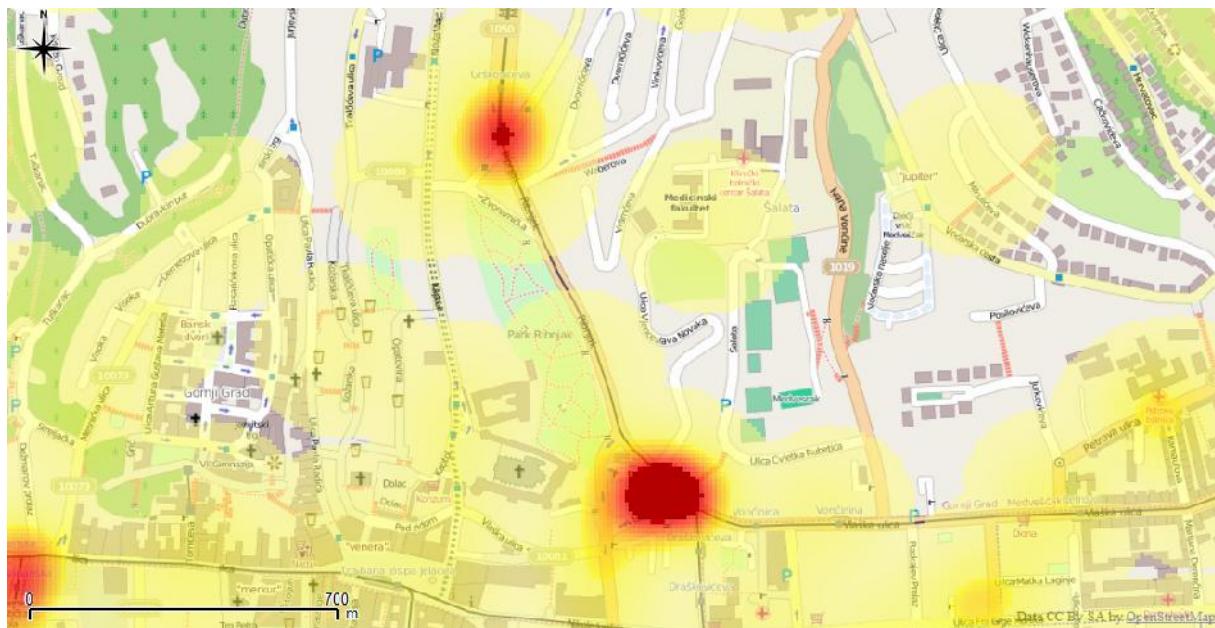
Slika 46. Prikaz žarišta prometnih nesreća s poginulim i ozlijeđenim pješacima (11 nesreća a)

Slika 46 prikazuje žarište na križanjima Selske ceste i Jadranskog mosta. Na donjem žarištu prevladavaju nesreće uzrokovane nepoštivanjem pravila prometa od strane pješaka i brzinom neprimjerenom uvjetima. Većina nesreća se dogodila tijekom noći u stabilnim atmosferskim prilikama. Potrebno je istaknuti žarište zbog velike koncentracije nesreća.



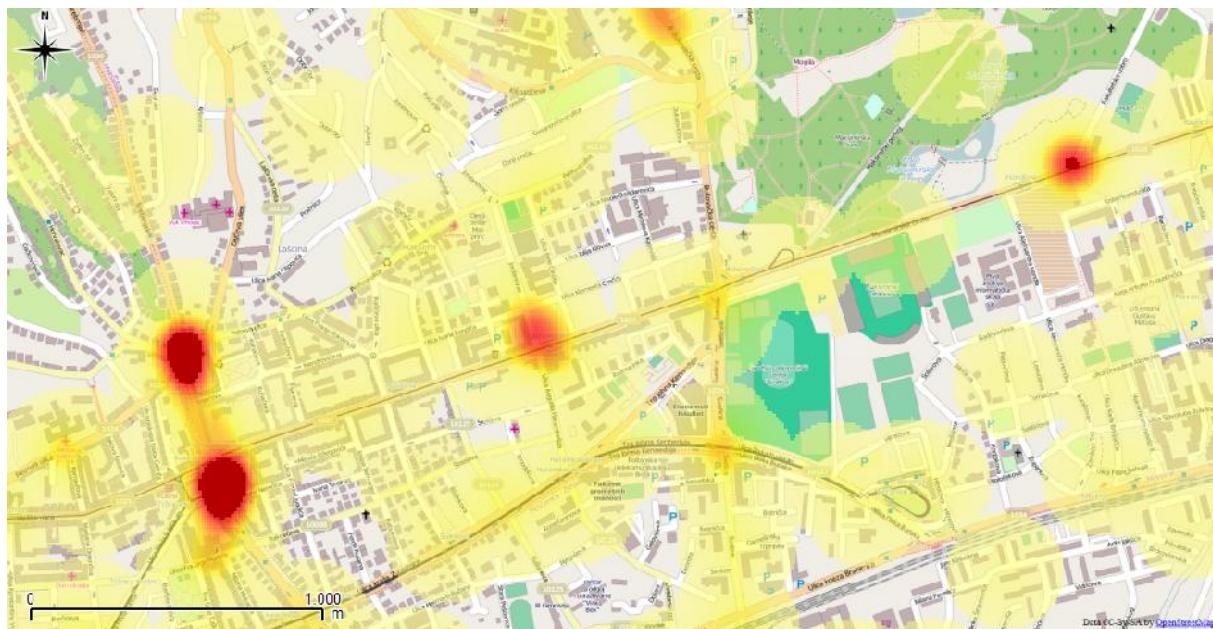
Slika 47. Prikaz žarišta prometnih nesreća s poginulim i ozlijeđenim pješacima (lijevo žarište - 24 nesreća a; desno žarište - 9 nesreća a)

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



*Slika 48. Prikaz žarišta prometnih nesreća s poginulim i ozlijedjenim pješacima
(donje žarište - 11 nesreće a; gornje žarište - 8 nesreće e)*

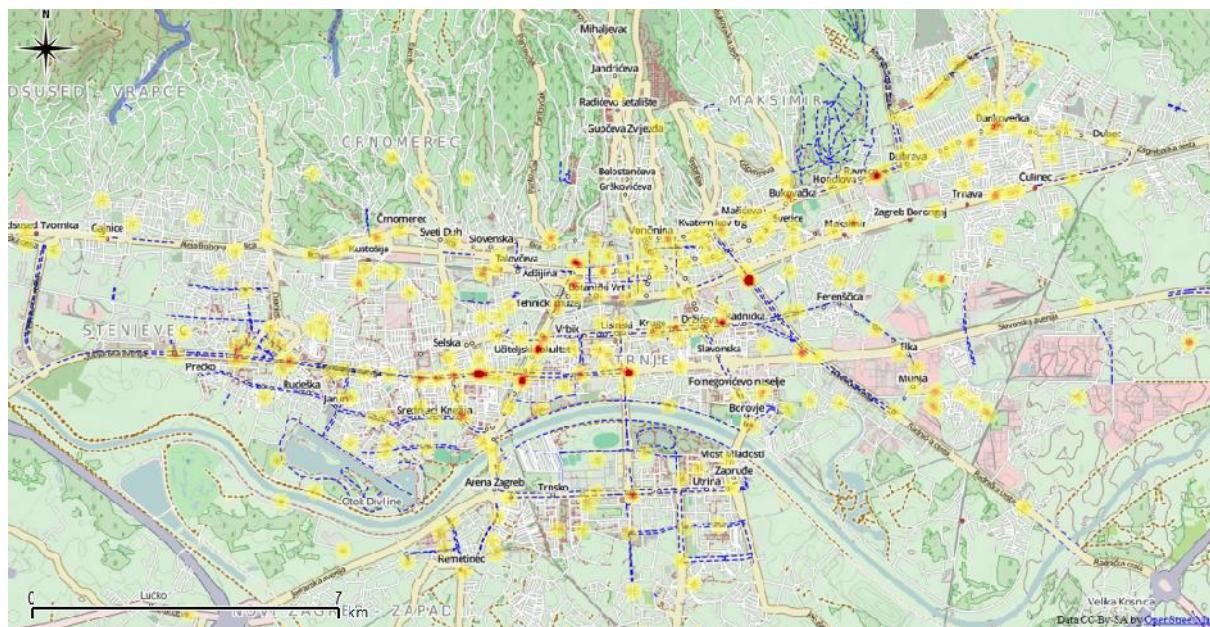
Slika 48 prikazuje dva žarišta. Prvo na Trgu Josipa Langa (dolje), a drugo na križanju Medveščaka i Ulice Nike Grškovića (gore). Na donjem žarištu prevladavaju nesreće uzrokovane brzinom neprimjerenom uvjetima i nepropisnim skretanjem. Dogodaju se tijekom cijelog dana i u stabilnim atmosferskim prilikama.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



*Slika 49. Prikaz žarišta prometnih nesreća s poginulim i ozlijednim pješacima
(donje lijevo žarište - 11 nesreća; gornje lijevo žarište - 11 nesreća;
gornje desno žarište - 7 nesreća)*

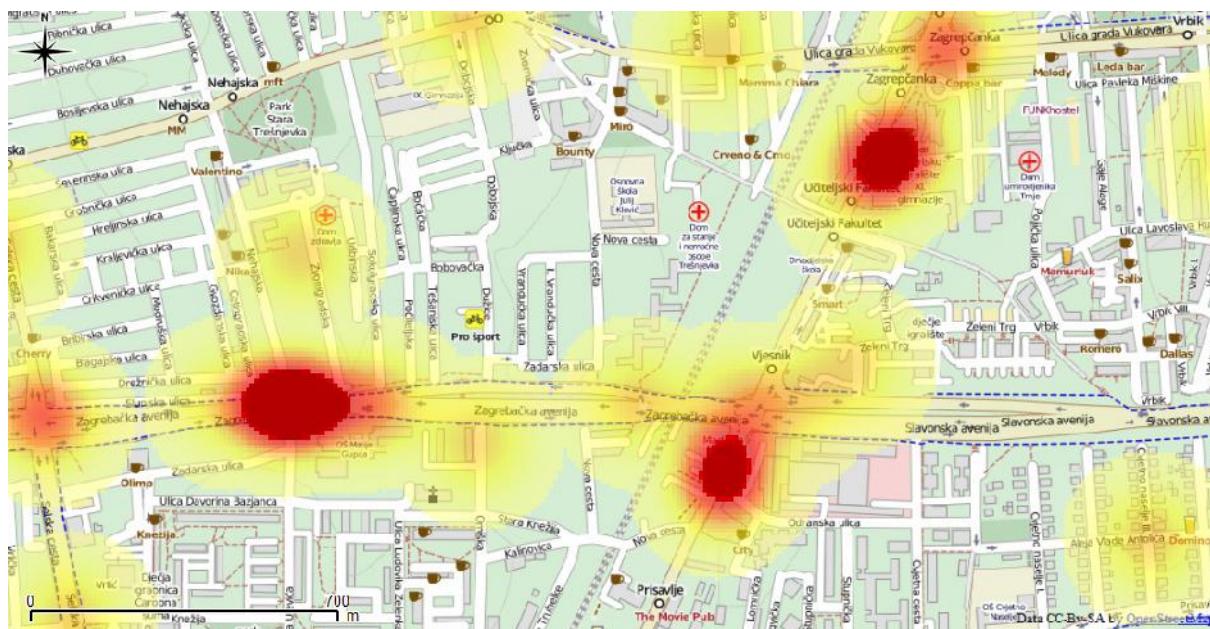
Slika 49 prikazuje tri žarišta. Prvo na kružnom toku gdje se spajaju Ulica Dragutina Domjanića, Petrova ulica, Vodarska ulica i Srebrnjak (gore lijevo), drugo na Trgu Eugena Kvaternika (dolje lijevo), a treće na križanju Maksimirske ceste i Ulice Fakultetskog dobra (gore desno). Na gornjem lijevom žarištu prevladavaju nesreće uzrokovane brzinom neprimjerenom uvjetima, a većina se dogodila tijekom 2013. godine. Tijekom 2013. godine sve su se nesreće dogodile na mokrom kolniku i u pretežito večernjim 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. Većina nesreća dogodila se tijekom dana na mokrom kolniku. Na toj lokaciji unutar zadnje dvije godine nije zabilježena nijedna nesreća te vrste. Slika je izdvojena jer prikazuje utjecaj vremenskih uvjeta na prometne nesreće i prikazuje napredak u regulaciji prometa.

4.4.4. Prometne nesreće s poginulim i ozlijeđenim biciklistima



Slika 50. Prikaz žarišta prometnih nesreća a s poginulim i ozlijeđenim biciklistima

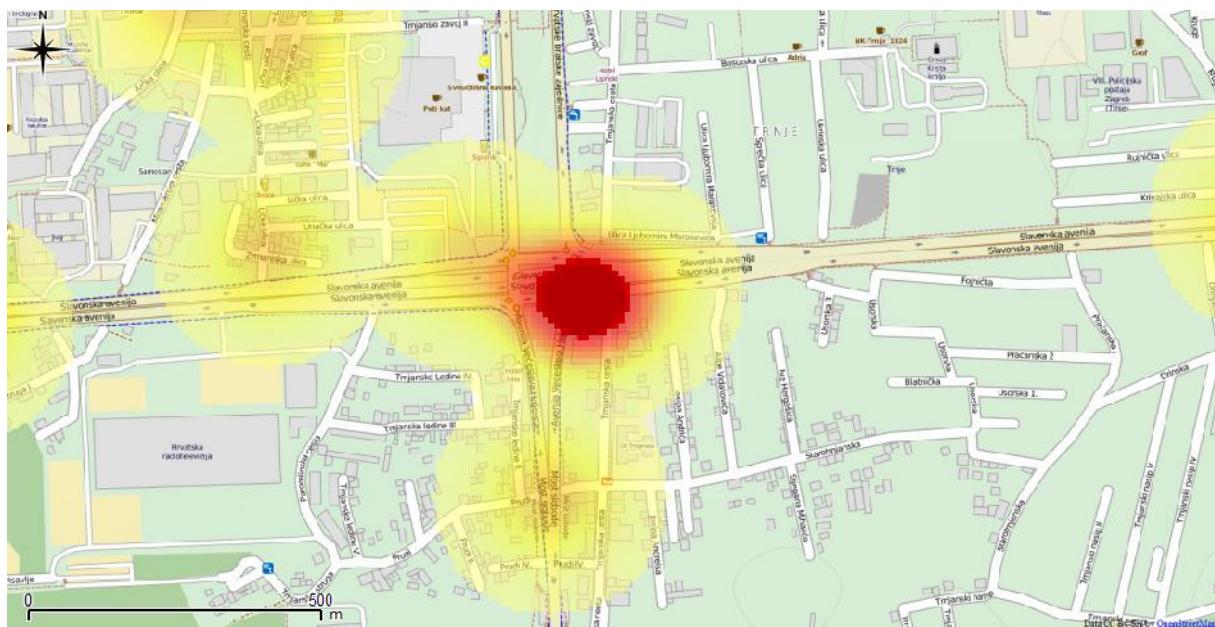
Druga najugroženija skupina sudionika u prometu su biciklisti. U nastavku se analiziraju prometne nesreće sa poginulim i ozlijeđenim biciklistima. Treba napomenuti da je kao pozadina korišten OpenCycleMap OpenLayersa zbog kartiranih biciklističkih staza (Slika 50). Slika služi za bolju orientaciju pri analizi žarišta koja su analizirana u nastavku.



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

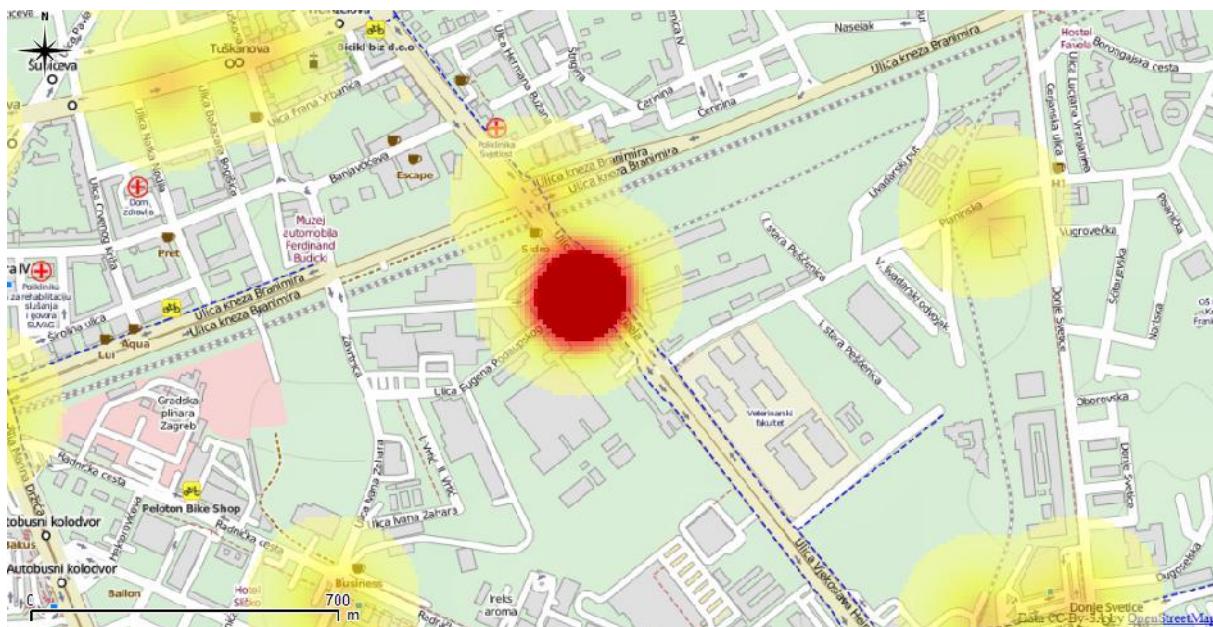
Rezultati

Slika 51 prikazuje tri žarišta. Prvo na križanju Zagreba ke avenije, Ulice Tina Ujevića i Nehajske ulice (dolje lijevo), drugo na križanju Savske ceste i Odranske ulice (dolje desno), a treće na Savskoj cesti kod Univerzitetskog fakulteta (gore). Na donjem lijevom žarištu prevladavaju nesreće uzrokovane nepoštivanjem svjetlosnog znaka i nepoštivanjem prednosti prolaska. Većina nesreća dogodila se 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.



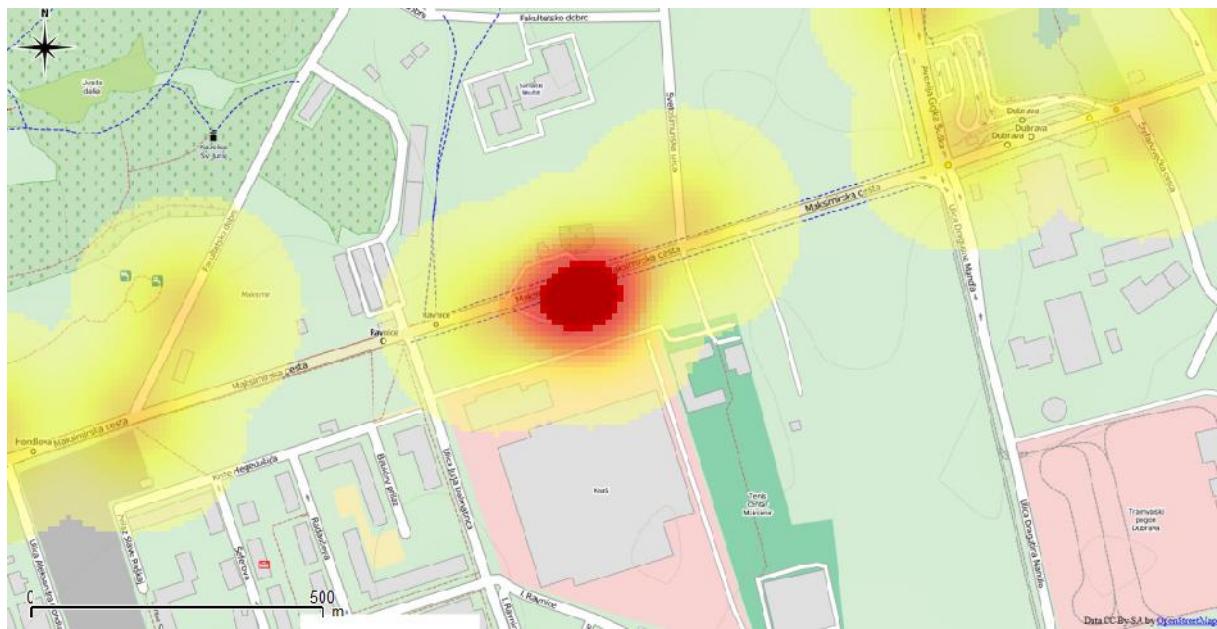
*Slika 52. Prikaz žarišta prometnih nesreća s poginulim i ozlijeđenim biciklistima
(5 nesreće a)*

Slika 52 prikazuje žarište na križanju Slavonske avenije i Avenija Vojislava 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 nesreća s poginulim i ozlijeđenim biciklistima
(6 nesreća)

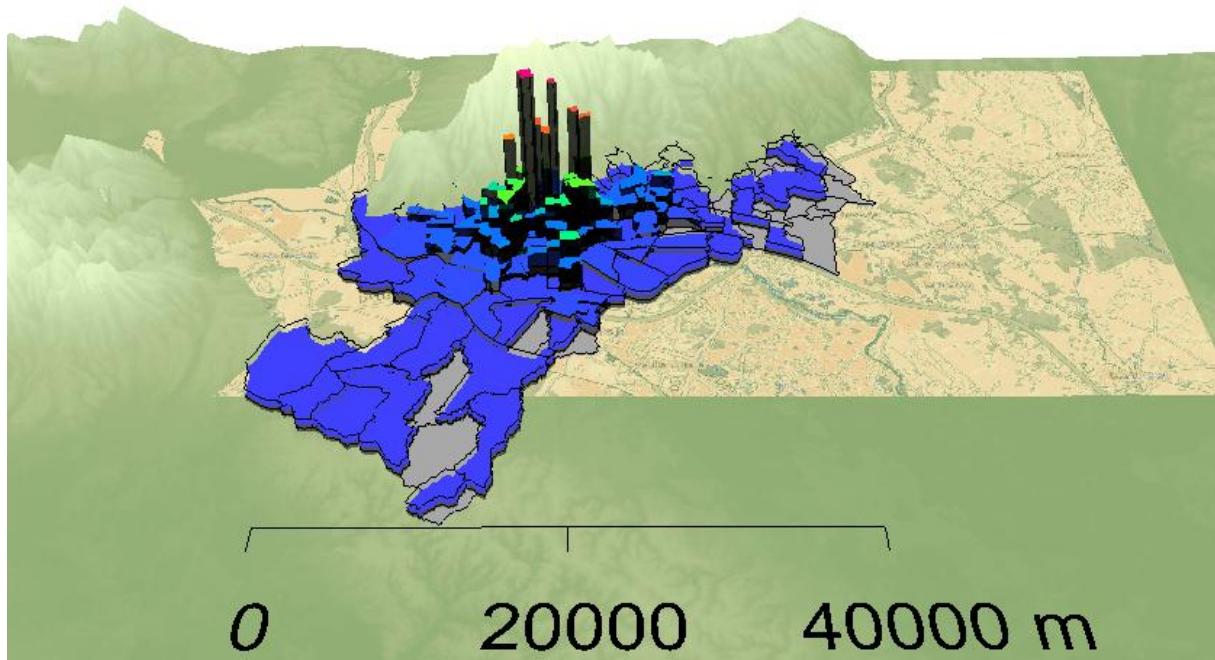
Slika 53 prikazuje žarište na križanju Ulice Vjekoslava Heinzela i Ulice Eugena Podaupskog. Većina nesreća je uzrokovana nepropisnim uključivanjem u promet. Sve su se dogodile tijekom dana, između svibnja i kolovoza te u stabilnim atmosferskim prilikama. Slika je izdvojena zbog velike koncentracije nesreća.



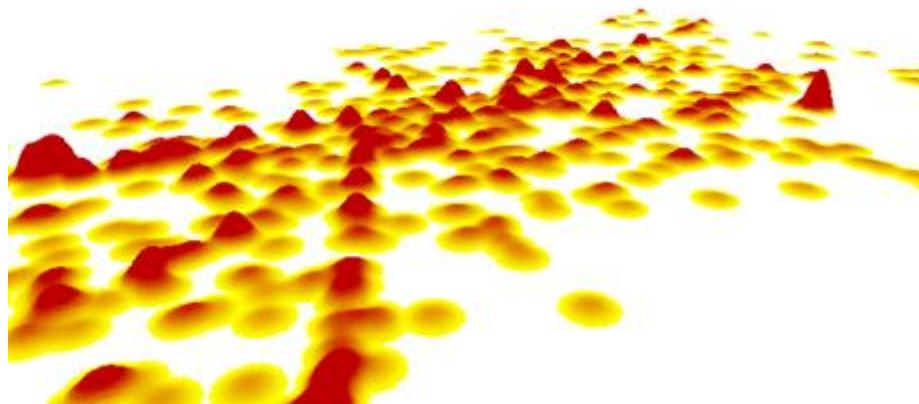
*Slika 54. Prikaz žarišta prometnih nesreća s poginulim i ozlijeđenim biciklistima
(5 nesreće)*

Slika 54 prikazuje žarište na Maksimirskoj cesti južno od Šumarskog fakulteta. Sve su nesreće uzrokovane nepropisnim uključivanjem u promet. Dogodile su se u stabilnim atmosferskim prilikama, a većina se dogodila tijekom dana. Slika je izdvojena zbog velike koncentracije nesreća.

4.5. Dodatne mogunosti vizualizacije rezultata



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



Slika 56. 3D prikaz žarišta prometnih nesreća s pješacima

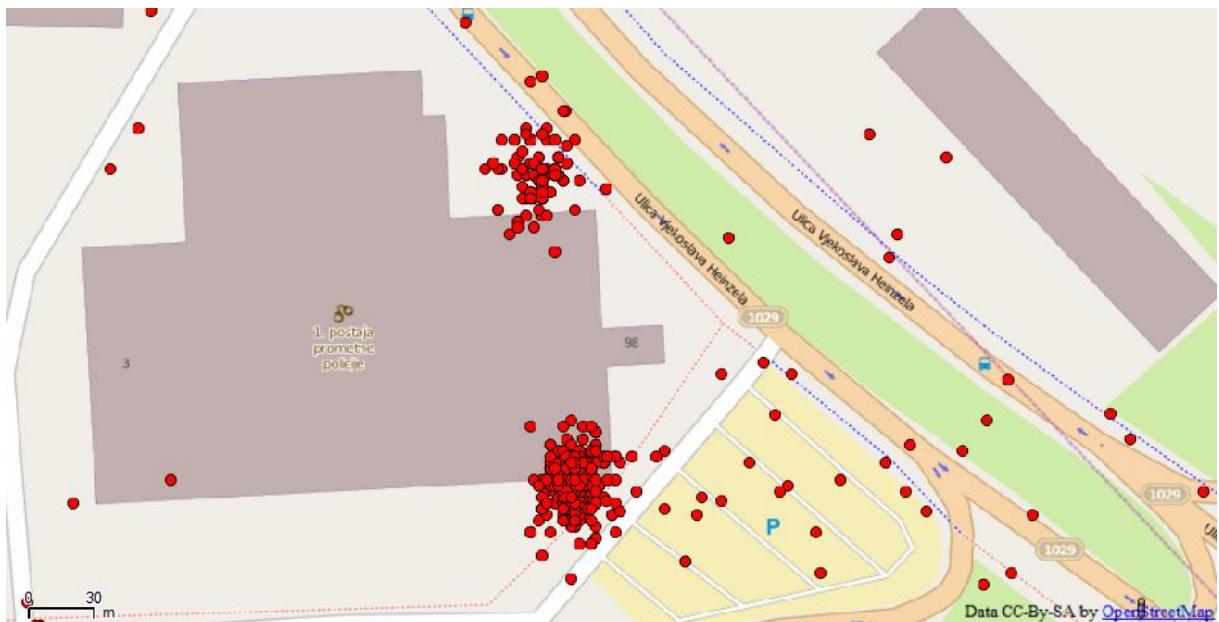
Slika 55 prikazuje 3D prikaz distribucije gustoće prometnih nesreća s teško ozlijedjenim sudionicima. Prikaz je izrađen u ILWIS aplikaciji. Slika 56 prikazuje 3D prikaz žarišta prometnih nesreća s pješacima izrađen u SAGA GIS aplikaciji. Ta dva primjera predstavljaju dodatne mogunosti vizualizacije koji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

4.6. Problemi uo eni tijekom obrade i analize podataka



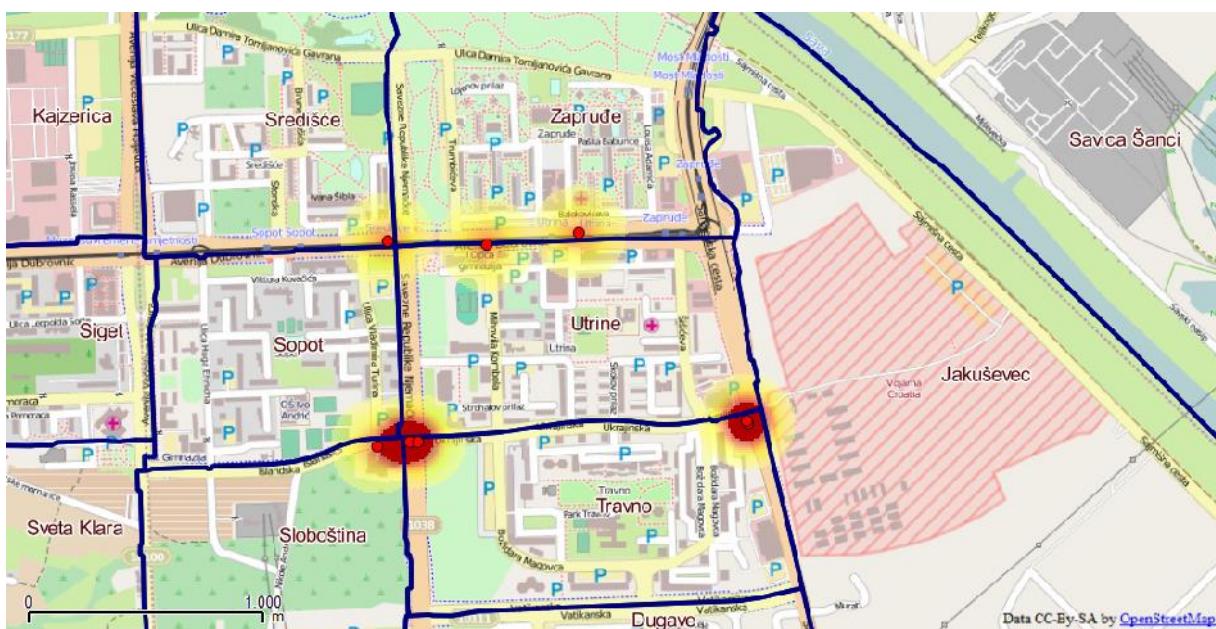
Slika 57. Prikaz loše pozicioniranih prometnih nesreća

Slika 57 predstavlja jedan od nedostataka fiksiranja GNSS prijamnika 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 unesreće ena vozila iz prometa. U poglavlju *Rasprava* je taj problem detaljnije analiziran.

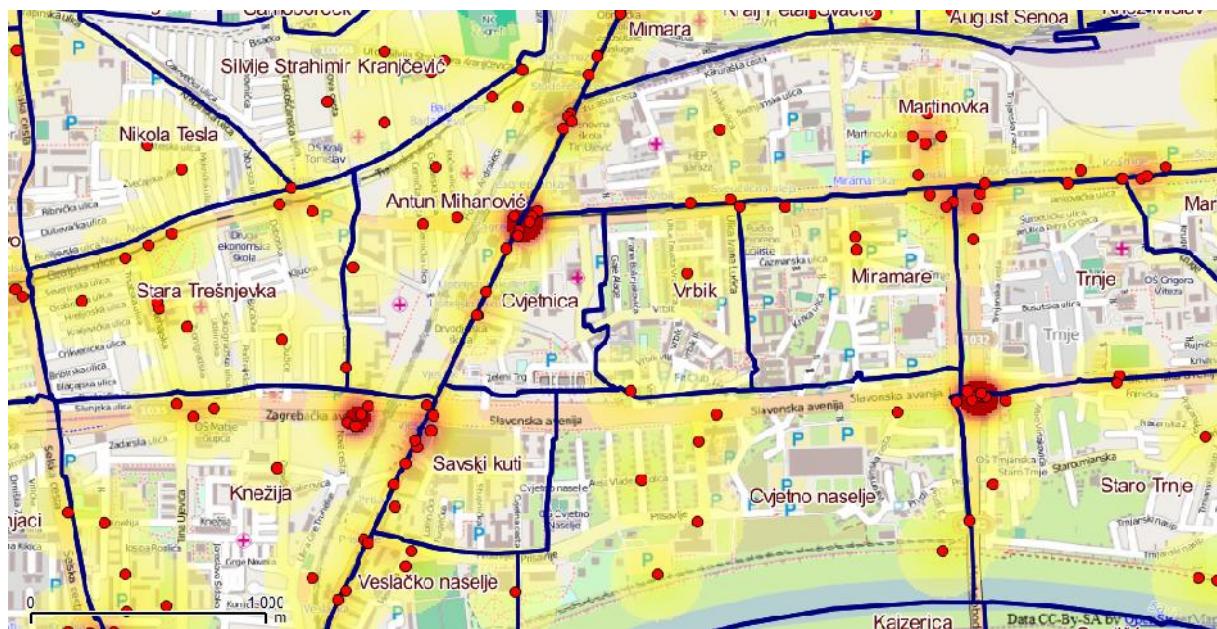


Slika 58. Prikaz pogrešno pozicioniranih prometnih nesreća

Slika 58 prikazuje krivo pozicionirane nesreće na 1. postaji prometne policije. Na toj je lokaciji zabilježeno oko 700 prometnih nesreća. Za većinu nesreća lako se može odrediti da su krivo pozicionirane na osnovi atributa ulica, ali problem su nesreće koje su se dogodile u Ulici Vjekoslava Heinzela ili na Slavonskoj aveniji. Budući da je riječ o dugačkim prometnicama, situacija je još gora ako ne piše kući ni broj. U tom slučaju postoji vjerojatnost da su nesreće koje su se dogodile na ili blizu križanja Ulice Vjekoslava Heinzela i Slavonske avenije dobro pozicionirane, ali se gube u oblaku krivo pozicioniranih točaka. S godinama se trend krivo pozicioniranih točaka na toj lokaciji smanjuje, s oko 350 nesreća u 2010. godini na 50 u 2013. Slika prikazuje napredak MUP-a u popravljanju uočenih nedostataka.



Slika 59. Prikaz prometnih nesreća s poginulim sudionicima pomoći u točkastih oznaka unutar pojedinih mjesnih odbora Grada Zagreba



Slika 60. Prikaz prometnih nesreća s teško ozlijedima jednim sudionicima pomoći u to kastim oznaka unutar pojedinih mjesnih odbora Grada Zagreba

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

5. Rasprava

Pou eni iskustvom istraživanja iz Hong Konga (Loo 2006), prije obrade podataka utvr ena je kvaliteta i iskoristivost podataka. Kvaliteta zapisnika daleko je od idealne i najve i izazov predstavlja georeferenciranje. Napisan je algoritam u Pythonu koji ispravlja prostorne koordinate pomo u naziva ulica iz zapisnika. Algoritam se pokazao uspješnim u tom zadatku. Pojavio se manji broj prometnih nesre a optere en grubim pogreškama koje nisu mogle biti ispravljene. Rije je o zanemarivom broju koji je isklju en iz daljnje analize.

Nužno je otkriti uzroke loših zapisa prostornih koordinata. To ne uklju uje samo pregled postoje e baze, nego i na in unosa podataka. Kao što je ranije navedeno, prometni službenici koriste GNSS prijamnike fiksirane u službenim vozilima. To se pokazalo kao loša praksa. esto moraju isklju iti unesre ena vozila iz prometa i to ih udaljava od mesta nesre e. Policijski službenici nisu u mogu nosti uzeti GNSS prijamnik i zabilježiti pravu lokaciju nesre e, ve o itavaju položaj svog automobila nakon isklju ivanja iz prometa. To rezultira grubim pogreškama koje zahtijevaju ru no ispravljanje.

Sljede i nedostatak je korištenje analognih zapisnika. Ru no moraju unijeti golemu koli inu podataka u kratkom vremenu. Unos podataka u bazu podataka naj eš e provodi osoba koja nije bila na mjestu nesre e i ne može provjeriti pouzdanost podataka. Ne postoji sustav provjere pouzdanosti podataka. Prostorne koordinate omogu uju vizualizaciju lokacije nesre e, a policijski službenici rade skice na mjestu nesre e. Logi no 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 taj jednostavan postupak.

Mogu e rješenje je ukidanje dosadašnjeg na ina rada i uspostava automatiziranog sustava registracije podataka. Tako er treba preispitati smisao atributa i na in unosa podataka. Nema potrebe optere ivati policijske službenike unosom prostornih koordinata i ostalih atributa koje sustav može u pozadini registrirati umjesto njih. Zapisnik trenutno sadrži ogromnu koli inu atributa koji nisu prikladni za provo enje analiza. Važni atributi poput datuma i vremena nesre e podijeljeni su 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 zna i da MUP može tijekom itanja rada zapo eti na uspostavi smislenog i kvalitetnog sustava.

Prva metoda statisti ke obrade podataka korištena u radu je hi-kvadrat test homogenosti. Provedena su tri testa s ciljem ispitivanja proporcionalnosti broja prometnih nesre a s razli itim posljedicama po odre enim atributima. Ti su atributi: karakteristike ceste, dijelovi dana te mjeseci u godini.

Vrlo bitan korak u provedbi testa je postavljanje nul-hipoteze. Kako je rije o testovima homogenosti, sve su nul-hipoteze postavljene na sli an na in. U sva tri slu aja nul-hipoteza nalaže da su omjeri broja posljedica prometnih nesre a jednako raspore eni s obzirom na analizirane attribute (karakteristike cesta, dijelovi dana i mjeseci). Temeljem rezultata svih triju testova, koji su detaljnije izloženi u poglavlju *Rezultati*, navedene nul-hipoteze su odbrane.

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

Metoda ima mnogo nedostataka. Postoje odre eni uvjeti, navedeni u poglavlju *Materijal i metode*, koji moraju biti ispunjeni kako bi metoda uop e bila primjenjiva. Metoda nije prikladna za detaljniju analizu prostorne distribucije prometnih nesre a jer ne omoguava vizualizaciju njihovog smještaja.

Iako se metodom ne može detaljnije analizirati prostorna distribucija prometnih nesre a, mogu e 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 na in provedbom hi-kvadrat testova homogenosti ostvaren op i uvid u odnose izme u podataka.

Metoda normalizacije uvedena je kao primjer esto korištene metodologije prouavanja prostorne distribucije prometnih nesre a (LaScala i dr. 2000, Noland, Quddus 2004, Quddus 2008, Wang i dr. 2009, Wier i dr. 2009). Rije je o podjeli promatranog podru ja na prostorne jedinice i brojanju opažanja unutar pojedine jedinice. Zbroj opažanja unutar jedinice može se podijeliti s ukupnim brojem opažanja ili s površinom jedinice na koju se odnosi. Sli ni se rezultati mogu posti i podjelom promatranog podru ja na pravokutnu mrežu gdje svaka elija predstavlja odre enu prostornu jedinicu.

Ukoliko su dostupni podaci iz drugih izvora, mogu se uklju iti u analizu s ciljem identifikacije prostornih korelacija. Primjerice, podaci iz popisa stanovništva mogu ili bi usporedbu odre ene vrste prometne nesre e s demografskim karakteristikama prostorne

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

Kao što je obra eno u poglavlju *Materijal i metode*, autori Smith, Goodchild i Longley (Smith i dr. 2007) istaknuli su odreene nedostatke te metode. Među njima su naveli da treba uzeti u obzir podatke koji nedostaju te da korištenje neprikladnog djelitelja ili podataka neizvjesne vrijednosti može dovesti do nesuvislih rezultata. Ti su problemi vezani za metode koje provode podjelu prostora na diskretne jedinice i ovise o veličini, obliku i smještaju prostorne jedinice. Pri definiciji ovih parametara treba uzeti u obzir gusto u i raspored opažanja.

Rezultati te 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 širi se na cijelu prostornu jedinicu bez obzira što je riječ o opažanjima na granici jedinice. To onemogućuje identifikaciju istinskih opasnih lokacija. Slika 59 predstavlja primjer tog slučaja. Velik broj nesreća dogodio se na raskrižju koje se nalazi na granici dva odbora, ali sve su pridružene jednom odboru dok je drugi 'prazan'. Prvi se odbor ocjenjuje iznimno opasnim, a drugi kao siguran.

U radu je korištena površina mjesnog odbora kao djelitelj. Također se kao djelitelj mogao koristiti ukupni broj prometnih nesreća, ali to onda drastično mijenja konstrukciju rezultata. Izbor djelitelja 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 prometnika, nije poznato koje su prometnice zaista opasne. To iziskuje puno pregledavanje podataka što je dug i mukotrpni posao podložan pogreškama.

Da bi se izbjegli navedeni nedostaci, zadnjih godina prometne nesreće su ujedno identificirane na žarištu (Anderson 2009, Depaire i dr. 2008, Erdogan i dr. 2008, Prasannakumar i dr. 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 različitih operacija nije pogodna za pravnu unanju, posebno ako se uzme u obzir koliko podataka za obradu. Duljina trajanja izrađena raste primjenom složenijih funkcija. Rezultat izrađena je rasterska datoteka. Velika datoteka ovisi o parametrima definiranim prije izvršavanja operacije. To se odnosi na

veli inu piksela i zadanu širinu podru ja. Ovisno o skupu podataka koji se obra uje, veli ina izlaznih rastera može zauzeti i do 5 GB memorije na ra unalu.

Prednosti te metode su zaista neizmjerne. Ne postoje hipoteze koje se moraju postaviti prije obrade podataka, ve 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 podru ja na prostorne jedinice ili elije. Time se izbjegavaju nedostaci povezani s dimenzijama elija i njihovim rasporedom.

Kao što je navedeno u poglavlju *Materijal i metode*, metoda obra uje to ku po to ku. Samo se analizira prostorna distribucija opažanja, a rezultati najviše ovise o izboru funkcije i definiranoj širini podru ja od strane korisnika. Metoda naglašava samo podru ja s velikim koncentracijama opažanja, odnosno žarišta. Teoretski, žarište je mreža elija položena preko podataka. Ali za razliku od metoda koje prije obrade definiraju mrežu koja pokriva cijelo podru je, ta 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 elija tih mreža popunjavaju se vrijednostima koje najbolje prikazuju rezultate metode.

Analiza rezultata jednostavno je izvediva u GIS-u. Nesre e se pohranjuju u vektorskom sloju kao to ke, a žarišta u rasterskom sloju. Rasterski sloj prikazuje opasne lokacije, a preklapanjem njega s vektorskим slojem vidi se koje nesre e pripadaju odre enim žarištima. Ozna avanjem odre enenog žarišta ozna avaju se nesre e koje mu pripadaju. One se mogu izdvojiti za daljnju analizu. Nad atributnim podacima izdvojenih nesre a mogu se provoditi analize koje omogu uju identifikaciju uzroka nesre a na odre enim lokacijama.

6. Zaklju ci

Cilj rada je uspostava metodologije obrade i analize prometnih nesre a. 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 statisti ke metode koje omogu uju analizu tog skupa podataka. Istaknuti su nedostaci postoje eg sustava i predložena potencijalna rješenja.

Neophodno je na po etku provesti ispitivanje kvalitete dostupnih podataka. To se najviše odnosi na kvalitetu prostornih koordinata prometnih nesre a, ali i na atrIBUTE poput naziva ulica iz zapisnika. S ciljem provo enja kvalitetnih analiza, važno je u i u obradu sa što to nijim podacima.

Naglasak rada je na primjeni prostorne statistike u analizi distribucije prometnih nesre a. Detaljno su obra ene sve korištene statisti ke metode tako da korisnici mogu odabrat 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 nesre a. Ako je potrebno analizirati podatke na ve em podru ju, metoda normalizacije daje zadovoljavaju e rezultate. Daleko najbolje rezultate pruža KDE metoda. U relativno kratkom vremenu mogu se identificirati žarišta vrsta nesre a koja korisnici zahtijevaju. Ne zahtijeva ulazne pretpostavke i hipoteze te ne uklju uje podjelu podru ja od interesa na prostorne jedinice. U potpunosti se oslanja na prostornu distribuciju prometnih nesre a, a korisnici definiraju kvalitetu rezultata. To im je omogu eno kroz izbor funkcija, veli ine piksela i definiranje širine podru ja.

Analize tog tipa neopisivo su važne za društvo. U kona nici je rije o ljudskim životima, pove anju sigurnosti u prometu i smanjenju materijalnih troškova koje donose prometne nesre e. Nadamo se da e ovaj rad poslužiti kao polazna osnova za daljnja istraživanja i pridonijeti poboljšanju sigurnosti u prometu.

Zahvale

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

Nadalje, htjeli bismo se zahvaliti asistentu s Fakulteta prometnih znanosti, mr. sc. Marku Ševrovi u, ija su nam iskustva i preporuke uvelike olakšale obradu podataka i dale uvid u stvarno stanje. Zahvalujemo mu se na pruženoj prilici i podacima koje nam je proslijedio. Nadamo se da će rezultati ovog rada još više zbližiti naše struke i omogu iti donošenje kvalitetnijih zaklju aka po pitanju sigurnosti u prometu. Tako er se zahvalujemo asistentu dr. sc. Ivanu Medvedu, ije je znanje iz domene prostorne statistike uvelike olakšalo analizu danih podataka.

Na kraju bismo se htjeli zahvaliti asistentu Draženu Odobaši u, dipl. ing. geod. te kolegama Franu Glasinovi u, mag. ing. geod. et geoinf., Franu Peru i u, mag. ing. geod. et geoinf., i Mauriciju Vidulinu, univ. bacc. ing. geod. et geoinf., na tehni koj podršci tijekom obrade podataka i moralnoj podršci koju su nam pružili tijekom izrade rada.

Želimo naglasiti da nam izrada ovog rada predstavlja neprocjenjivo iskustvo koje nam je otvorilo brojne mogu nosti i nepovratno proširilo vidike. Sura ivali smo s divnim ljudima, koji su spremni podijeliti svoje znanje i vrijeme s neiskusnim, ali ambicioznim studentima. Nadamo se da smo jednog dana biti u stanju pružiti drugima priliku koja je sada pružena nama te da smo 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, (8.1.2014.)

URL 2: OpenStreetMap Wiki,

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 (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://datasciencerules.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, (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 Šimunović, Filip Todić

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

Prometne nesreće su jedan od svakodnevnih problema suvremenog društva. Identifikacija žarišta prometnih nesreća ključna je pri provođenju u inkovitih strategija smanjenja područja sa visokom gustoćom prometnih nesreća. U radu su obrađene prometne nesreće na području Grada Zagreba za razdoblje od 2010. do 2013. godine. Prometna mreža je preuzeta sa slobodne karte svijeta OpenStreetMap, a podaci iz zapisnika prometnih nesreća obrađeni su u programskom jeziku Python. Posebna pozornost posvećena je provjeri kvalitete dostupnih prostornih podataka te otkrivanju i uklanjanju pogrešaka. Obrađeni su podaci uneseni u sustav za upravljanje objektno-relacijskim bazama podataka PostgreSQL koji uključuje proširenje za geoprostorne podatke i operacije – PostGIS. Nakon provjere tonosti dostupnih podataka testirane su tri najčešće korištene statističke metode za tu vrstu opažanja: hi-kvadrat test, normalizacija i Kernel Density Estimation (KDE). U radu se detaljno obrađuju prednosti i nedostaci svake od metoda i predlaže model geostatističke analize prometnih nesreća. Za područje Grada Zagreba precizno su određena kritika na mesta na kojima je u predmetnom razdoblju najčešće dolazilo do nesreća s teškim posljedicama, s naglaskom na posebno ugrožene sudionike u prometu – pješake i bicikliste. Model će poslužiti kao temelj za daljnja istraživanja i kao primjer Ministarstvu unutarnjih poslova Republike Hrvatske za identifikaciju žarišta prometnih nesreća s ciljem provođenja kvalitetnijih protumjera.

Ključne riječi: žarišta prometnih nesreća, OpenStreetMap, PostgreSQL, geostatistika

Summary

Tomislav Šimunović, Filip Todić

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 (KDE). 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