

Tartu Ülikool
Loodus- ja täppisteaduste valdkond
Ökoloogia ja maateaduste instituut
Geograafia osakond

Magistritöö geoinformaatikas ja kartograafias (maht 30 EAP)

**Riigimaanteedel toimunud liiklusõnnetuste koondumiskohtade
leidmise meetodid**

Anton Štšeglakov

Juhendajad: PhD Evelyn Uuemaa

PhD Raivo Aunap

Kaitsmisele lubatud:

Juhendaja:

Osakonna juhataja:

Tartu 2016

Riigimaanteedel toimunud liiklusõnnetuste koondumiskohtade leidmise

meetodid

Lühikokkuvõte:

Käesolevas töös käsitletakse liiklusõnnetuste koondumiskohtade leidmise meetodikaid ning püütakse nende abil välja selgitada, kas pindadele välja töötatud meetodeid on võimalik kasutada ka joontel toimuvate sündmuste analüüsimiseks. Töö käigus katsetati nelja erinevat meetodit liiklusõnnetuste koondumiskohtade leidmiseks.

Meetoditeks olid kilomeetri pikkuste lõikude analüüs, optimeeritud tulipunktide analüüs, hierarhilise lähima naabri analüüs ja punktihedus kernelis. Töö viimaseks osaks oli parima selgunud meetodi abil koondumiskohtade leidmine Eesti põhimaanteedel, kasutades aastatel 2013 - 2015 toimunud õnnetuste andmestikku.

Töö näitab, et ükski kasutatud meetoditest ei ole täiuslik ning kõigil esineb teatud puudusi. Kõige paremaid tulemusi vaadeldavate hulgast andis punktiheduse meetod, mis mõõdab õnnetustevahelist kaugust mööda teed.

Märksõnad:

Liiklusõnnetus, koondumiskoht, võrgustik, CrimeStat, SANET

CERCS:

P510 Füüsiline geograafia, geomorfoloogia, mullateadus, kartograafia, klimatoloogia

P175 Informaatika, süsteemiteooria

Methods for Identification of Accident Black Spots on National Highways

Abstract:

The aim of this thesis was to test different traffic accident black spot detection techniques and find out if methods, which were in first place developed for analysing events in planar space, would give accurate results for events, which only happen in network space.

Four different methods were used: road segments with length of 1 kilometre, Optimized Hot Spot Analysis, Nearest Neighbor Hierarchical Clustering and Network Constrained Kernel Density Estimation. Traffic accident black spots on main roads of Estonia were detected using the best of previously named methods and data from years 2013 to 2015.

None of the tested methods was perfect but results with network constrained kernel density estimation method were the best.

Keywords:

Traffic accident, black spot, network distance, CrimeStat, SANET

CERCS:

P510 Physical geography, geomorphology, pedology, cartography, climatology

P175 Informatics, systems theory

Sisukord

Sissejuhatus	5
1 Teoreetilised lähtekohad.....	8
1.1 Liiklusõnnetuste koondumiskoha ehk tulipunkti olemus	8
1.2 Varasemad õnnetuste koondumiskohtade uurimused Eestis	10
1.2.1 Maanteeameti tellimusel tehtud uurimused.....	10
1.2.2 Muud uurimused.....	12
1.3. Koondumiskohtade leidmiseks kasutatavad meetodid	14
1.3.1 Optimeeritud tulipunktide analüüs	17
1.3.2 Hierarhiline lähima naabri analüüs.....	24
1.3.3 Punktihedus kernelis	30
2 Andmed ja meetodika.....	38
2.1 Lähteandmed.....	38
2.2 Meetodika	40
2.2.1 Teedevõrgu parandamine	42
2.2.2 Liiklusõnnetuste arv kilomeetri pikkustel teelõikudel.....	42
2.2.3 Optimeeritud tulipunktide analüüs	42
2.2.4 Hierarhiline lähima naabri analüüs.....	43
2.2.5 Punktihedus kernelis	46
3 Tulemused ja arutelu	47
3.1 Liiklusõnnetuste arv kilomeetri pikkustel teelõikudel	47
3.2 Optimeeritud tulipunktide analüüs.....	50
3.3 Hierarhiline lähima naabri analüüs	55
3.4 KDE (<i>Kernel Density Estimation</i>)	61
3.5 Meetodite võrdlus	68
3.6 Järeldused.....	69

3.7	Eesti põhimaanteedel 2013–2015 toimunud õnnetuste koondumiskohad.....	71
3.7.1	Metoodika.....	71
3.7.2	Tulemused	72
	Kokkuvõte	78
	Summary.....	80
	Kasutatud kirjanduse loetelu	83

Sissejuhatus

Liiklusõnnetustes kaotab igal aastal oma elu rohkem kui 1,2 miljonit inimest ning see on üheks peamiseks surmapõhjuseks kogu maailmas (Maailma Terviseorganisatsioon, 2015). Seetõttu on liiklusohutuse temaatika tänapäeval väga aktuaalne kogu maailmas. Suurem osa õnnetustest juhtub arengumaades, kus kiire majanduskasvuga on kaasnenud ulatuslik autostumine ning sellest tulenevalt ka liiklusõnnetuste arv. Lisaks liiklussurmadele ja vigastustele mõjutavad õnnetused ka majandust ja selle arengut. Arengumaad kaotavad nende tõttu ligikaudu 3% SKTst (Maailma Terviseorganisatsioon, 2015). Hukkunute ja vigastatute arvu on aga võimalik vähendada või õnnetusi koguni ära hoida.

Paljud riigid, kaasa arvatud Eesti ja teised Euroopa Liidu liikmed, näevad vaeva, et liiklus oleks ohutu ja turvaline (Euroopa Komisjon, 2011; Majandus- ja Kommunikatsiooniministeerium, 2013). Euroopa Liidu eesmärgiks on viia maanteeliikluses hukkunute arv 2050. aastaks võimalikult lähedale nullile ning 2020. aastaks vähendada liiklusohvrite arvu poole võrra võrreldes aastaga 2010 (Majandus- ja Kommunikatsiooniministeerium, 2013). Liiklusohutuse all mõeldakse seda, et liikluskeskkonnas, milleks on maantee, tänav või muu liikluspind, toimuv ei põhjusta kahju seal viibivate elusolendite elule, tervisele ega varale (Jairus, 2015). Parandades liiklusohutust, väheneb õnnetuste arv ning nende tagajärjed on leebemad või langeb õnnetuste suhe liikluspinda kasutatavusse (Elvik et al., 2009; Jairus, 2015).

Nende eesmärkide saavutamiseks on vaja välja selgitada, kus ja mis põhjusel õnnetused üldse juhtuvad. On täheldatud, et teatud kohtades juhtub neid rohkem. Enamasti on põhjus halvasti projekteeritud maantees või puudulikus liikluskorralduses (Elvik et al., 2009). Sellisel juhul on võimalik õnnetusi nendes kohtades ennetada või nende arvu vähendada, parandades liiklustingimusi. Teelõike, kus toimub keskmisest rohkem õnnetusi, nimetatakse tulipunktideks (inglise keeles *hot spot/hotspot* või *black spot/blackspot*) ning eestipärasemalt liiklusõnnetuste koondumiskohtadeks. Sellel mõistel ei ole ühest konkreetset definitsiooni, kuid Elviku (2007) järgi on see teelõik, kus õnnetuste arv kohalikest riskifaktoritest tingituna on suurem kui teistes sarnaste liiklustingimustega kohtades.

Mitmesuguste nähtuste koondumiskohtade leidmiseks on välja töötatud palju erinevaid meetodeid. Suurem osa neist analüüsib kogu pinda ning sellisel juhul on võimalik vaadata naabruses olevaid alasid ja nähtusi igas suunas. Teatud sündmused saavad aga toimuda ainult kitsamates kanalites ehk joonobjektidel, nagu näiteks teedel, elektriliinidel, gaasitrassidel ja

jõgedel (Okabe, 2015). Väljaspool kanaleid on nende sündmuste toimumise tõenäosus null. Seepärast on selliste nähtuste analüüsimiseks vaja kasutada tööriistu, mis arvestavad tingimust, et sündmus saab toimuda vaid joonobjektidel. Vastasel juhul võivad pinnal töötavad meetodid anda väärade tulemusi.

Kuna seni on suurem osa koondumiskohtade leidmiseks kasutatavad tarkvarad ja meetodid loodud analüüsima kogu pinnal isotroopselt toimuvaid sündmusi, siis selle magistritöö käigus püütakse välja selgitada, kuidas ja mil määral on võimalik kogu pinnal töötavaid meetodeid korrektselt tööle saada ka nende sündmuste jaoks, mis saavad toimuda ainult kitsamates kanalites või nende kõrval. Käesolev magistritöö on jätk autori bakalaureusetööle (Štšeglakov, 2014), kus teede kurvilisuse temaatika kõrval leiti ohtlikud teelõigud teede kilomeetristeks lõikudeks segmenteerimise abil.

Erinevate meetodite katsetamiseks kasutatakse Politsei- ja Piirivalveameti poolt aastatel 2013–2015 riigimaanteedel fikseeritud liiklusõnnetuste andmeid. Meetodeid katsetatakse esmalt väiksematel aladel ning erinevates tingimustes. Parima selgunud meetodi abil leitakse Eesti põhimaanteedel toimunud õnnetuste koondumiskohad eelnevalt mainitud ajaperioodi kohta.

Lähtuvalt töö eesmärgist püstitati järgnevad uurimisküsimused:

- Kuidas on võimalik pindadele välja töötatud meetodeid rakendada joontele?
- Millised on selles töös kasutatud meetodite head ja halvad küljed?
- Kuidas erinevad meetodite tulemused teineteisest?
- Millised on vahemikus 2013–2015 juhtunud liiklusõnnetuste koondumiskohad Eesti põhimaanteedel?

Töös kasutatavad põhimõisted:

Liiklusõnnetuste koondumiskoht - üldjuhul kuni 1 kilomeetri pikkune teelõik, kus kolme järjestikuse aasta jooksul on registreeritud kolm või enam liiklusõnnetust (Stratum, 2013 järgi).

Liiklusohutus - meetmed ja meetodid, vähendamaks liikleja riski hukkuda või saada vigastada (Äär, 2014).

Liiklusõnnetus - juhtum, kus vähemalt ühe sõiduki teel liikumise või teelt väljasõidu tagajärjel saab inimene vigastada või surma. Siia hulka ei loeta juhtumeid, kus jalakäija saab kannatada kokkupõrkes rongiga; juhtumeid võistluste tõttu avalikuks kasutamiseks suletud teelõigul võistlusel osalenud sõidukite vahel; juhtumeid töötamisel tee ehitus-, remondi- või hooldustöödel ainult selles töös osalenud sõidukite ja inimeste vahel (Maanteeamet, 2015).

Hukkunu - inimene, kes suri liiklusõnnetuses saadud vigastuste tagajärjel sündmusekohal või 30 päeva jooksul pärast liiklusõnnetust. Kui inimene suri saadud vigastustesse rohkem kui 30 päeva peale õnnetust, loetakse ta vigastatuks. Suitsiidi sooritanud loetakse hukkunute hulka (Maanteeamet, 2015).

Vigastatu - inimene, kellele liiklusõnnetuses saadud vigastuse tõttu antakse meditsiinilist esmaabi, määratakse ambulatoorne või statsionaarne ravi. Vigastuse olemasolu määrab meditsiinitöötaja. Erinevalt mitmetest Euroopa riikidest ei tehta vahet kergete ja raskete vigastuste vahel. Suitsiidikatse tulemusel vigastatu loetakse vigastatute hulka (Maanteeamet, 2015).

1 Teoreetilised lähtekohad

1.1 Liiklusõnnetuste koondumiskoha ehk tulipunkti olemus

Nagu varasemalt mainitud, on koondumiskoht Elviku (2007) järgi teelõik, kus õnnetuste arv on kohalikest riskifaktoritest tingituna suurem kui teistes sarnaste liiklustingimustega kohtades. Maanteeameti liiklusohutuse spetsialisti Sirje Lilleoru (Lilleorg, 2011) järgi on Eestis toimunud liiklusõnnetuste koondumiskohtadeks kõrgendatud riskiga alad, kus õnnetuse toimumise tõenäosus võib olla suurem või kuhu kontsentreeruvad konfliktid. Selliste teelõikude tuvastamine ja nendega tegelemine on teedeasutuste üks peamisi ülesandeid teedel ohutuse tagamiseks (Lilleorg, 2011).

Kui Elviku (2007) järgi seisneb teelõigu ohtlikkus halvas liikluskorralduses ja halvasti projekteeritud maantees, siis Lilleoru (2011) järgi on lisaks nendele faktoritele ohtlikud ka sellised paigad, kus ei ole arvestatud teiste liikujate vajadustega. Näiteks kohad, kus teel liigub palju jalakäijaid, kuid puuduvad jalgteed ning valgustus on halb. Samuti kohad, kus on raske märgata teisi ristuvatel teel liiklejaid. Samas tasub arvestada ka teadmise, et raskeid liiklusõnnetusi võib juhtuda ka väga hea liikluskorraldusega paikades, seda eelkõige juhi hooletusest, oskamatusel, joobes olekust, sõiduki tehnilisest seisukorrast ja paljudest muudest faktoritest, mis ei ole tingitud liikluskorraldusest (Lilleorg, 2011).

Seetõttu tulekski välja selgitada kõige liiklusohutumad kohad ning seejärel analüüsida, mis on nendel kohtadel ühist ja millest tulenevalt teevad inimesed sõiduki juhtumisel vigu, mis lõppevad õnnetustega. See võimaldab kindlaks teha, millised võiksid olla kõige efektiivsemad liiklusohutust parandavad meetmed nendes kohtades (Bills, 2009). Kõige ohtlikumate ja rohkem tähelepanu vajavate kohtade leidmine võimaldab ressursse ratsionaalsemalt kasutada ja seeläbi paraneb ka liiklusohutus. Vastasel juhul võidakse ressursse kasutada kohtades, kus nende rakendamine ei ole esmatähtis, ning tähelepanu vajavad teelõigud jääksid endiselt ohtlikuks (Mungimit et al., 2009).

Mõistele tulipunkt ei ole konkreetset igal pool kehtivad definitsiooni. Üldjuhul peetakse selleks ala, kus on keskmisest suurem õnnetusjuhtumite arv (Bills, 2009). Andersoni (2009) järgi on liikluses tulipunktiks ala, kus on suur liiklusõnnetuste arv. Samuti on tulipunkte moodustavatele õnnetustele iseloomulik koondumine ning see vihib nende omavahelisele ruumilisele sõltuvusele. Samuti ei ole koondumiskohad suuruse, antud juhul ohtliku lõigu pikkuse, poolest identsed (Anderson, 2009).

Ka eri riigid defineerivad tulipunkti moodustavat õnnetuste arvu ja teelõigu pikkust erinevalt (tabel 1):

Tabel 1. Eri riikide parameetrid, mis defineerivad liiklusõnnetuste koondumiskohta (Mungnimit et al., 2009)

Riik	Teelõigu pikkus	Õnnetuste arv
Austraalia	Võrdlemisi lühike	Vähemalt 3 õnnetust 5 aasta jooksul
Inglismaa	300 meetrit	12 õnnetust 3 aasta jooksul
Saksamaa	300 meetrit	8 õnnetust 3 aasta jooksul
Norra	100 meetrit	4 õnnetust 3 aasta jooksul
Portugal	200 meetrit	5 õnnetust 3 aasta jooksul
Tai	Erinev klassifikatsioon	Vähemalt 3 õnnetust ühel aastal

Lisaks õnnetuste arvule ja teelõigu pikkusele on oluline määratleda, millise ajaperioodi jooksul registreeritud õnnetusi kasutatakse koondumiskohtade leidmiseks (Moreira, 2011). Võttes aluseks Chengi ja Washingtoni (2005) töö, oli May (1964) esimene, kes arutles selle üle, mitme aasta õnnetuste statistikat tuleks koondumiskohtade leidmiseks kasutada. Ta kasutas nende kohtade leidmiseks suurenevat t -väärtust, kus t tähendas aastate arvu ning nende aastate jooksul registreeritud õnnetusi. Ta leidis, et t kasvades erinevused vähenesid ning väga suure ajaperioodi andmestiku kasutamine ei ole mõttekas, kuna sellest saadav kasu on väike (May, 1964 cit. Cheng ja Washington, 2005). Maksimaalse t -väärtusena kasutati selles töös 13 aastat. Töö järeldusena soovitati kasutada koondumiskohtade leidmiseks kolmeaastast perioodi (May, 1964 cit. Cheng ja Washington, 2005). Kolmeaastase ehk võrdlemisi lühikese ajaperioodi kasutamise eelis on see, et selle aja jooksul ei ole väga suuri muutusi keskkonnas, liikluskorralduses, tee pinnakattes ja autode ehituses (Moreira, 2011; Swedish National Road Consulting AB, 2001). Just sellised tegurid ning nende muutlikkus tingivad lühema ajaperioodi kasutamise. Cheng ja Washington (2005) soovivad kasutada 3-6 aastast perioodi, kus 3 aastat on optimaalne aeg ning sellest suurema ajaperioodi kasutamine on õigustatud, kui vaadeldavas piirkonnas (keskkonnas või liikluskorralduses) ei ole toimunud olulisi muutusi. Ka õnnetuste koondumiskohtade käsiraamatu (Swedish National Road Consulting AB, 2001) järgi on kolm aastat mõistlik ajaperiood.

Paraku ei piisa liiklusohutuse parandamiseks ainult koondumiskohtade leidmisest. Kogu protsess on märksa ajanõudlikum ja keerulisem. Tuginedes õnnetuste koondumiskohtade käsiraamatule (Swedish National Road Consulting AB, 2001), tuleb peale ohtlike lõikude leidmist liiklusohutuse parandamiseks teha mitmeid toiminguid:

- hinnata, millised võisid olla liiklusõnnetuse põhjused, liikluskorralduse ja keskkonna puudujäägid;
- leida eelmisele punktile tuginedes meetmeid, et kõrvaldada need puudujäägid;
- analüüsida, milline on eelmises punktis leitud meetmete maksumus ning efekt liiklusohutusele;
- reastada tähtsuse järjekorras valitud meetmed, et leida parim rakendusplaan. Otsuse tegemisel tuleks arvestada meetme maksumust ja kasuteguri suurust;
- rakendada eelmises punktis valitud meetmed;
- viimase ja väga tähtsa toiminguna on vaja jälgida meetmetest tingitud muutusi ning anda hinnang nende efektiivsusele ja tulemustele (Swedish National Road Consulting AB, 2001).

1.2 Varasemad õnnetuste koondumiskohtade uurimused Eestis

1.2.1 Maanteeameti tellimusel tehtud uurimused

Eestis on juba alates 1999. aastast tegeletud liiklusõnnetuste koondumiskohtade otsimisega ning esimeseks sellealaseks uurimuseks oli TTÜ Teedeinstituudis 1999. aastal valminud töö „Maanteede liiklusõnnetuste kontsentratsioonikohtade väljaselgitamine”, kus analüüs tugines 1996-1998. aastate liiklusõnnetuste andmebaasile (Metsvahi, 2002). Sarnase sisuga uurimusi on ilmunud põhimõtteliselt iga kolme aasta tagant, kus analüüsitakse koondumiskohti kolmeaastase perioodi lõikes (Metsvahi, 2002, 2005; AS Teede Tehnokeskus, 2008; Stratum, 2013). Kõigi eelnevalt nimetatud tööde puhul on õnnetuste koondumiskohad jaotatud omakorda kolme rühma:

- liiklusohulik koht – kuni 100 meetri pikkune teelõik, kus kolme järjestikuse aasta jooksul on registreeritud kolm või enam inimkannatanuga liiklusõnnetust;
- eriti liiklusohulik teelõik – üldjuhul kuni 500 meetri pikkune teelõik, kus kolme järjestikuse aasta jooksul on registreeritud kolm või enam inimkannatanuga liiklusõnnetust. Seda teelõiku pikendatakse nii ette- kui ka tahapoole kohani, kust algab 500 meetri pikkune teelõik, kus kolme aasta jooksul ei ole registreeritud ühtegi inimkannatanuga liiklusõnnetust;
- liiklusohulik teelõik – üldjuhul kuni 1 kilomeetri pikkune teelõik, kus kolme järjestikuse aasta jooksul on registreeritud kolm või enam inimkannatanuga liiklusõnnetust, seda teelõiku pikendatakse nii ette- kui ka tahapoole kohani, kust

algab 500 meetri pikkune teelõik, kus kolme aasta jooksul ei ole registreeritud ühtegi inimkannatanuga liiklusõnnetust. (Stratum, 2013)

Võttes aluseks Stratumi (2013) uurimuse, teostati arvutused nn libisevate ühekilomeetriste lõikude kaupa vastavalt eelnevalt nimetatud definitsioonidele. Samuti kasutati kilomeetriseid ja kümne kilomeetri pikkuseid teelõike, millele leiti nendel toimunud õnnetuste arv. Viimasele sarnast metoodikat kasutas käesoleva magistr töö autor oma bakalaureusetöös (Štšeglakov, 2014), kus leiti lisaks kilomeetrisel lõigu kurvilisusele ka sellel toimunud õnnetuste arv. Tõsisemat õnnetuste analüüsi see bakalaureusetöö ei hõlmanud.

Stratumi uurimuses (Stratum, 2013) oli kolm lähenemist:

1. Leiti liiklusohhtlikud kohad, eriti liiklusohhtlikud teelõigud ja liiklusohhtlikud teelõigud lähtuvalt eelnevalt esitatud definitsioonidele. Siin on võetud arvesse vaid inimkahjuga liiklusõnnetusi, kuid nn libisevate teelõikude pikkustega. Antud kriteerium ei võtnud arvesse ei liiklussagedust ega liiklusõnnetuse maksumust.
2. Teostati põhimaanteede lõikude analüüs, mis arvestab ka aasta keskmist ööpäevast liiklussagedust teelõigul.
3. Põhimaanteede lõikude (ühe ja kümne kilomeetri pikkused) analüüs, mis arvestab ka liiklusõnnetuse arvestuslikku maksumust. Viimasel juhul on arvutused teostatud nii ainult inimkahjuga liiklusõnnetusi arvestades kui ka lisades viimastele materiaalse kahjuga registreeritud liiklusõnnetuste maksumuse. (Stratum, 2013)

Tööst selgus, et erinevad meetodid annavad ka üsna erinevad tulemused, eelkõige puudutab see kõige ohhtlikumate teelõikude leidmist, mis tõstatab küsimuse, milline nendest meetoditest on nii-öelda õige. Üheselt on sellele küsimusele keeruline vastata, kuid töö autorid tõid välja mõned olulisemad aspektid, mis mõjutavad arvutuste tulemusi.

Teise lähenemise puhul arvestab meetod lisaks toimunud õnnetustele ka aasta keskmist ööpäevast liiklussagedust. Paljud uuringud on näidanud, et liiklussageduse ja liiklusohutuse vahel on teatav seos ehk mida suurem on liiklussagedus, seda suurem on ka õnnetuse tõenäosus (Stratum, 2013). Samas ei ole see seos lihtsalt ülekantav liiklusõnnetuste prognoosimisele, sest näiteks liiklusõnnetuste arvu ja liiklussageduse ööpäevase jagunemise võrdlus ei kinnita seda, et nende vahel tuleks tugev statistiline seos, kuigi õhtusel tippajal on üldjuhul tegu kokkulangemisega (Stratum, 2013). Kuigi võib olemas olla teatud statistiline seos liiklusõnnetuste arvu ja liiklussageduse vahel, siis on seda usaldusväärsem kasutada

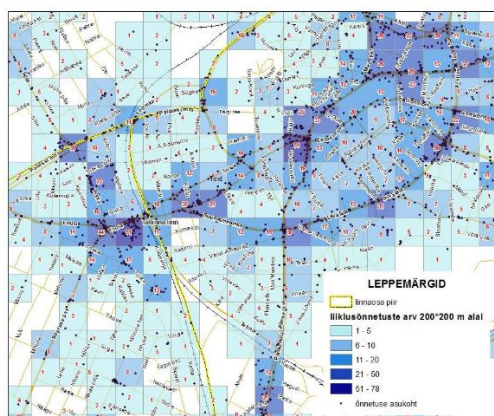
vaid üldistatud tasemel, mitte detailsemas analüüsis nagu seda on kahe teelõigu võrdlus. Teiseks probleemiks selle meetodi puhul on see, et liiklussageduse arvesse võtmine tõstab erilise kaaluga esile väga väikese liikluskoormusega teelõike. Lisaks erinevad ühe ja sama põhimaantee liiklussagedused isegi kuni 24 korda, mis tingib liikluskoormuse kaalu ülehindamise võrreldes registreeritud õnnetustega (Stratum, 2013).

Meetodi puuduseks võib olla ka see, et isegi kilomeetriste lõikude puhul, rääkimata 10-kilomeetristest, võib esineda olukordi, kus ühe ohtliku koha poolt põhjustatud registreeritud liiklusõnnetused jagunevad omakorda kahe erineva täiskilomeetri vahel. Nii jagatakse üks ohtlik lõik, millel toimus kindla ajaperioodi jooksul näiteks kümme õnnetust, kaheks osaks ning neid eraldi vaadates ei pruugi ühe tee-elementi ohtlikkus silma paista (Stratum, 2013).

1.2.2 Muud uurimused

Lisaks nendele uurimustele, mida käsitleti eelnevas alapeatükis, on Eestis tehtud ka muid õnnetuste koondumiskohti hõlmavaid uurimusi või visualiseeringuid. Näiteks on veebis olemas avalikult nähtav veebikaart, mille aluseks on AS Teede Tehnokeskuse poolt valminud uurimuse (AS Teede Tehnokeskus, 2008) tulemused. Kaardil on kujutatud Eesti ohtlikumad teelõigud perioodil 2003 kuni 2007 (Eesti ohtlikumad teelõigud, 2009).

Alphagis on visualiseerinud liiklusõnnetuse koondumiskohti ruutkaardil (Trasberg, 2014; joonis 1). Andmetena kasutati Eesti Liikluskindlustuse Fondi (LKF) poolt kogutud liiklusõnnetuste asukoha koordinaatide infot. Tulipunktide leidmiseks jaotati uuritav ala ruutudeks, mille külje pikkus oli 200 meetrit, ning loeti kokku igasse ruutu jäävate liiklusõnnetuste arv.



Joonis 1. Liiklusõnnetuste koondumiskohtade visualiseerimine (Trasberg, 2014)

Bakalaureuse- ja magistritööna on õnnetuste tulipunkte uurinud Deniss Muraško TTÜ Teedeinstituudist. Bakalaureusetöö kaitses ta 2005. aastal ning selle pealkirjaks on „Aastatel 2001-2003 Eesti maanteedel toimunud liiklusõnnetuste kohtanalüüs“. Magistritöö kaitsi 2008. aastal ning selle pealkirjaks on „Liiklusõnnetuste koondumiskohtade analüüs Eesti põhimaanteedel“. Paraku ei ole need tööd veebipõhiselt kättesaadavad ning seetõttu ei ole võimalik leida infot, millist metoodikat on kasutatud, millised olid nende tööde nõrgad ja tugevad küljed ning millised olid tulemused.

Lisaks Muraško töödele on magistritööna uurinud suurulukitega toimunud õnnetuste koondumiskohti Toomas Eilat (Eilat, 2011). Tulipunktide leidmiseks kasutas ta hierarhilise lähima naabri analüüsi (HLNA), mida võimaldas teha vabavaraline tarkvarapakett CrimeStat III (Smith ja Bruce, 2008). Hierarhiline lähima naabri analüüs on hierarhiline klasterdamise meetod, mis grupeerib ruumiliselt üksteise lähedal asuvad punktid eelnevalt määratletud kriteeriumite alusel (Levine, 2013c). Kriteeriumiteks on otsinguraadius ja minimaalne õnnetuskohtade arv. Esimese suuruseks määras Eilat oma magistritöös kaks kilomeetrit. Minimaalne õnnetuskohtade arv erines ulukiliigist sõltuvalt. Kõigi suurulukite puhul määrati arvuks 10, metskitsel 8, põdral 4 ning metsseal 3 õnnetust (Eilat, 2011).

Kõige värskema töö, mis hõlmab õnnetuste koondumiskohtade leidmist, on Eestis teinud Peet ja Padrik (2015) ning see kannab pealkirja „Eesti riigimaanteedevõrgu loomaõnnetuste registri loomine ning liiklusohlike lõikude selgitamine“. Tulipunktide leidmiseks kasutati ArcGISi enda vahendeid, milleks kõigepealt oli *Kernel Density*, mille abil loodi maanteede kaupa rasterformaadis tiheduspinnad õnnetuste koondumiskohtadele (Peet ja Padrik, 2015). *Kernel Density* ehk *Kernel Density Estimation (KDE)* meetodit kasutatakse ka selles töös ning lähemalt tutvustatakse seda peatükis 1.3.3. Infot otsinguraadiuse ja kerneli suuruse kohta aruandes ei olnud. ArcGIS kasutab *Kernel Density* puhul eukleidilist kaugust.

Tiheduspinnast statistiliselt oluliste koondumiskohtade eristamiseks viidi läbi *Hotspot*-analüüs, kasutades samanimelist tööriista (Peet ja Padrik, 2015). Lisaks tavalistele sisendpunktidele kasutati ka nende kaalu, mis loeti külge eelnevalt saadud KDE tulemustest. Kaalud suurendavad väärtuste varieeruvust. Tööriist arvestab sisendpunktide küljes olevaid varieeruvusega väärtusi ja arvutab seejärel välja, millise olulisustõenäosusega kuulub intsident kas kõrgete või madalate väärtuste ruumilisse klastrisse (Peet ja Padrik, 2015). Automaatselt määratakse ka optimaalne arvutuskaugus (otsinguraadius), mille ulatuses lähedalasuvaid punkte vaadeldakse.

Tulemuseks on igale punktile arvatud Getis-Ord statistik (Z -väärtus), mis sisuliselt on standardhälve. See näitab lokaalse punkt pilve keskmise erinevust üldisest keskmisest. Punkt kuulub õnnetuste koondumiskohta, kui tal on väga kõrge Z -väärtus (Peet ja Padrik, 2015). Kauguse tüübina võimaldab tööriist kasutada eukleidilist või Manhattani kaugust. Infot selle kohta, millist kauguse tüüpi töö tegemisel kasutati, ei olnud. Et tulemused ei oleks ristmikel ja paralleelselt kulgevatel teedel moonutatud, analüüsiti iga maanteed eraldi.

Enne ArcGISi vahendite kasutamist prooviti koondumiskohti leida ka KDE+ (*Kernel Density Estimation* +) töövahendiga, millega leitakse eelkõige loomaõnnetuste koondumiskohti, kuid mis on kasutatav ka kõigi liiklusõnnetuste analüüsimiseks. Peet ja Padrik (2015) leidsid, et programmil on mitmeid kitsaskohti. Tarkvara suudab vajalikud arvutused teha ainult väikeste andmehulkadega ning see eeldaks maanteede jagamist väiksemateks mõne kilomeetristeks lõikudeks. Lisaks tahtsid nad oma töös kasutada kaalusid, kuid KDE+ ei võimalda neid õnnetustele määrata. Tulemused väljastati Exceli tabelina, kuid selle alusel ei olnud võimalik teha järeldusi õnnetuste intensiivsuse kohta maantee igas punktis. Samuti ei saanud leitud tulipunkte lihtsa ja automaatse meetodiga teisendada ruumiandmeteks, mida hiljem esitada kaardil (Peet ja Padrik, 2015). Sellest kõigest tulenevalt hinnati KDE+ tarkvara nende analüüsi jaoks mittesobivaks.

1.3. Koondumiskohtade leidmiseks kasutatavad meetodid

Tulipunktide leidmiseks kasutatakse mitmeid erinevaid meetodeid (Mungnimit et al., 2009). Üldiselt jagunevad need neljaks:

- õnnetuste arvu meetod – leitakse õnnetuste arv konkreetses asukohas, et hinnata selle koha liiklusohutuse taset. Tulipunktiks on sellisel juhul kohad, kus õnnetuste arv on teatud ettemääratud arvust suurem;
- punktiheduse meetod – selleks kasutatakse juhtunud õnnetuste statistikat ning arvutatakse punktihedus kindla pikkusega teelõigu kohta. Tulipunktiks on sellisel juhul kohad, kus õnnetuste arv on teatud ettemääratud arvust suurem;
- õnnetuste määra meetod – liiklusõnnetuste arv jagatuna liiklussagedusega. Tulemuseks saadakse suhe, mis iseloomustab õnnetuste arvu teatud paika läbiva miljoni sõiduki kohta või liiklusõnnetuste arvu ühe miljoni autokilomeetri kohta;
- raskusastme meetod – õnnetustele, kus hukkus või sai vigastada inimesi, antakse suurem kaal. Seejärel summeeritakse kindla asukoha või teelõigu kaalud. Kaalude

määramisel tuleks aluseks võtta hukkunu või vigastatu sotsiaalmajanduslik väärtus (Mungnimit et al., 2009).

Võttes aluseks raskusastme meetodi, on ka Eestis välja arvatud liiklusõnnetuste maksumus, kus hukkunul, vigastatul, invaliidistunud või varakahjuga õnnetusel on oma hind. Õnnetuse maksumus põhineb Tallinna Tehnikaülikooli Logistikainstituudi poolt 2012. aastal valminud uurimustööl „Liiklusõnnetustest ühiskonnale põhjustatud kahjude määramise meetodika täiustamine, kahjudes suuruse hindamine ja prognoosimine“ (Ernits ja Koppel, 2012). 2012. aastal valminud prognoosi järgi on 2016. aastaks riskiväärtust arvestamata hukkunu hind 644 852, invaliidistunu 517 800, vigastatu 12 705 ja varakahjuga õnnetuse hind 9 883 eurot (Ernits ja Koppel, 2012; tabel 2).

Tabel 2. Liiklusõnnetuste kogukahju prognoos aastateks 2012–2016 riskiväärtust arvestamata (Ernits ja Koppel, 2012)

Näitaja	Mõõtühik	2012	2013	2014	2015	2016
Jooksevhinnad						
Hukkunu	eurot hukkunu kohta	582 156	593 178	608 195	624 777	644 852
Invaliidistunu	eurot invaliidistunu kohta	441 120	460 290	479 460	498 630	517 800
Vigastatu	eurot vigastatu kohta	11 387	11 729	12 045	12 371	12 705
Varakahju	eurot õnnetuse kohta	8 858	9 124	9 370	9 623	9 883

Paljudes GIS-tarkvarades on võimalik kasutada mitmesuguseid statistilisi vahendeid punktobjektide omavaheliste seoste leidmiseks. ArcGISi puhul võimaldavad seda teha *Spatial Analysti* või *Spatial Statisticsi* töövahendid. Esimete hulka kuuluvad tihedust arvutavad tööriistad, nagu *Point Density* ja *Kernel Density*. Teine tööriistapakett võimaldab punktid jagada eelnevalt määratletud näitajate abil klastritesse, kasutades *Optimized Hot Spot Analysis* või *Hot Spot Analysis (Getis-Ord Gi*)* töövahendeid.

Lisaks tarkvarasse sisseehitatud moodulitele on võimalik kasutada ka laiendusi, mis suurendavad GIS-tarkvara funktsionaalsust. Sellisteks tarkvarapakettideks on näiteks SANET (*Spatial Analysis Along Networks*) ja CrimeStat, mis ühilduvad ka ArcGISiga, kuna tulemusi on võimalik salvestada otse *shp* formaati. See võimaldab tulemusi visualiseerida ning jätkata nende analüüsi ArcGISi enda vahenditega.

SANET (Okabe et al., 2008) on loodud analüüside tegemiseks spetsiaalselt võrgustikel, milleks võivad olla teed, jõed või muud joonobjektid. Üks töövahenditest selles tarkvarapakettis on *Kernel Density Estimation* (KDE), mille eestikeelne tõlge võiks olla

punktitihedus kernelis ehk tuumikus (Remm et al., 2012) ning mille arvutamine on kohandatud võrgustikule. Mõiste „kohandatud võrgustikule“ tähendab, et kaugus kahe uuritava punkti vahel arvutatakse mööda joonobjekte liikudes. Kernel on kas korrapärase kujuga (enamasti ruut, ring, sektor või sõõrik) või kujutise eelneva segmenteerimise abil saadud suvalise kaju ja suurusega eraldis ehk geograafiline kernel (Merchant, 1984 cit. Remm, 2012).

Kuigi erinevaid töövahendeid punktide ruumianalüüsiks on palju, leiab Anderson (2009), et KDE on nendest kõige paljutootavam ning omab eeliseid teiste meetodite ees (Chainey ja Ratcliffe, 2005; Sabel, 2006 cit. Anderson, 2009). Peamiseks eeliseks on see, et selle abil on võimalik leida ka alasid, kus on kõrgem õnnetuse risk ning õnnetusse sattumise tõenäosus suurem. Eeliseks olemine sõltub aga kasutajast ja püstitatud eesmärgist. Lisaks ohutsoonidele võimaldab KDE kasutajal ise määrata, millise ruumilise üksuse või ühiku kohta analüüsi tehakse. Selline funktsionaalsus tagab võrreldavad tulemused üle kogu uuritava ala (Anderson, 2009). Lisaks Andersonile (2009) on tasapinnalist KDE (*planar KDE*) või võrgustikule kohandatud KDE (*Network constrained KDE*) erivarianti kasutanud liiklusõnnetuste koondumispunktide leidmiseks Xie ja Yan (2008, 2013), Okabe et al. (2008), Tang et al. (2015), Erdogan et al. (2008), Prasannakumar et al. (2011), Yu et al. (2014) jpt.

CrimeStat (Levine, 2013a) võimaldab tulipunktide leidmiseks kasutada mitmeid töövahendeid. Erdogan et al. (2008) sõnul võimaldab see kõigist tarkvaradest kasutada enim erinevaid meetodeid mitmesuguste punktide koondumiskohtade leidmiseks. Kasutada saab näiteks punktobjektide hägust klassifikatsiooni (*Fuzzy mode*), hierarhilist lähima naabri analüüsi (*Nearest Neighbor Hierarchical Clustering*), ette antud rühmade arvuga klasteranalüüsi (*K-means clustering*), Getis-Ord lokaalstatistikut (*Getis-ord Local „G“*) jm. Hierarhilist lähima naabri analüüsi kasutas oma magistritöös ka Toomas Eilat (Eilat, 2011). Getis-Ord statistikut, küll ArcGISi tööriistaga, kasutasid loomaõnnetuste koondumiskohtade leidmiseks ka Peet ja Padrik (2015). Kuigi CrimeStat on algselt välja töötatud kuritegude ja nende ruumiliste seoste uurimiseks, siis on sobilik seda tarkvara kasutada ka teiste juhtumite puhul, mida on võimalik kaardile märkida punktobjektidena. CrimeStati võimalusi on liiklusõnnetuste uurimiseks lisaks Toomas Eilatile (2011) kasutanud ka Moreira (2011), Kundakci (2014) jt.

1.3.1 Optimeeritud tulipunktide analüüs

Optimeeritud tulipunktide analüüsi meetod (*Optimized Hot Spot Analysis*) valiti töösse, et selgitada välja, millised on tulemused, kasutades ArcGISi automaatset vahendit, kus tööriist teeb sisendandmestiku põhjal parameetrite valikud ise. Sarnast tööriista, kuid kus kasutajal endal on võimalik määrata rohkem parameetreid, kasutasid Peet ja Padrik (2015) loomaõnnetuste koondumiskohtade leidmisel. Nemad kasutasid oma uurimuses *Hot Spot Analysis (Getis-Ord G_i^*)* tööriista, mille töötamiseks peavad sisendpunktidel olema ka kaalud, mis tagavad väärtuste suurema variatsiooni.

Kuna käesolevas magistritöös kaalusid ei kasutata, sobis paremini optimeeritud tulipunktide analüüsi töövahend, millel on lisaks olemas ka ümbritseva polügoni (*bounding polygon*) funktsioon. Selle abil on võimalik ära määrata piirkonnad, kus õnnetused üleüldse toimuda saavad. Üks selle töö eesmärkidest on välja selgitada, kas pindadele arendatud töövahendeid on võimalik rakendada nii, et need annaksid kasutajat rahuldavaid tulemusi ka joontel toimuvate intsidentide uurimisel.

Optimeeritud tulipunktide analüüs (ka optimeeritud kuumpunktanalüüs) on ArcGISi integreeritud töövahend, mille abil leitakse statistiliselt olulised klastrid, kus on vaadeldava piirkonna keskmisest kõrgemad ja madalamad väärtused. Tööriist valib automaatselt parameetrid vastavalt sisendandmestikule, et tulemus oleks optimaalne. Tulemuseks on tunnusklass koos Z -, p - ja G_i _bin- väärtuste ning usaldusnivooga. Meetod arvutab Getis-Ord G_i^* statistikut, mis peaks kasutajale väljundina andma statistiliselt usaldusväärseid piirkonnad, kus uuritavate punktide arv on keskmisest suurem. Tööriista kasutatakse juhtumite või objektide puhul, mida on ruumis võimalik märkida punktobjektina. Näiteks võivad juhtumiteks olla kuriteopaigad, liiklusõnnetused, taksotellimused ning objektideks teatud liiki hooned (ArcGIS, 2015c).

Kuna käesolevas töös kasutatakse sisendandmestikuna punkte, siis selgitatakse toimimise põhimõtet nendest lähtudes. Antud töövahendis on võimalik, aga ei ole kohustuslik, sisendina kasutada ka punktide kaalusid (*analysis field*), kuid antud töö keskendub eelkõige õnnetuste ruumilisele paiknemisele ning seetõttu on iga õnnetuse kaaluks 1.

Tööpõhimõtet on selgitatud järgnevates peatükkides. Kokkuvõtvalt on etapid ArcGISi tulemuste akna sõnumite põhjal sellised: kontrollitakse, kas analüüsiks on piisavalt sisendpunkte. Leitakse ning eemaldatakse punktid, mis on väga kaugel teistest punktidest ning võivad seetõttu olla mõõtmisvead (*locational outlier*). Leitakse kalavõrgu (*fishnet*)

polügonide suurus ning tekitatakse selle suurusega polügonidest võrk uuritavale alale. Summeeritakse igasse võrguelementi jäävate õnnetuste arv ning hinnatakse saadud tulemust, tagastades kasutajale keskmine, miinimum- ja maksimumväärtus, mis saadi õnnetuste summeerimisel igas võrguelemendis. Määratakse analüüsi skaala ehk otsinguraadius, arvutatakse Getis-Ord G_i^* statistik ning tulemused korrigeeritakse valeavastusmäära meetodi alusel.

1.3.1.1 Valikud

Esmalt valitakse sisendandmestik, mis selle töö puhul on liiklusõnnetuste punktid. Järgmisena määratakse algoritm, kus ja mille järgi punktid loendatakse. Valikuid on kolm (ArcGIS, 2015a):

1. punktide loendamine tekitatud kalavõrgu polügonides;
2. punktide loendamine sisendina antud polügonides;
3. lähestikku asuvate punktide liitmine üksteise külge, et tekitada kaalutud punktid.

Kasutades esimest meetodit, on võimalik lisafunktsioonina ära määratleda alad, kus õnnetused üldse toimuda saavad, kasutades nn ümbritsevaid polügone (*bounding polygon*). Teedevõrgu puhul on selline määratlemine otstarbekas, sest liiklusõnnetused saavad toimuda vaid teedel. Kui seda lisafunktsiooni ei kasutata, siis ei tea töövahend, kas mõni piirkond ilma õnnetuseta peaks olema väärtusega 0, kuid õnnetuse tõenäosus on seal küll olemas, aga seda ei toimunud. Või ei ole teatud piirkondi üldse mõtet uurimusse kaasata, kuna intsidentide seal toimuda ei saagi. Teise kahe valiku puhul, mis on nimetatud selle alapeatüki alguses, ei ole võimalik kasutada ümbritseva polügoni funktsiooni ning seetõttu ei oma need selle uurimuse seisukohalt tähtsust.

1.3.1.2 Kalavõrgu polügoni suuruse leidmine

Et leida kalavõrgu polügoni suurus, arvutatakse esmalt välja sisendpunktide keskmine ja mediaankaugus lähima naabrini. Keskmise lähima naabri kauguse (*average nearest neighbor distance* ehk ANN) puhul liidetakse kõigi punktide kaugused lähima naabrini ning jagatakse see sisendandmestikus olevate punktide arvuga (ArcGIS, 2015a). Mediaan lähima naabri kauguse (*median nearest neighbor distance* ehk MNN) puhul reastatakse lähima naabri kaugused kasvamise järjekorras ning valitakse mediaanväärtus. Seejärel valitakse kalavõrgu polügonide esialgne suurus (*cell size* ehk CS) ANN või MNN järgi sõltuvalt sellest, kumb väärtus oli suurem (ArcGIS, 2015a). Kui näiteks NNA väärtus oli 380 ja MNN 400, valitakse suuruseks viimane. Järgmisena kohandatakse seda valemiga (valem 1).

$$CS_{kohandatud} = CS_{esialgne} \times Scalar \quad (\text{Valem 1})$$

kus:

väiksem = MIN(ANN, MNN);

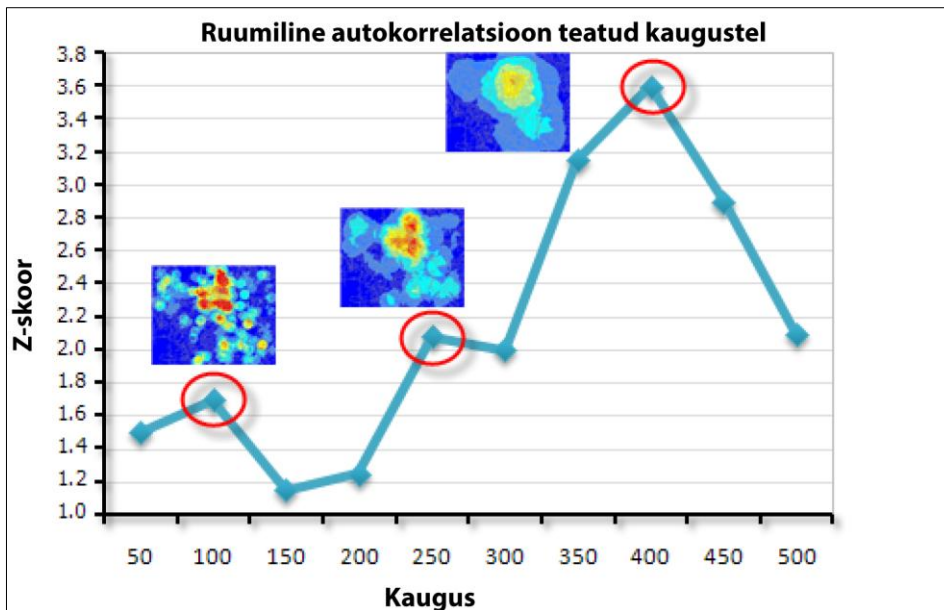
suurem = MAX(ANN, MNN);

scalar = MAX((Suurem/Väiksem), 2).

Saadud väärtust kasutades tekitatakse sama küljepikkusega ruudukujulised kalavõrgu polügonid, asetatakse need võrgustikuna uuritavale alale ning loetakse kokku, mitu punkti tekitatud pindade sisse jääb (ArcGIS, 2015a).

1.3.1.3 Uuringuskaala

Kui punktid on loetud, leitakse analüüsiks sobiv skaala ehk otsinguraadius (*Scale of Analysis*). Näiteks haiguspuhangu puhul, kus viiruse kandjaks on moskiito ning on teada, et tema lennukaugus on 10 kilomeetrit, siis oleks sobilik kasutada sedasama lennukaugust (ArcGIS, 2015a). Liiklusõnnetuste puhul on raske iseseisvalt otsustada, milline võiks olla optimaalne uurimisskaala, mis annab piisavalt täpsed ja objektiivsed tulemused. Esimene strateegia, mida meetod kasutab mõistliku skaala leidmiseks, on astmeline ruumiline autokorrelatsioon (*Incremental Spatial Autocorrelation*), mis kasutab globaalset Morani I statistikut (*Global Moran's I statistic*) teatud suurenevatele kaugustele ning mõõdab ruumilise klasterdumise intensiivsust nendel kaugustel. Klasterdumise intensiivsus tehakse kindlaks Z-väärtuse järgi, mille tagastab globaalne Morani I statistik (ArcGIS, 2015a). Üldjuhul kauguse kasvamisel suureneb ka Z-väärtus, mis näitab klasterdumise intensiivistumist, kuid teatud kaugusel saavutab z-väärtus oma suurima väärtuse (joonis 2). Sellised tipud kajastavad kaugusi, kus ruumiline klasterdumine on veel kindlasti olemas, kuid kauguse edasisel suurenemisel ei pruugi klasterdumist enam olla. Neidsamu kaugusi otsib ka tulipunktide analüüsi töövahend, kasutades astmelist ruumilist autokorrelatsiooni. Kui selline kaugus leitakse, muutub see analüüsi skaalaks (ArcGIS, 2015a).



Joonis 2. Astmeline ruumiline autokorrelatsioon (ArcGIS, 2016)

Kui selliseid kaugusi ei leita, vaadatakse punktide ruumilist paigutust ning arvutatakse keskmine kaugus, mis annaks igale punktile K naabrit (ArcGIS, 2015a). K arvutatakse kui $0,05 \cdot N$, kus N on sisendpunktide arv. Näiteks 100 punkti puhul oleks $K = 100 \cdot 0,05$ ehk 5. K kohandatakse selliselt, et see ei oleks väiksem kui 3 ja suurem kui 30. Kui keskmine kaugus, mis annaks K naabrit, on suurem kui standardkaugus (*standard distance*), siis määratakse analüüsi skaalaks üks standardkaugus. Vastasel juhul on otsinguraadiuseks sama kaugus, mis annab igale sisendpunktile K naabrit (ArcGIS, 2015a).

1.3.1.4 Getis-Ord G_i^* statistik ja selle arvutamine

Nüüdseks on kõik ettevalmistused tehtud ning vajalikud parameetrid valitud. Järgmise sammuna arvutatakse Getis-Ord G_i^* statistik (valem 2),

$$G_i^* = \frac{\sum_{j=1}^n w_{i,j} x_j - \bar{X} \sum_{j=1}^n w_{i,j}}{S \sqrt{\frac{n \sum_{j=1}^n w_{i,j}^2 - \left(\sum_{j=1}^n w_{i,j} \right)^2}{n-1}}} \quad (\text{Valem 2})$$

kus x_j on väärtus kohas j , w_{ij} on ruumiline kaal vaatluse j ja i vahel ning n on sisendpunktide arv (ArcGIS, 2015b; Remm et al., 2012). Kui koht j on etteantud kauguse sees, siis $w_{ij} = 1$, muul juhul $w_{ij} = 0$ (Remm et al., 2012).

\bar{X} arvutatakse järgnevalt (valem 3):

$$\bar{X} = \frac{\sum_{j=1}^n x_j}{n} \quad (\text{Valem 3})$$

S arvutatakse järgnevalt (valem 4):

$$S = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - (\bar{X})^2} \quad (\text{Valem 4})$$

G_i^* statistik on Z -väärtus ning seetõttu ei ole edasised arvutused vajalikud (ArcGIS, 2015b).

Tulemuseks on Z - ja p -väärtused, mis näitavad, kas teatud piirkonnas esineb punktide ruumilist klasterdumist. Statistiku arvutamiseks vaadatakse kalavõrgu polügoni koos tema naabritega. Suure väärtusega polügon võib olla huvitav, kuid see ei pruugi veel olla statistiliselt usaldusväärse tulipunkti osa. Selleks, et olla statistiliselt usaldusväärne, peavad läheduses olema ka teised suure väärtusega polügonid (ArcGIS, 2015a). Z -väärtus on sisuliselt standardhälve, mis näitab lokaalse punktiple keskmise erinevust üldisest keskmisest (Peet ja Padrik, 2015).

1.3.1.5 Valeavastusmäär ja tulemuste korrigeerimine

Selleks, et saada usaldusväärsed tulemused, tuleb neid korrigeerida, kasutades valeavastusmäära (*FDR – False Discovery Rate Correction*). See on statistiline meetod, mis seab ülempiiri valepositiivsete testitulemuste osakaalule (Koppel, 2015). ArcGISi puhul vähendab see meetod kriitilise p -väärtuse künnist, et leida valepositiivsed tulemused.

Valepositiivsete nähtuste arvu hinnang leitakse järgmiselt. Näiteks, kasutades 95% usaldusnivood, ütleb tõenäosusteooria, et viie võimaluse puhul sajab paiknevad punktmustrid struktureeritult (klasterdunud või hajutatud) ning on statistiliselt olulise p -väärtusega, kuid tegelikkuses paiknevad need punktid juhuslikult (ArcGIS, 2015d). Viis

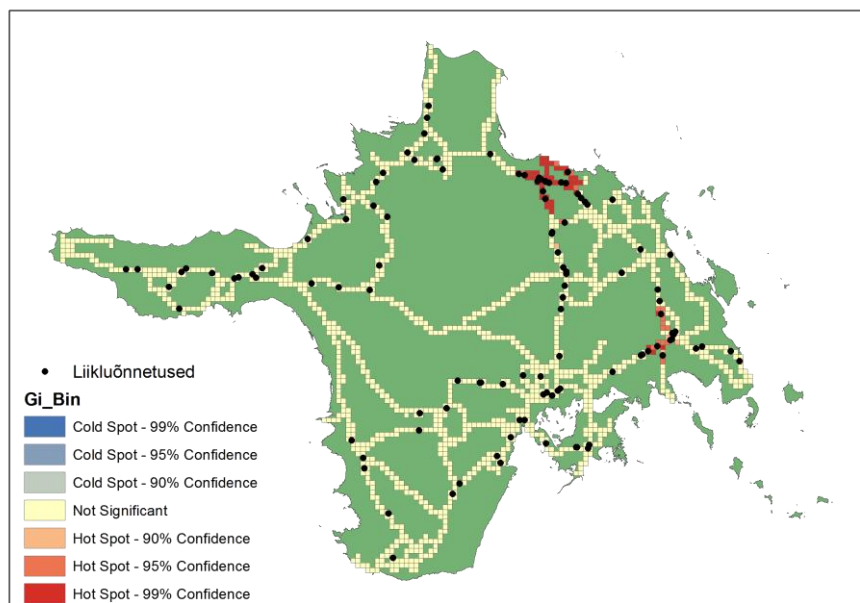
võimalust sajast tundub esmapilgul tühine hulk, kuid tuleb arvestada asjaolu, et statistikud arvutatakse igale punktile andmestikus. Kui selles on 10 000 punkti, siis on tõenäoline, et leitakse ka 500 valepositiivset tulemust.

FDR korrigeerib tulemusi, arvestades ka punktide ruumilist sõltuvust (*Spatial Dependency*) autokorrelatsiooni tõttu – üksteisele lähedal paiknevad kohad kipuvad olema sarnased ainuüksi selle pärast, et nad paiknevad lähestikku (Remm et al., 2012). Sellest hoolimata eeldavad paljud statistilised testid, et vaatlusalused punktid või nähtused oleksid sõltumatud (ArcGIS, 2015d). Autokorrelatsioon võib tehislikult suurendada statistilist usaldusväarsust. Ruumiline sõltuvus mõjutab tulemusi, kuna iga punkti väärtus arvutatakse arvestades naabrite väärtusi ning punktid, mis paiknevad üksteisele lähedal, omavad paljusid samu naabreid.

Selleks, et eemaldada valepositiivsed tulemused ja vabaneda autokorrelatsiooni mõjust, reastab FDR statistiliselt olulised p -väärtused kasvamise järjekorras ning põhinedes valepositiivsete väärtuste arvu hinnangule, eemaldatakse kõige väiksema p -väärtustega nähtused (ArcGIS, 2015d). Allesjäänud väärtused eristatakse G_i _Bin atribuudiga, kus selle väärtused ± 3 , ± 2 ja ± 1 näitavad statistilist olulisust vastavalt 99%-se, 95%-se ja 90%-se usaldusnivooga (ArcGIS, 2015c).

1.3.1.6 Väljund

Tööriista lõpptulemuseks on andmekiht, kus on tekitatud kalavõrgu polügonid koos atribuutidega, milleks on Z - ja p -väärtused ning G_i _Bin väärtus. Vastavalt G_i _bin atribuudile on polügonid juba erinevalt värvitud (joonis 3).



Joonis 3. Optimeeritud tulipunktide analüüsi tulemus Hiiumaa näitel

1.3.1.7 Getis-Ord varasemates uurimustes

Peet ja Padrik (2015) leidsid oma töös loomaõnnetuste koondumiskohad kahte meetodit kasutades. Esmalt leiti ArcGIS *Kernel Density* töövahendiga tiheduspind, tekitati 100-meetrised lõigud ning lõikude tsentroididesse punktid, millele loeti külge punkti alla jääva punktutiheduse piksli väärtus. Seejärel kasutati neid punkte sisendina *Hotspot*-analüüsi (*Getis-Ord Gi**) töövahendis. Tulemuseks olid statistiliselt olulised punktid. Sellise meetodiga on võimalik leida *Kernel Density* tulemuse statistiliselt olulised punktid.

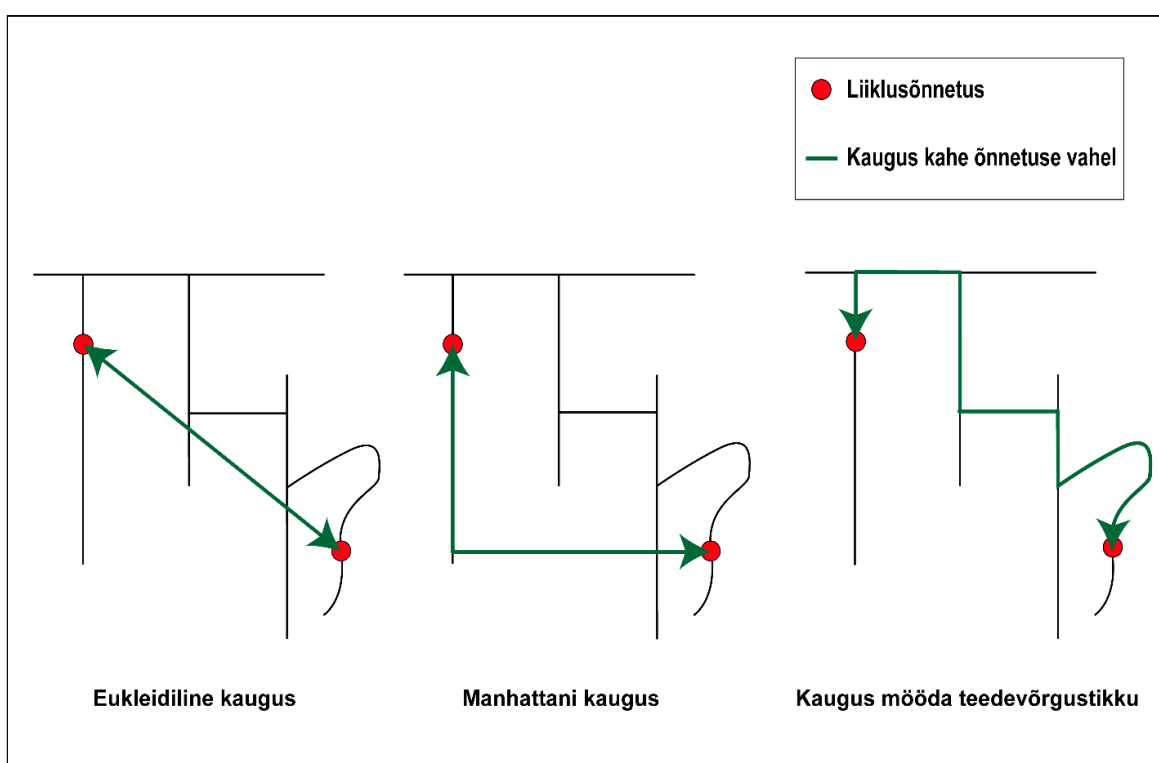
Nie et al. (2015) kasutasid sarnast lähenemist nagu Peet ja Padrik (2015). Nad arvutasid esmalt võrgustikule kohandatud KDE väärtused ning kasutasid neid sisendina Getis-Ord statistiku arvutamiseks. Kusjuures, statistiku arvutamine oli samuti kohandatud võrgustikule. See tähendab, et kaugust naabruses olevate teiste punktideni hinnatakse kui kaugust mööda teed. Nad tõdevad, et KDE ilma statistilise olulisuse kontrollita on mõeldud pigem visualiseerimiseks. Kombineerides seda Getis-Ord statistikuga, on võimalik leida just statistiliselt olulised liiklusõnnetuste koondumiskohad. Erdogan et al. (2015) leidsid oma töös, et tasandiline KDE meetod koos Getis-Ord statistikuga leiab edukalt üles koondumiskohad, mis asuvad eelkõige ristmikel.

1.3.2 Hierarhiline lähima naabri analüüs

Meetodi kirjeldamiseks võetakse aluseks vabavaraline tarkvara CrimeStat IV. Hierarhiline lähima naabri analüüs (*Nearest Neighbor Hierarchical Clustering*) ehk HLNA on hierarhiline klasterdamise meetod, mis grupeerib ruumiliselt üksteise lähedal asuvad punktid vastavalt kasutaja poolt määratletud kriteeriumite alusel. Punkte grupeeritakse niikaua, kuni kõik punktid on grupeeritud ühte klastrisse ehk kobarasse või kuni klasterdamise kriteerium ebaõnnestub (Levine, 2013c).

HLNA kasutamiseks on vaja määratleda kaks kriteeriumi. Esiteks punktide otsimise raadius ning minimaalne punktide arv (Eilat, 2011; Levine, 2013c). Raadiuse abil määratakse ära, kui kaugel võivad punktid üksteisest maksimaalselt ruumiliselt paikneda, et need grupeeritaks ühte klastrisse. Intsidentide arvu abil pannakse paika, kui palju punkte on minimaalselt vaja, et neist moodustataks klaster (Eilat, 2011; Levine, 2013c).

Kaugusi ehk vahemaid ühest punktist teiseni on CrimeStat'is võimalik arvutada kolme erineva meetodiga ehk kasutades eukleidilist kaugust, Manhattani kaugust või kaugust mööda teedevõrgustikku (joonis 4).



Joonis 4. Erinevate kauguste olemus (modifitseeritud autori poolt Wikibooks, 2016 järgi)

Eukleidilise kauguse puhul arvutatakse kahe punkti vaheline vahemaa kui lühim kaugus nende punktide vahel. Projitseeritud koordinaatide puhul on selleks sirgjoon kahe punkti vahel (Levine, 2013a). Manhattani kaugus on liikumiste summa, mis on vaja teha, et jõuda lühimat teed pidi mööda horisontaalset ja vertikaalset telge liikudes ühest punktist teiseni (Levine, 2013a).

Võrgustikukauguse puhul on tegu kahe punkti vahelise lühima vahemaaga mööda teed liikudes (Levine, 2013a). Sellise kauguse tüüpi kasutamiseks on CrimeStatil vaja ka sisendkihti, milles oleks olemas teedevõrgustik, mis on kauguse arvutamise aluseks. Teedevõrgustikke on kahte tüüpi – ühesuunalised, kus teel saab liikuda ainult ühes konkreetses suunas, ja kahesuunalised, kus liikumine on võimalik mõlemas suunas. Programm toetab *shp* formaadis faile, mis sisaldavad joonobjekte. Teede sõlmpunktid ehk ristmikud leiab CrimeStat iseseisvalt üles. Kasutades kahesuunalist teedevõrku, on võimalik määrata ära ka ühesuunalised teelõigud, kasutades kahte atribuutvälja (Levine, 2013a). See on väli, kus teelõigu lõpus olevatele sõlmpunktidele määratakse väärtused 1 või 0. Väärtus 0 tähendab, et sellest sõlmpunktist on võimalik liikuda järgmisele selle sõlmpunkti küljes olevale teelõigule ning samas on võimalik ka liikumine mööda sama teelõiku tagasi. 1 tähendab, et lõigu lõpusõlmpunktist ei ole võimalik liikuda mööda sama teelõiku tagasi. Jättes need atribuutväljad tühjaks, eeldab programm, et teelõik on kahesuunaline (Levine, 2013a).

Võrgustikukauguse kasutamisel on kaks peamist puudust: tihtipeale on digitaliseeritud teedevõrgustikes palju vigu. Andmed võivad olla vananenud ehk ei sisalda infot uute teede või ümberehituste kohta, kuigi inimesed kasutavad neid puuduolevaid teid (Levine, 2013b). Esineb ka topoloogiavigu ehk teed on üksteise peale digitaliseeritud või esineb rippuvaid otsasid. Rippuvad otsad kujutavad endast olukorda, kus teed ei ole ettenähtud kohtades üksteise külge snäpitud (vt peatükk 2.2.1). Teiseks võivad arvutused teedevõrgul võtta väga kaua aega. Algoritm peab leidma mitmest võimalikust teekonnast ühe ja lühima ning igaks alternatiivse teekonna arvutuseks kulub aega (Levine, 2013b). Suurendades uuritavate punktide arvu, kasvab arvutusteks kuluv aeg eksponentsiaalselt. Lisaks sõltub kiirus ka sellest, kui komplitseeritud on kasutatav teedevõrk. Seetõttu soovitavad CrimeStat tarkvara loojad võimalusel kasutada eukleidilist või Manhattani kaugust, sest võrgustikukaugust kasutades võib saadav lisatäpsus olla väike. Arvutusteks kuluv ooteaeg võib aga kujuneda väga pikaks (Levine, 2013b).

1.3.2.1 Otsinguraadiuse määramine

Kasutaja saab ise määrata, kui suur on otsimise raadius ning millistes ühikutes (miil, meeter, kilomeeter). Selle eeliseks on see, et määrata saab konkreetse numbrilise suuruse ja see võimaldab hiljem võrrelda erinevate alade tulemusi. Puuduseks on aga see, et raadiuse valimine on subjektiivne ning kasutaja tihtipeale ei tea või ei ole kindel, milline suurus valida, et tulemus vastaks tema uurimisküsimusele. Mida suurem valitud raadius, seda suurem on tõenäosus, et punktid grupeeritakse juhuslikult. Statistiliselt usaldusväärset selliselt grupeeritud rühmad ei ole (Levine, 2013c).

Võimalik on kasutada ka juhuslikku lähima naabri kaugust (*random nearest neighbor distance*), kus tarkvara valib vastavalt sisendpunktidele ja kasutaja poolt määratud uurimisala pindala järgi ise otsinguraadiuse (Levine, 2013c). Kasutaja saab määrata ka olulisuse nivoo eeldatavale lähima naabri kaugusele.

Keskmine juhuslik lähima naabri kaugus arvutatakse järgmiselt (Levine, 2013c; valem 5):

$$d_{NN(ran)} = 0.5 \sqrt{\frac{A}{N}} \quad (\text{Valem 5})$$

kus:

A - vaadeldava ala pindala;

N - uuritavate intsidentide arv.

Standardviga leitakse järgmiselt (valem 6):

$$SE_{d(ran)} \cong \sqrt{\frac{(4 - \pi)A}{4\pi N^2}} = \frac{0.26136}{\sqrt{\frac{N^2}{A}}} \quad (\text{Valem 6})$$

kus:

A - vaadeldava ala pindala;

N - uuritavate intsidentide arv.

Kauguse usaldusvahemik on määratletud järgmiselt (valem 7):

$$Usaldusvahemik = d_{nn(ran)} \pm t * SE_{d(ran)} \quad (\text{Valem 7})$$

kus t on t -väärtus t -jaotusest, mis vastab olulisuse niivoole ja vabadusastmete arvule, milleks on 120. Alumine usalduspiir on arvutatav, kui lahutada keskmisest juhuslikust kaugusest standardviga ning ülemine usalduspiir on arvutatav, kui keskmisele juhuslikule kaugusele liita standardviga (Levine, 2013c).

Leitud usaldusvahemik määrab ära tõenäosuse kahe punkti vahelise kauguse jaoks. Näiteks kasutades ühepoolset tõenäosust p , vähem kui $p\%$ intsidentidest omaksid lähima naabri kaugust, mis oleks väiksem kui see valitud piir (Levine, 2013c). Seda aga juhul, kui punktide jaotus on ruumiliselt täiesti juhuslik. Kui uuritavad punktid paiknevad ruumiliselt juhuslikult ning otsinguraadiusena kasutatakse keskmist juhuslikku kaugust ja olulisuse nivoo liuguri asend on vaikimisi, siis ligikaudu 50% punktipaaridest on üksteisele lähemal kui see keskmine juhuslik kaugus. Valides tõenäosuseks 0,05, oleks ainult 5% paaridest üksteisele lähemal kui keskmine juhuslik kaugus. Liuguril on 12 positsiooni, millega on võimalik määrata tõenäosus, et kaks juhuslikult valitud punkti on üksteisele lähemal kui keskmine juhuslik kaugus (Levine, 2013c).

Otsinguraadiusena saab juhuslikku lähima naabri kaugust kasutada vaid juhul, kui uuritava ala pindala on tarkvaraaknas defineeritud. Pindala peab olema õigesti määratud, sest sellest sõltub, milline juhuslik kaugus valitakse. Kui kasutada vale pindala, siis on tulemused väärad, mis võib viia valede järelduste tegemiseni. Kui kasutaja ise ei määra pindala, kuid kasutab otsinguraadiusena juhuslikku lähima naabri kaugust, siis leiab tarkvara pindala ise, kasutades sisendpunktide miinimum ja maksimum koordinaate (X, Y). Nende alusel leitakse moodustunud ristküliku pindala, mis enamasti on pindalalt suurem kui tegelik uuritav ala (Levine, 2013c).

1.3.2.2 Minimaalse punktide arvu määramine

Minimaalne intsidentide arv tuleb määrata selleks, et vähendada väga väikese punktide arvuga kobaraid (Levine, 2013c). Mahuka andmestiku puhul võidakse leida sadu, kui mitte tuhandeid paari punkti sisaldavaid klastreid, mis vastavad otsinguraadiuse kriteeriumile. Väikeste gruppide arvu ja juhuslike klastrite leidmise tõenäosuse vähendamiseks saab kasutaja määrata punktide miinimumarvu, mida on vaja grupi moodustamiseks (Levine, 2013c).

1.3.2.3 Esimest järku klastrid ja nende visualiseerimine

Kahe eelnevalt kirjeldatud kriteeriumi, otsinguraadiuse ning minimaalse punktide arvu, alusel leiab programm esimest järku klastrid. Kuna selle töö eesmärk on leida liiklusõnnetuste koondumiskohad, siis ei ole teist ja kõrgemat järsku kobarad vajalikud, kuigi mõne teise temaatikaga analüüsid võivad ka kõrgemat järku klastrid osutada kasulikeks. CrimeStat võimaldab leitud rühmasid visualiseerida kahel moel – ellipsi väljund ja kumer kate (*convex hull*) väljund (Levine, 2013c). Kumer kate, mis tekitab klastrid moodustavate punktide ümber polügoni, vastab kujult täpselt sellele klastrile. Ellips on aga üldistus ning võib kujutusviisina olla liialdav ja viia valede järelduste tegemiseni (Levine, 2013c).

1.3.2.4 Statistilise olulisuse simulatsioon

HLNA leitud klastrite statistilise olulisuse testimine on keeruline. Kasutades juhuslikku lähima naabri kaugust otsinguraadiusena ning määrates tõenäosuse, et kaks punkti grupeeritakse juhuslikult, teab kasutaja, kui suur on tõenäosus, et leitud klastrid on juhuslikult grupeeritud ning seega statistiliselt mitteolulised (Levine, 2013c). Probleem on aga selles, et HLNA ei grupeeri vaid punktipaare, vaid kõiki punkte, mis jäävad määratud otsinguraadiuse sisse, kuna määratud on ka minimaalse punktide arvu kriteerium, alates millest loetakse punktide kogumit klastriks. Sellisel juhul ei ole punktide tõenäosusjaotus teada. Sellest tulenevalt on vajalik kasutada Monte Carlo simulatsiooni (Levine, 2013c).

Monte Carlo simulatsiooni tulemuseks on usaldusvahemikud HLNA poolt leitud klastritele. Selle meetodi põhimõte seisneb selles, et simulatsioon paigutab juhuslikult N arvu punkte ristkülikukujulisele alale, mille pindala on sama suur, kui alal, kus eelnevalt klastrid leiti. Seejärel leitakse neid punkte ning juba eelnevalt määratud kriteeriume (otsinguraadius, minimaalne punktide arv) kasutades uued grupid. Algoritm kordab sama simulatsioon K korda, kus arvu K määrab kasutaja (Levine, 2013c).

Simulatsiooni mitmeid kordi korrates on kasutajal võimalik hinnata, millised on ligilähedased usutavad vahemikud kasutatud kriteeriumite ja andmestiku jaoks. Tulemusena saadakse 5 veergu ning 12 rida. Veergudeks on protsentiil ning sellele leitud klastrite arv, pindala, punktide arv ja punktihedus (punktide arv jagatud nendest punktidest moodustatud ellipsi pindalaga). Ridadeks on simulatsioonis leitud minimaalne ja maksimaalne klastrite arv ning nende arvu protsentiilid (0,5., 1., 2,5., 5., 10., 90., 95., 97,5., 99. ja 99,5. protsentiil). Protsentiilid arvutatakse järgmiselt: iga simulatsiooni käigus leitakse klastrite arv ning sorteeritakse nad kasvavas järjekorras. Lõpptulemusena on kõigi K arvu simulatsioonide

käigus leitud klastrite arvud sorteeritud kasvavas järjekorras ühe pingereana ning seejärel määratakse selle pingerea alusel protsentiilid. Seega miinimumiks on kõigi simulatsioonide käigus leitud kõige väiksem arv klastreid, maksimumiks suurim arv klastreid ning protsentiilideks neile vastavad gruppide arvud (Levine, 2013c).

Simulatsiooni tulemust hinnates on vaja valida kindel protsentiil ning enamjaolt on selleks 95. protsentiil. See tähendab, et I liiki vea tegemiseks on maksimaalne tõenäosus 5%, kuna kasutaja tahab leida rohkem klastreid võrreldes selle arvuga, mis on leitud simulatsiooni käigus ehk juhuslikult (Levine, 2013c). CrimeStati juhendis käsitletud nädisandmestikul (1181 punkti, $p \leq 0,5$ punktide juhuslikuks grupeerimiseks, minimaalne punktide arv 5) leidis HLNA 69 gruppi ning simulatsiooni tulemusena leiti minimaalselt üks klaster, maksimaalselt 7 klastrit ning 95. protsentiil oli 3. Simulatsiooni tulemus näitab, et juhuise tõttu on grupeeritud väga väikene arv punktidest. 95. protsentiili valides eeldab kasutaja, et maksimaalselt 3 gruppi 69st võiksid olla juhuslikult leitud (Levine, 2013c). Paraku ei ole võimalik kindlaks teha, millised need kolm rühma on. CrimeStat arvutab väljundisse ka iga klasteri punktiheduse ning võib oletada, et juhuslikud on väikseima punktihedusega grupid.

1.3.2.5 Hierarhiline lähima naabri analüüs varasemates uurimustes

Moreira (2011) kasutas HLNA puhul oma töös eukleidilist kaugust, mille tulemusena osa leitud koondumiskohtadest sisaldasid ka paralleelsetel teedel toimunud õnnetusi. See tuleneb kasutatud kauguse tüübist. Sellised tulemused aga võivad viia selleni, et mõned ohtlikud lõigud tuvastatakse valesi, kuna eri teedel võivad liiklus- ja keskkonnatingimused olla erinevad ning ühe grupina selliseid kohti analüüsida ei tohiks (Moreira, 2011). Meetodi jaoks peab iga järgmine õnnetus asuma eelmisest lähemal kui määratud otsinguraadius. See võib aga viia selleni, et koondumiskohana leitakse ulatuslik õnnetusteahel, kuna iga järgmine õnnetus on eelmisele lähemal kui määratud otsinguraadiuse kriteerium. Tulemuseks on ohtlik koht väga pikal lõigul ning ühe klasteri sisse võivad kuuluda mitmete paralleelsete teede ja ristmike õnnetused (Moreira 2011). Seetõttu toob ta põhiliseks miinuseks selle meetodi juures selle, et eukleidilise kauguse asemel võinuks olla võrgustikukaugus. Plussideks on Moreira (2011) sõnul see, et õnnetusi grupeeritakse statistilisel alusel (juhuslik lähima naabri kaugus ja olulisuse nivoo).

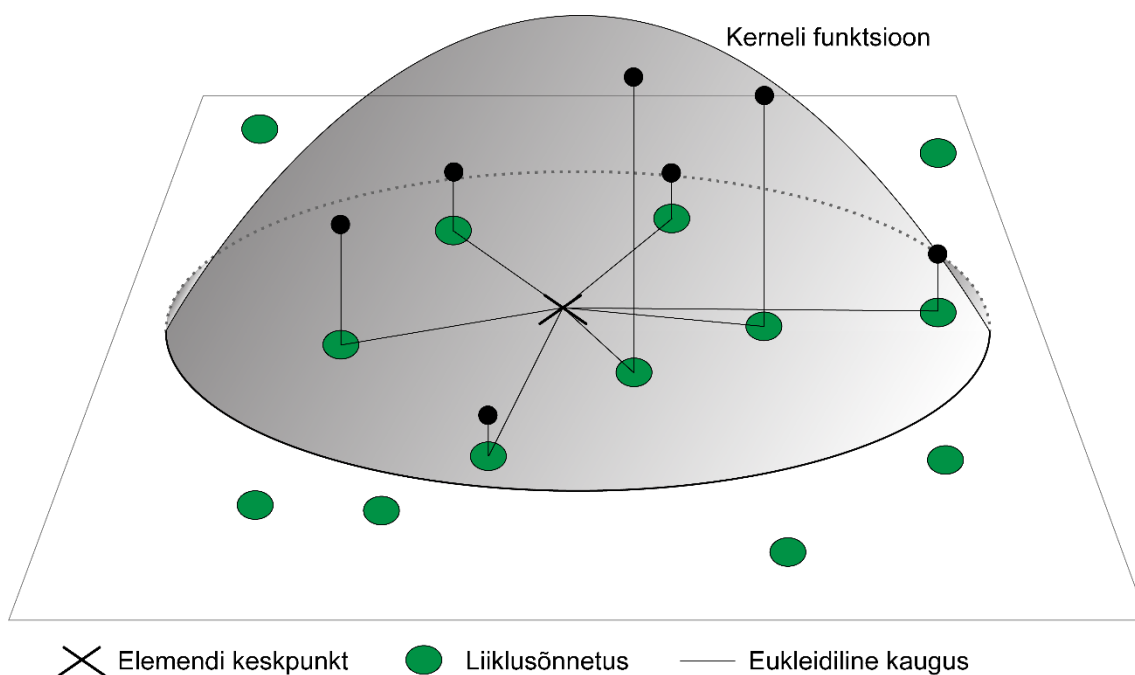
Smith ja Bruce (2008), CrimeStat III juhendi autorid, märgivad, et juhusliku lähima naabri kauguse kasutamise asemel on autoril võimalik ise määrata otsinguraadiuse kriteeriumi suurus, mille tulemused küll ei põhine tõenäosusel, vaid analüüsija enda nägemusel

uurimusest ja selle eesmärgist. Tihti peale on aga vaja katsetada mitmesuguste suurustega, et saada rahuldav tulemus. Samas märgivad nad, et juhusliku lähima naabri kauguse kasutamine annab statistiliselt kõige usaldusväärsemad tulemused (Smith ja Bruce, 2008).

HLNA tugevuseks on see, et meetod leiab nii suured kui ka väikesed õnnetuste koondumiskohad. Samuti sobib see kasutamiseks suurel ja väikesel tasandil (Smith ja Bruce, 2008). Plussiks on ka see, et tarkvara on vabavaraline (Eilat, 2011). Selle meetodi peamiseks puuduseks on see, et valesti määratud kriteeriumid võivad viia ebatäpsete tulemusteni (Smith ja Bruce, 2008). Eilat (2011) töö andmestikuna kasutati suhteliselt lühikese ajavahemiku jooksul kogutud andmeid loomaõnnetuste kohta ning seetõttu oli andmestikus vähe põdra- ja metsseaõnnetusi. Sellest tulenevalt tuli kasutada suurt otsinguraadiust ning väikest minimaalset õnnetuskohtade arvu, mis võib tähendada suuremat juhuslikult leitud koondumiskohtade arvu (Eilat, 2011).

1.3.3 Punktihedus kernelis

Kernel density estimation ehk KDE (eestipäraselt punktihedus kernelis) on meetod, kus uuritava ala kohale paigutatakse sümmeetrilised pinnad ning punktiheduse arvutamiseks summeeritakse nende punktide väärtused, mis jäävad otsinguraadiuse sisse (Anderson, 2009). Kernel on kas korrapärase kujuga (enamasti ruut, ring, sektor või sõõrik) või kujutise eelneva segmenteerimise abil saadud suvalise kuju ja suurusega eraldi ehk geograafiline kernel (Merchant 1984 cit., Remm et al., 2012). Sümmeetrilise pinna väärtus on suurim uuritava elemendi keskkohas ning sellest eemalduses väärtus kahaneb (joonis 5). Väärtuse kahanemise kiirus sõltub kasutatavast kerneli funktsioonist.



Joonis 5. Sümmeetriline pind tasapinnalise kerneli kohal (Produit et al., 2010 järgi)

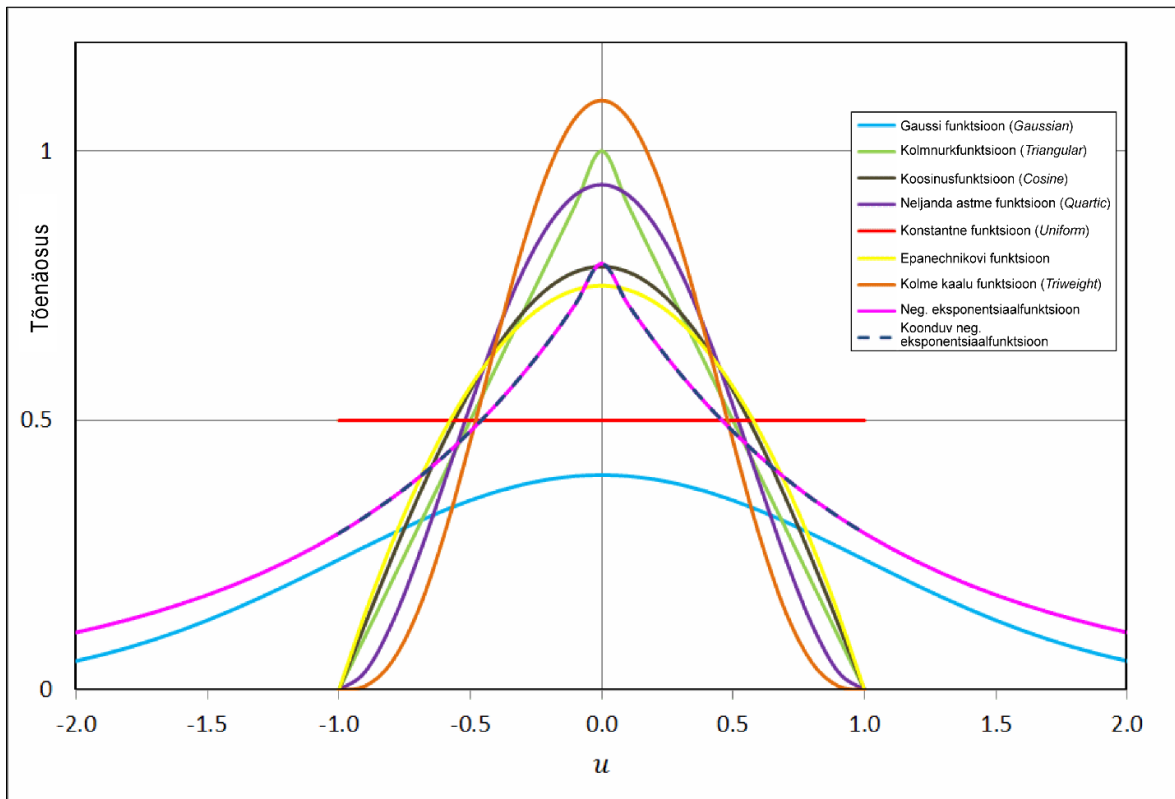
1.3.3.1 KDE tasandil

Seni on punktitihedust kernelis enamasti arvatud tasandil, kus esmalt jaotatakse ala võrdse suurusega ruutudeks (piksliteks), mille punktitiheus arvutatakse otsinguraadiuse sisse jäänud punktide alusel. Punktide väärtus sõltub kasutatavast kerneli funktsioonist ning kaugusest piksli keskmest. Valemi üldkuju tasandil arvutamiseks on järgmine (Xie ja Yan, 2008; valem 8):

$$\lambda(s) = \sum_{i=1}^n \frac{1}{\pi r^2} k\left(\frac{d_{is}}{r}\right) \quad (\text{Valem 8})$$

kus $\lambda(s)$ on punktitiheus kohas s , r on otsinguraadius (*bandwidth*), kuid teisiti öeldes ka silumislaius. Punktitiheduse arvutamiseks kasutatakse ainult neid punkte, mis jäävad otsinguraadiuse sisse. K on punkti i kaal ehk väärtus teatud kaugusel kohast s (Xie ja Yan, 2008). K kaal sõltub funktsioonist (kerneli funktsioon), mis väljendub kauguse ja koha s suhtest. Selle funktsiooni tõttu ei saa kogu sümmeetriline pind sama väärtust kogu otsinguraadiuse ulatuses, vaid eemaldudes kohast s selle pinna väärtus väheneb. Ehk mida suurem on punkti a kaugus koha s keskpunktist, seda väiksem on selle väärtus kogu punktitiheduse arvutamisel (Xie ja Yan, 2008). Kaugusena kasutatakse tasandil eukleidilist kaugust.

Kaugusest tuleneva kaalu vähenemise efekti mõõtmiseks on kasutusel mitmeid kerneli funktsioone. Põhilised on Gaussi (*Gaussian*), neljanda astme (*Quartic*), kooniline funktsioon (*conic*), negatiivne eksponentsiaalfunktsioon (*negative exponential*) ja Epanechnikovi funktsioon (joonis 6).



Joonis 6. Põhilised kerneli funktsioonid (Spatial Analysis, 2015 järgi)

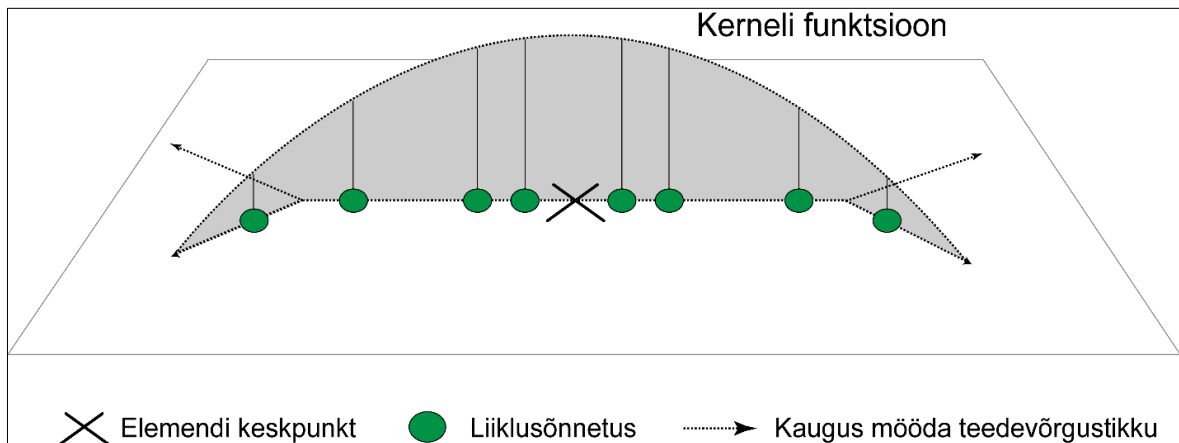
Tuginedes kirjandusele, mis on võrrelnud eri funktsioonide mõju ja tulemusi, levib arusaam, et kõige olulisem parameeter KDE juures on otsinguraadius (*bandwidth* ehk r) ning funktsioon ise ja kerneli suurus on vähem olulised (Xie ja Yan, 2008). Otsinguraadiusest sõltub otseselt ka see, kui sujuv on saadud tulemus. Liiga väikese r väärtuse korral on tulemus väga konarlik ja liiga detailne, kuid liiga suur r toob kaasa paljude konarluste väljasilumise ning väikesed, kuid olulised erinevused punktiheduses, ei pruugi välja tulla (Xie ja Yan, 2008).

1.3.3.2 KDE võrgustikul

Nähtuste analüüsimiseks, mis saavad toimuda ainult ühemõõtmelises ruumis, on otstarbekas kasutada kaugust mööda võrgustikku ehk selle töö puhul kaugust mööda maanteid. Valemi üldkuju KDE võrgustikul arvutamiseks on järgmine (Xie ja Yan, 2008; valem 9):

$$\lambda(s) = \sum_{i=1}^n \frac{1}{r} k\left(\frac{d_{is}}{r}\right) \quad (\text{Valem 9})$$

Kui tasandil arvutati tihedus pinnauhiku kohta, siis võrgustikul arvutatakse tihedus lineaarse ühiku kohta. Lineaarseks ühikuks on teatud pikkusega teelõigud. Kasutada on võimalik samu kerneli funktsioone, mis on nimetatud peatükis 1.3.3.1. Sarnaselt tasandil arvutamisega, jagatakse kõigepealt uuritav ala, mis võrgustiku puhul on teed, väiksemateks võrdse pikkusega lõikudeks. Seejärel leitakse igale lõigule KDE väärtus, summeerides otsinguraadiuse sisse jäävate punktide väärtused sõltuvalt kasutatavast kerneli funktsioonist ning nende punktide kaugusest teelõigu keskmest (joonis 7). Punktid, mis on teelõigu keskkohale lähemal, annavad lõigu punktihedusele suurema väärtuse ehk kaalu. Selles magistritöös uuritakse liiklusõnnetuste asukohtasid kujutavate punktide tihedust Eesti riigimaanteedel ning iga õnnetuse kaaluks on 1. Sellise kaalu annavad aga ainult need õnnetused, mis asuvad täpselt selle teelõigu keskmes, millele punktihedust arvutatakse. Keskosast eemal olevate õnnetuste kaal on väiksem. Mida suurem on kaugus vaadeldava punkti ja teelõigu keskosa vahel, seda väiksema väärtuse annab see punkt kogu lõigu punktihedusele.



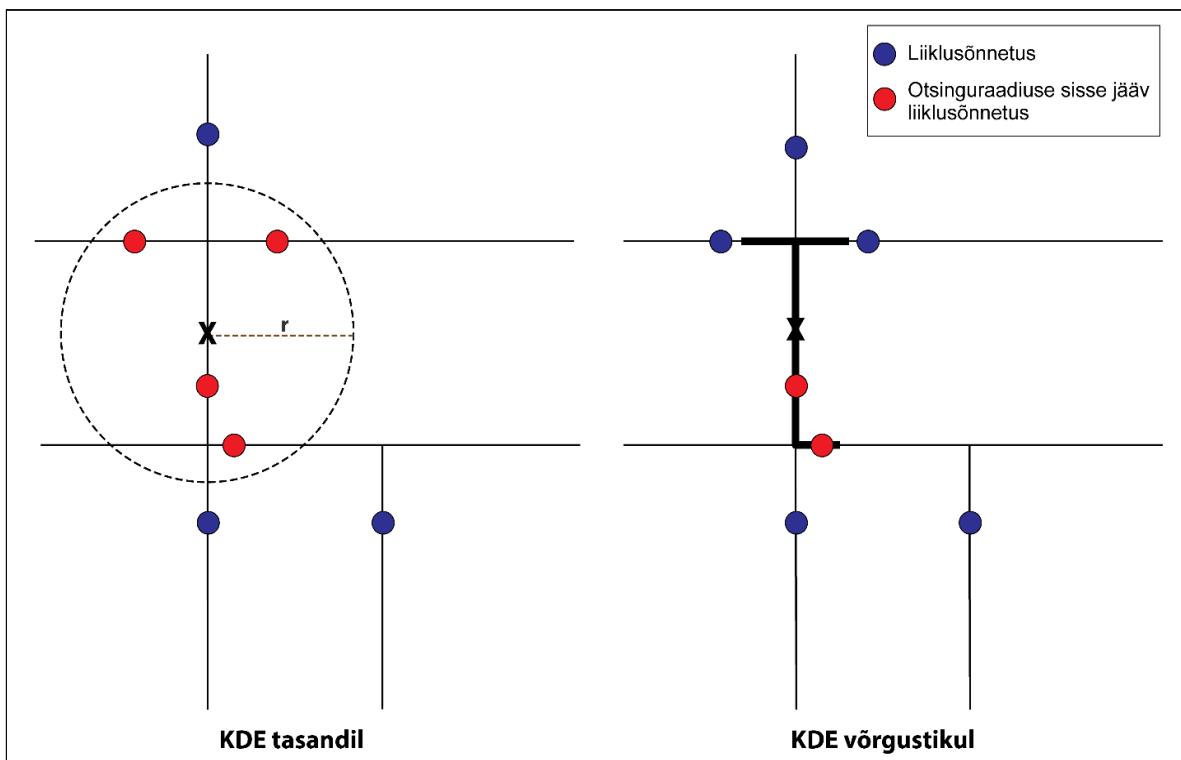
Joonis 7. Sümmeetriline pind kindla pikkusega teelõigu kohal (Produit et al., 2010 järgi)

On oluline rõhutada, et KDE arvutamine tasandil ja võrgustikul on erinev mitme aspekti poolest:

- KDE väärtust võrgustikul saab arvutada vaid siis, kui kasutada intsidente, mida on ruumis võimalik kujutada punktina (tasandil on võimalik arvutada punktihedus ka joonte kohta);

- kerneli funktsioon ning otsinguraadius põhineb kaugusel mööda võrgustikku mitte eukleidilisel kaugusel;
- punktihedus arvutatakse lineaarse ehk ühemõõtmelise, mitte mitmemõõtmelise ühiku kohta (Xie ja Yan, 2008).

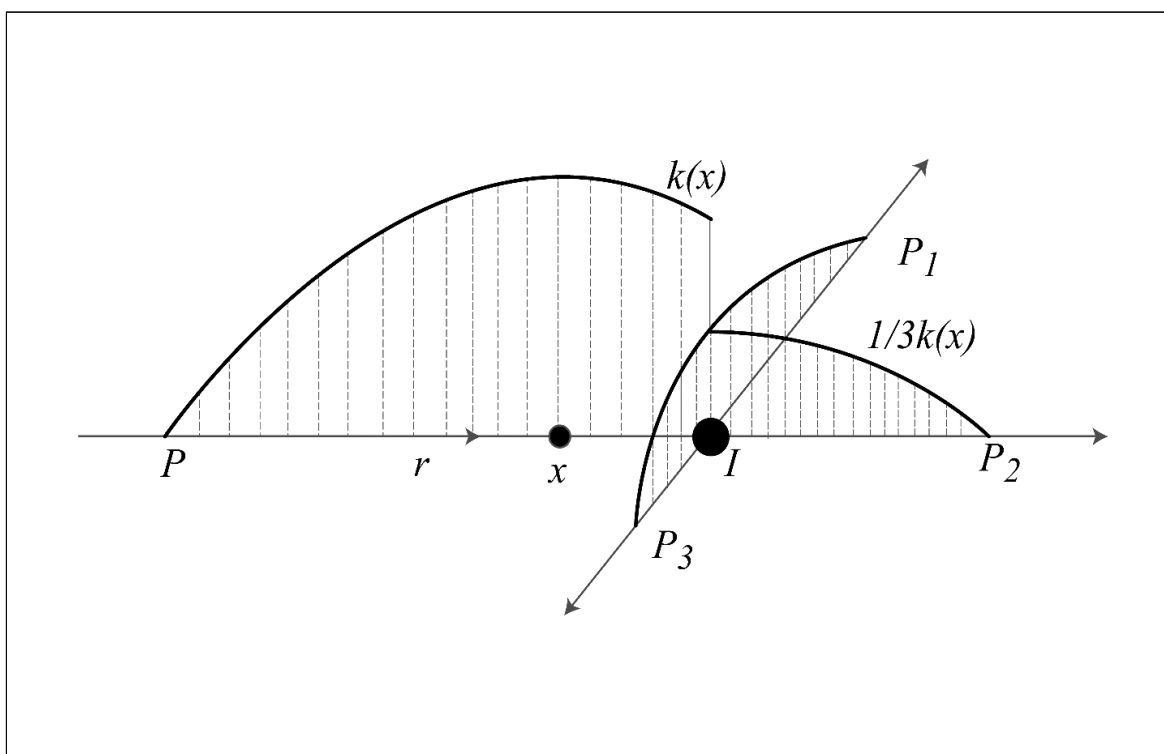
KDE väärtus võrgustikul arvutatakse ühemõõtmelises ruumis, kuid tasandilise KDE väärtus kahemõõtmelises. Kõigi eespool kirjeldatud erinevuste tagajärjeks on see, et kahe meetodi tulemused on väga erinevad, kui kasutada sama andmestikku (Xie ja Yan, 2008). Näiteks ühe elemendi keskpunktist (olgu selleks tasandilise KDE puhul ruudukujuline kernel küljepikkusega 500 meetrit ning võrgustiku KDE puhul teelõik pikkusega 500 meetrit) võib KDE tasandil kilomeetrist otsinguraadiust ja eukleidilist kaugust kasutades leida 4 õnnetusepunkti, kuid tegelikkuses mööda teedevõrku jääb kilomeetrise raadiuse sisse ainult kaks õnnetust (joonis 8).



Joonis 8. KDE arvutamisel kasutatavate kauguste erinevused tasandil ja võrgustikul (Xie ja Yan, 2008 järgi)

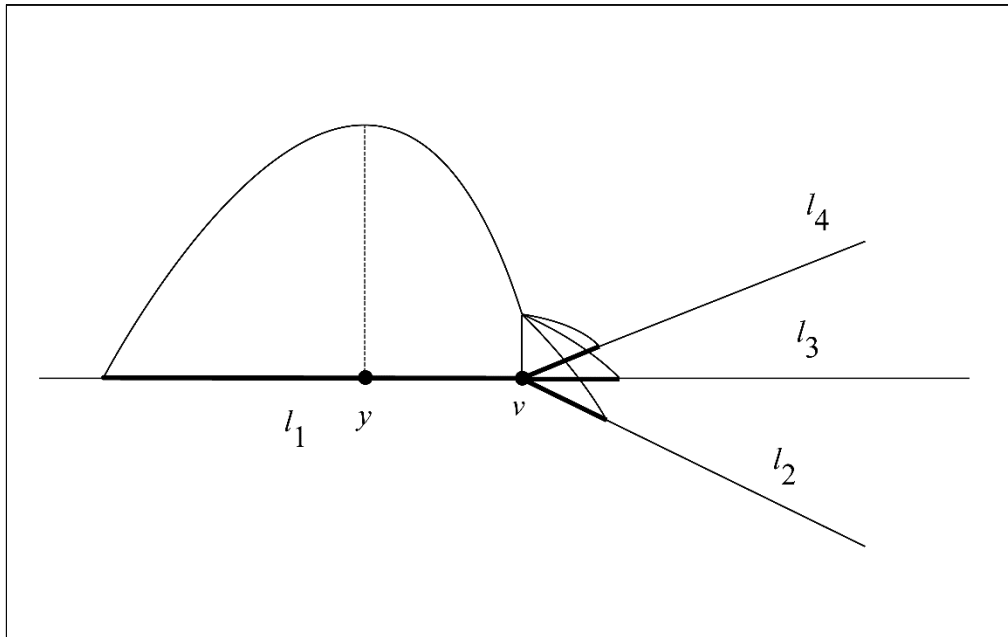
KDE väärtuse arvutamine teede sõlmpunktidel ehk ristmikel on mõnevõrra keerulisem kui ühel teesegmendil. Sõlmpunktides on reeglina suurem lõikude arv, mis pikendab otsinguraadiust, millest tulenevalt on ristmikel punktihedust keerulisem arvutada ning tulemus ei pruugi olla veatu ja usaldusväärne (Tang et al., 2015). Okabe et al. (2008), kes on

SANETi töövahendi loojad, töötasid välja kerneli funktsiooni, mis tagab õige KDE väärtuse ka ristmikel. Selleks on *equal-split discontinuous* funktsioon. Siirdepinna, mis on paigutatud segmendi kohale, väärtus kahaneb kauguse kasvades tavapäraselt. Väärtuse kahanemise kiirus vastab kasutatavale kerneli funktsioonile. Kohas, kus siirdepind on jõudnud sõlmpunkti kohale, jaguneb selle väärtus sõlmpunktiga külgnevate lõikude vahel ära (joonis 9). Näiteks kui ristmiku kohal on kerneli funktsiooni väärtus kahanenud kauguse tõttu 0,6 peale ning sõlmpunktiga külgnub 3 teist segmenti, siis saab iga külgnev segment ristmiku kohal väärtuse 0,2 ehk $0,6/3$. Ristmikust eemaldudes hakkab väärtus taas kahanema vastavalt kasutatavale kerneli funktsioonile.



Joonis 9. Kerneli *equal-split discontinuous* funktsiooni väärtused ristmikul (Tang et al., 2015 järgi)

SANETi tarkvara võimaldab KDE arvutamiseks kasutada ka teist kerneli funktsiooni, milleks on *equal split continuous*. Kui katkendliku funktsiooni puhul (*discontinuous*) jagatakse ristmiku kohal väärtus külgnevate segmentide vahel võrdselt, siis pideva funktsiooni (*continuous*) puhul saab iga segment selle väärtuse, mis on ristmiku kohal (joonis 10).



Joonis 10. Kerneli *equal-split continuous* funktsiooni väärtused ristmikul (Okabe et al., 2008 järgi)

Mahukat andmestikku kasutades on üheks piiravaks teguriks SANETi KDE puhul arvutuste kiirus. See sõltub eelkõige otsinguraadiuse suurusest ning teesegmendi pikkusest (Okabe et al., 2008). Eelkõige puudutab arvutuskiiruse probleem *equal-split continuous* funktsiooni, kus kasutades suuremat otsinguraadiust ning väikest segmendi pikkust, kasvab arvutusteks kuluv aeg mitmekordseks. Seetõttu soovitavad SANETi autorid kasutada *equal-split discontinuous* funktsiooni, kui tegu on andmetega, kus on rohkelt lühikesi segmente ning kasutatakse suurt otsinguraadiust (Okabe et al., 2008).

1.3.3.3 KDE varasemates uurimustes

Xie ja Jan (2008) kinnitasid oma uurimuses, et kõige suuremat mõju KDE tulemustele avaldab silumislaiuse ehk otsinguraadiuse parameeter. Sama seaduspära kinnitas oma töös ka Anderson (2009), kuigi tema kasutas tasandile mõeldud kerneli punktiheduse meetodit. Lõigu pikkus avaldab mõju vaid lokaalsel tasandil (Xie ja Jan, 2008). Väiksema silumislaiusega leitakse lokaalsed erinevused, kuid suur otsinguraadius leiab eelkõige suurima väärtusega lõigud. Tang et al. (2013) leidsid, et KDE, nii tasandil kui ka võrgustikule kohandatuna, on koondumiskohtade leidmiseks teiste meetoditega võrreldes efektiivsem. Väga hästi sobib see maanteede koondumiskohtade leidmiseks ehk suuremas mastaabis. Võrgustikule kohandatud KDE annab täpsemad tulemused kui tasandiline punktihedus kernelis (Tang et al., 2013).

Kuigi optimaalse lõigu pikkuse pakkumine teadlaste poolt on väga ahvatlev, siis tegelikkuses ei pruugi iga andmestiku jaoks ainuõiget suurust eksisteerida (Xie ja Jan, 2008). Xie ja Jan (2008) kasutasid oma töös 10-meetriseid lõike. Kui kümne meetri pikkune lõik võiks linnas toimunud õnnetuste jaoks olla optimaalne, siis teiste punktobjektide analüüsimiseks võivad sobida hoopis teised suurused. Sarnastel põhjustel ei paku töö autorid välja ka optimaalset silumislaiust, kuna ka see peaks sõltuma konkreetsest uurimisteedest. Tang et al. (2015) peab silumislaiuse ja lõigu pikkuse parameetreid lausa piiranguteks, kuna need väärtused peavad olema hoolikalt valitud. Eri andmestike puhul võivad aga nende parameetrite optimaalsed väärtused olla erinevad.

Xie ja Jan (2008) tõdevad ka KDE meetodi puudujäärke. Esimene neist on see, et kõik teesegmendid on võrdsed, kuigi nende liiklustingimused on erinevad. Teed erinevad üksteisest läbilaskevõime, kiiruspiirangute ja sõiduridade arvu poolest. Seetõttu oleks huvitav ka katse, kus arvestatakse ka teede liiklustingimusi. Kerneli punktutiheduse üheks puudujärgiks on see, et statistilist olulisust ei ole võimalik kontrollida ning seetõttu ei ole võimalik kindlalt öelda, millisest punktutiheduse väärtusest on tegu statistiliselt olulise koondumiskohaga (Xie ja Jan, 2008). Statistilise olulisuse puudujärgi rõhutavad oma töös ka Erdogan et al. (2015), Bills (2009), Anderson (2009) ja Nie et al. (2015).

Viimastel aastatel on püütud kerneli punktutiheduse meetodi statistilise olulisuse probleemi lahendada, töötades välja tarkvarasiseseid meetodeid selle kontrollimiseks (tarkvara KDE+). Bil et al. (2013, 2015) on kasutanud Monte Carlo simulatsiooni, mille abil tekitatakse uuritavale alale sama arv punkte, mis on ka sisendandmestikus. Seejärel leitakse juhuslikult genereeritud punktide alusel tiheduspinnad. Detailsem kirjeldus meetodi põhimõttest on CrimeStat IV näitel esitatud ka selles töös (peatükk 1.3.2.4). Bil et al. (2013) võrdles välja töötatud KDE+ tulemusi teiste koondumiskohtade leidmiseks kasutatavate meetoditega. Bil et al. (2013, 2015) väidavad, et nende KDE+ annab kõigist meetoditest kõige objektiivsema tulemuse ning on stabiilne ja efektiivne töövahend joonobjektidel toimunud liiklusõnnetuste analüüsimiseks. Seda meetodit tahtsid oma töös kasutada Peet ja Padrik (2015), kuid nende eesmärgile see ei vastanud. Täpsemad põhjused, miks KDE+ kasutusse ei läinud, on selgitatud peatükis 1.2.2. Xie ja Jan (2013) kasutasid KDE väärtuste statistilise olulisuse kontrolliks Morani I lokaalstatistikut. Sisendandmestikuna kasutati juba KDE poolt arvutatud punktutiheduse väärtusi. Sarnase meetodikaga, kuid ArcGISi *Kernel Density* töövahendi tulemuste alusel, arvasid Peet ja Padrik (2015) Getis-Ord lokaalstatistiku väärtused, et leida statistiliselt olulised KDE punktid.

2 Andmed ja metoodika

2.1 Lähteandmed

Töös kasutati andmestikuna Politsei- ja Piirivalveameti poolt riigimaanteedel fikseeritud liiklusõnnetuste andmeid, mis on leitavad Riiklikust Teeregistrist (Jairus, 2015). Õnnetuste andmebaasi väljavõte saadi *shp* formaadis punktobjektidena Maanteeameti GIS-juhtivhaldurilt Tanel Jairuselt 26. oktoobril 2015. aastal. Andmebaasi väljavõttes oli 18526 õnnetust. Neist varaseim oli fikseeritud 1. jaanuaril 2000. aastal ning hiliseim 29. juunil 2015. aastal.

Selles magistritöös kasutati ainult aastatel 2013–2015 toimunud õnnetusi ning neid oli saadud andmestikus 1633. 22.04.2016 saadi Maanteeametist täiendav väljavõte liiklusõnnetuste andmekogust MALIS. Väljavõttest saadi varasemast andmestikust puudu olnud õnnetused perioodist 01.07.2015 kuni 31.12.2015. Erinevused kahe andmestiku vahel seisnevad selles, et Riiklikus Teeregistris on kõik riigimaanteedel toimunud õnnetused, kuid MALISE andmebaasis vaid need, kus inimesed said vigastada või hukkusid.

Iga õnnetuse kohta olid andmebaasis järgmised atribuutandmed:

- | | |
|------------------|--------------------|
| 1) Tee number | 7) Hukkunute arv |
| 2) Nimetus | 8) Vigastatute arv |
| 3) Kuupäev | 9) Aasta |
| 4) Kellaaeg | 10) X-koordinaat |
| 5) Õnnetuse tüüp | 11) Y-koordinaat |
| 6) Osalejate arv | 12) M-koordinaat |

Kuna nende õnnetuste puhul on tegu juhtumitega, mida registreerib politsei, siis puuduvad andmestikust kergemad liiklusõnnetused, milleks võivad olla plekimõlkimised ning intsidendid, kus osapooled lepivad omavahel kokku kahju hüvitamise (Äär, 2014).

Lisaks õnnetustele kasutati Eesti Topograafilise Andmekogu põhjal loodud teede ruumikujusid, mida on täiendatud Riikliku Teeregistri andmetega (Jairus, 2015). Ka see *shp* formaadis andmestik on saadud Tanel Jairuselt. Iga tee joonobjekti kohta olid andmebaasis järgmised atribuutandmed:

- 1) Tee number
- 2) Nimetus
- 3) Alguse m-koordinaat
- 4) Lõpu m-koordinaat
- 5) Keskmine ööpäevane liiklussagedus
- 6) Liiklussageduse protsentuaalne jaotus sõiduki liigi järgi
(sõiduauto/veoauto/autorong)
- 7) Liiklussageduse arvuline jaotus sõiduki liigi järgi (sõiduauto/veoauto/autorong)
- 8) Loenduse aasta

Lisaks nendele atribuutidele leiti tee numbri järgi igale teele tema liik (põhi-, tugi- või kõrvalmaantee).

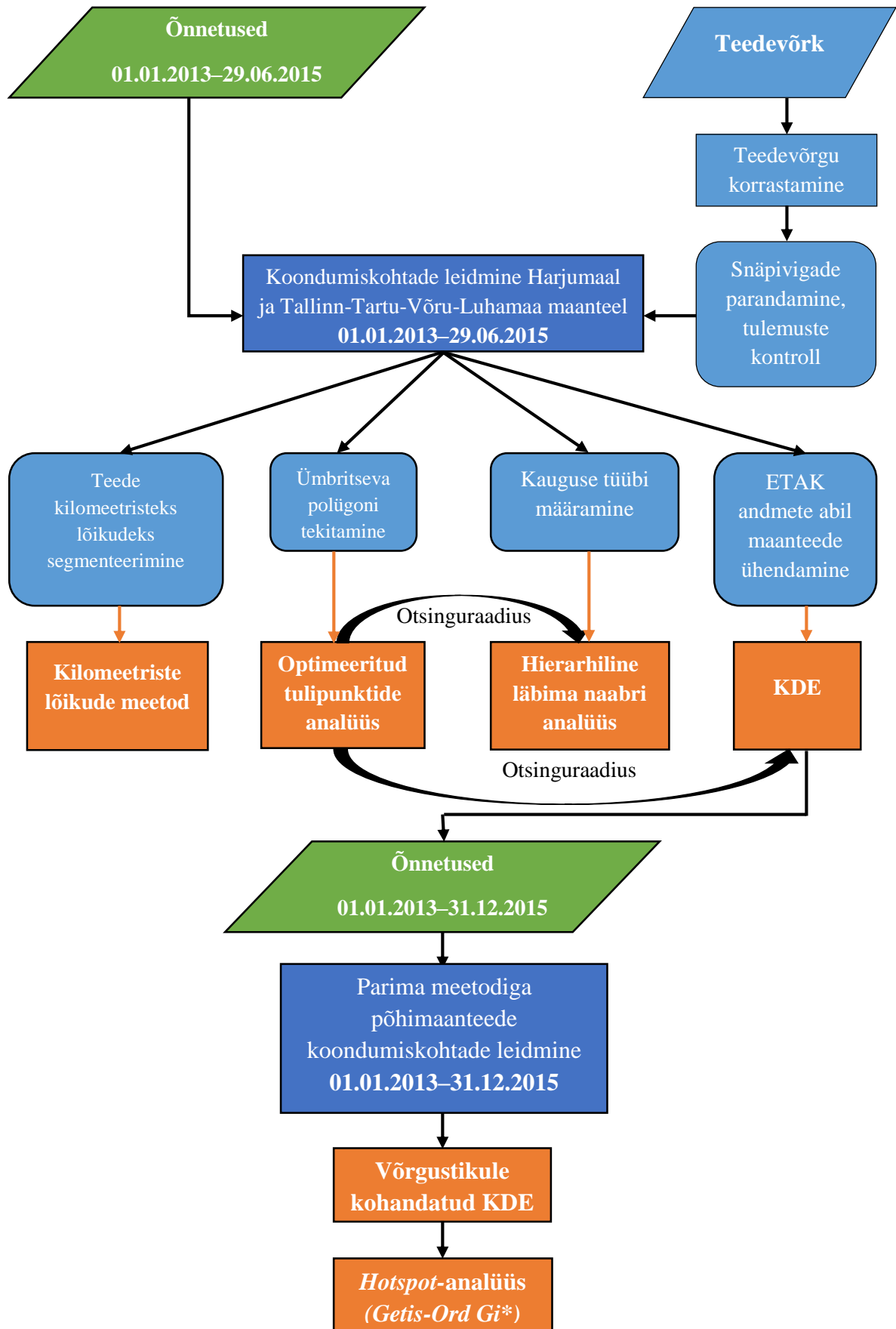
2.2 Metoodika

Metoodika koosneb kolmest etapist (joonis 11). Töö esimeses etapis korrastati teedekiht ja valmistati ette analüüsiks. Teises etapis leiti koondumiskohad ajavahemikus 01.01.2013 kuni 29.06.2015 toimunud liiklusõnnetustele Harjumaal (421 õnnetusjuhtumit) ja Tallinn – Tartu – Võru – Luhamaa maanteel (132 õnnetusjuhtumit). Liiklusõnnetuste koondumiskohtade leidmiseks kasutati nelja erinevat meetodit:

- õnnetuste arv kilomeetri pikkusel teelõikudel (ArcGIS);
- optimeeritud tulipuntide analüüs (ArcGIS);
- hierarhiline lähima naabri analüüs (CrimeStat IV);
- punktihedus kernelis (SANET).

Töö kolmandas etapis leiti parima selgunud meetodiga Eesti põhimaanteedel ajavahemikus 01.01.2013 kuni 31.12.2015 toimunud liiklusõnnetuste koondumiskohad.

Detailsemat õnnetuste analüüsi ja põhjusi ei ole selles töös analüüsitud. Samuti ei ole õnnetustele antud erinevaid kaalusid sõltuvalt tee liiklussagedusest või õnnetuse raskusastmest, kuna tegu on vaid ruumilise analüüsiga, mis näitab, kus esineb punktide koondumist. Et võimalikult palju välistada liiklustingimuste mõju õnnetustele, siis analüüsiti kolmandas etapis ainult põhimaanteedel toimunud intsidente. Seda põhjusel, et uuritavate teede hooldusklassid, projekteerimismid ja nõuded oleksid võimalikult sarnased.



Joonis 11. Metoodika plokkskeem

2.2.1 Teedevõrgu parandamine

Teeregistri andmestikuga tutvudes avastati, et mitte kõik teelõigud ei ole omavahel korrektselt ühendatud ehk snäpitud. Nendest vigadest tulenevalt võib mõni päring või töömeetod, mis eeldab sidusat võrku, mitte töötada (*Network Analyst*) või anda valesid tulemusi. Sellised vead tulid nähtavale väga suurt mõõtkava kasutades. Rippuvate otste kaugused selle teeni, millega nad peaksid ühendatud olema, jäid üldjuhul vahemikku 0,02 kuni 0,3 meetrit.

Snäpivigade parandamiseks kasutati *Extend Line* tööriista. Pikendamise parameetriks valiti 0,3 meetrit. Et hiljem kontrollida meetodi tulemust, leiti enne selle rakendamist üles kõigi teede rippuvad otsad. Selleks tekitati punktid iga joone algusesse ja lõppu. Igale punktile leiti joonega lõikumiste arv. Kui lõikumisi oli ainult üks, võis tegu olla varasemalt snäppimata jooneotsaga. Kontrollimise käigus kasutati ka Maa-ameti WMS teenuse ortofotosid ning Eesti Põhikaarti.

2.2.2 Liiklusõnnetuste arv kilomeetri pikkustel teelõikudel

Kilomeetriste lõikude tekitamiseks liideti teed üheks tervikuks tee numbril alusel ja genereeriti tee telgjoonele iga 1000 meetri tagant punkt (*Create Points on Lines*). Seejärel lõigati teed igas tekitatud punktis katki (*Split Line at Point*). Tulemuseks võimalikult kilomeetri pikkused lõigud, millele leiti ka nendel toimunud õnnetuste arv. Kuna teedevõrk on keerukas ja algandmestikuna kasutatakse ainult riigiteede kihti, tekkisid ka mõningad kilomeetrist lühemad teelõigud. Riigiteede kiht ei ole sidus sellepärast, et andmestikust puuduvad kohalikud teed ja erateed, mis ei kuulu riigiteede hulka. Sellest tulenevalt esines andmete tükeldatust ja ideaalseid kilomeetriseid lõike ei tekkinud. Lühemad segmendid tekkisid enamjaolt ristmikele ja kohtadesse, kus ka algandmestikus olid lühemad lõigud.

2.2.3 Optimeeritud tulipunktide analüüs

ArcGISi 10.2 optimeeritud tulipunktide analüüsi töövahendi ümbritseva polügoni (*bounding polygon*) saamiseks tekitati Harjumaa teedele puhver laiusega 1 meeter. Selline laiusega puhvrit kasutati seetõttu, et töövahend on mõeldud eelkõige pindadel kasutamiseks ning selleks, et kasutada seda lineaarsetel objektidel, tuleb hoida võimalikult väikese laiusega ümbritsevat polügoni. See määrab töövahendi jaoks ära kohad, kus intsidendid üleüldse

toimuda võivad. Katsetuste käigus andis väikese puhvri kasutamine kõige täpsema ja detailsema koondumiskohtade lõpptulemuse. Suurema puhvri puhul olid kalavõrk polügonid väga suured ning tulemus muutus väga üldistavaks.

Tallinn – Tartu – Võru – Luhamaa puhul andis meetrise puhvri kasutamine veateate ning tulemusteni tööriist ei jõudnud. Kuigi tööriist valib ise vastavalt sisendandmetele sobivad parameetrid, siis tegi ta meetrist puhvrit kasutades seda kuni analüüsiskaala valikuni. Skaala määramisel anti aga veakood, mis ütles, et valitud kaugus on uuritava ala jaoks liiga väike. Et hoida võimalikult kitsast kanalit meetodi toimimiseks joontel, prooviti meetri kaupa suurendada puhvri laiust. Seda tuli teha kuni 5 meetrini, kuna see oli esimene suurus, mis andis tulemusi.

Tallinn – Tartu – Võru – Luhamaa andmestiku puhul oli uuritava ala pindalaks 2 821 893 ning Harjumaa puhul 3 249 300 ruutmeetrit. Pindala saadakse teede puhvrist, mida kasutatakse ümbritseva polügonina. Kalavõrgu polügoni külje pikkuseks valis töövahend maantee puhul 207,5 ning Harjumaa andmestiku puhul 124,8 meetrit. Maanteele tekkis 1786 ning Harju maakonna teedele 15708 kalavõrgu polügoni. Maantee ja Harjumaa otsinguraadiused olid vastavalt 415 ja 1123 meetrit.

2.2.4 Hierarhiline lähima naabri analüüs

HLNA läbiviimiseks kasutati CrimeStat IV vabavaralist tarkvara. Minimaalseks punktide arvuks määrati 3 ehk antud juhul leiti klastrid, kus oli kolm ja rohkem õnnetust. Otsinguraadiusena kasutati eelkõige optimeeritud tulipunktide analüüsi tööriista poolt leitud väärtust (globaalne Morani I), mida ümardati mõningase lihtsustusena sajalisteni. Tallinn – Tartu – Võru – Luhamaa andmestiku puhul oli selleks suuruseks 400 ja Harju maakonna puhul 1100 meetrit.

Olenevalt nende raadiustega saadud tulemustest, kasutati Harju maakonna puhul ka autori subjektiivsel arvamusel põhinevat otsinguraadiust, et selle parameetri mõju oleks võimalik võrrelda. Sobivam meetod tulemuste visualiseerimiseks oli kumer kate (*convex hull*), mida kasutas oma magistritöös ka Eilat (2011). Olulisusenivoona kasutati 95%, mis näitab, et 0,05 on tõenäosus, et kaks juhuslikult valitud punkti on üksteisele lähemal kui keskmine juhuslik kaugus. Juhuslikku lähima naabri kaugust, mis HLNA puhul määrab ise sobiva otsinguraadiuse, ei kasutatud, kuna selle meetodi toimimise aluseks on õige uuringuala

pindala. Viimase suurust on aga antud juhul võimatu määrata, kuna tegu on ainult joontel toimuvate intsidentidega.

2.2.4.1 HLNA kauguse tüübi määramine ja sellega kaasnenud probleemid

Kauguse tüübina taheti kasutada võrgustikukaugust, kuid selgus, et seda kasutades tarkvara korrektselt ei tööta. Teedekihtidele tekitati juurde kaks atribuudivälja (*From one-way flag* ja *To one-way flag*), mis määravad ära, kas tee on ühe- või kahesuunaline. Jättes atribuutide väärtused tühjaks eeldab programm, et teelõigud on kahesuunalised (Levine, 2013a). Selliselt toimides ning algoritmi käivitades ilmus veateade (*irrecoverable error*). Määrares nende atribuutide väärtusteks 0, mis samuti tähendab, et teelõik on kahesuunaline, ilmus sama veateade ning algoritm ei lähe tööle. Tööle hakkas vaid siis, kui määrati mõlema atribuudivälja väärtusteks ükskõik mis arv, välja arvatud 0. Erinevaid arve kasutades tulemus ei muutunud. Kuigi näiliselt hakkas programm tööle, siis sõltus selle korrektne toimimine otsinguraadiuse suurusest ning sisendpunktide arvust. Suurt otsinguraadiust kasutades ei jõudnud algoritm tulemuseni.

Harjumaa õnnetuste puhul (421 punkti) oli võimalik otsinguraadiusena kasutada maksimaalselt 250 meetrit ning sellise väärtusega jõuti mõne sekundiga ka tulemuseni. Suurendades otsinguraadiust vähemalgi määral, näiteks 260 meetri peale, jäigi algoritm arvutama ning tulemusi ei tulnud. Sarnaselt käitus võrgustikukaugusega HLNA ka Tallinn – Tartu – Võru – Luhamaa puhul (132 sisendpunkti), kuigi väiksema punktide arvu ning väiksema summaarse teepikkuse tõttu oli maksimaalne kasutatav otsinguraadius mõnevõrra suurem, ulatudes kuni 900 meetrini. Sellise otsinguraadiusega leidis HLNA Tallinn – Tartu – Võru – Luhamaa maanteelt kaks klastrit. Visuaalselt kontrollides oli õnnetusepunkte, mis vastasid määratud kriteeriumitele (minimaalselt 3 punkti ning otsinguraadiuseks 900 meetrit mööda teedevõrgustikku) ja oleksid võinud moodustada klasteri, kindlasti rohkem kui kaks.

Ühe põhjusena arvati olevat seda, et teedevõrk ei ole topoloogiliselt päris korrektne, kuigi KDE kasutamiseks seda teatud määral juba korrastati (snäpivead). Katsetuseks võeti üks Eesti baaskaardi kaardileht topoloogiliselt korras teedevõrguga. Juurde tekitati kaks atribuudivälja ning väärtustega 0 programm tööle ei läinud, kuid väärtustega 1 läks. Et proovida, kas leitakse ka klastreid ja millise otsinguraadiusega, siis tekitati suvalistesse kohtadesse õnnetusi imiteerivad punktid, kord üksteise lähedale, kord kaugemale. Kokku tuli neid 85 ning programm töötas isegi suurte otsinguraadiustega. See aga tuleb tõenäoliselt sellest, et väikest punktide arvu ja väiksema suuruse ning lihtsama ülesehitusega võrgustikul toimib programm korrektselt, kuid kuni teatud punktide arvu ja otsinguraadiuseni. Selle

kontrollimiseks suurendati punktide arvu selle kaardilehe teedel kuni 421-ni ning seekord jäi programm juba 1000-meetrise otsinguraadiuse juures lõpmatult arvutama.

Viimase katsetusena muudeti Harjumaa teede ruumikujusid nii, et need oleksid ristmikel tükeldatud (*Feature To Line*) ning tehti väljavõte Harjumaa põhimaanteedest ja nendel toimunud õnnetustest, et vähendada punktide hulka ning teedevõrgu suurust. Põhimaanteedel toimus ajavahemikul 01.01.2013–29.06.2015 175 õnnetust. Paraku ei toonud ka sellise lihtsustuse tegemine kaasa programmi efektiivsuse paranemist ning maksimaalne võimalik otsinguraadius ei läinud oluliselt suuremaks kui eelnevalt töötanud 250 meetrit.

Kuna võrgustikukaugust kasutades ei toiminud CrimeStat korrektselt, siis ei välistatud ka seda, et kasutaja oli teinud vigu HLNA meetodi seadistamisel. Seetõttu kirjutas töö autor CrimeStat IV tarkvara loojale, Ned Levine'le. Selles kirjas selgitati lühidalt probleemi olemust ning lisati ka näidisandmestik. Ned Levine vastas, et ka tema ei ole võrgustikukaugust kasutades HLNAd korrektselt tööle saanud, kuigi teiste CrimeStat IV funktsioonidega pidi see töötama (Levine, 2016). Ta katsetas ka ise HLNA toimimist võrgustikukaugusega, kasutades andmeid, mis olid eelnevalt talle selle töö autori poolt kirjaga saadetud. Ka temale ei andnud algoritm tulemusi. Võrgustikukauguse asemel soovitas ta kasutada eukleidilist kaugust, mis peaks andma praktiliselt samad tulemused (Levine, 2016). Erinevused võivad olla kohtades, kus eukleidiline kaugus leiab klatri mitte ühel ja samal teel, vaid kahel lõikuval teesegmendil, kuid võrgustikukaugust kasutades oleks see klaster leidmata jäänud. Samas peaks selliste juhtumite arv olema väike.

Kuna võrgustikukaugust ei olnud võimalik kasutada, siis katsetati esmalt, kasutades juhuslikke kriteeriume, milliseid tulemusi annavad eukleidiline ja Manhattani kaugus ning millised on erinevused. Suurt otsinguraadiust ning eukleidilist kaugust kasutades oli juhuseid, kus HLNA moodustas klatri punktidest, millest kaks paiknesid ühel teesegmendil ning kolmas punkt hoopiski kilomeeter eemal teisel teelõigul. Õnnetusepunktid paiknesid nende teelõikude ristumiskohast oluliselt kaugemal ning nende segmentide liiklustingimused ja üldine liiklusohutus võisid olla väga erinevad. Samadel kriteeriumitel, kuid Manhattani kaugust kasutades, seal klatri ei leitud. Bills (2009) väidab, et Manhattani kaugus on sarnane mööda võrgustikku mõõdetud kaugusega. Kuna võrgustikukaugus ei toiminud korrektselt ning töö eesmärk on leida ohtlikud sarnaste liiklustingimustega teelõigud, siis katsetuste ja Billsi (2009) väidete põhjal sobis selle töö CrimeStati hierarhilise lähima naabri analüüsi jaoks kauguse tüübina paremini Manhattani kaugus.

2.2.5 Punktiühendus kernelis

KDE läbiviimiseks kasutati vabavaralist tarkvarakomplekti SANET. Andersoni (2009) sõnul on otsinguraadiuse suuruse määramine mõnevõrra subjektiivne. Tema valis oma töös rastrilemendi suuruseks, millele KDE arvutatakse, 100 meetrit ning otsinguraadiuseks 200 meetrit. Ehk otsinguraadius on kaks korda suurem teelõigu pikkusest. Kuna selles töös uuritakse õnnetuste koondumiskohti maanteedel, mitte linnas, siis tuleb valida suurem otsinguraadius ning mõnevõrra suurem lõigupikkus. Teistes töödes, kus on kasutatud KDE meetodit, on enamasti leitud linnasiseste õnnetuste koondumiskohti, kus vahemaad ning uuritav maa-ala on kordades väiksemad ning seetõttu on nendes uurimustes kasutatud parameetrite väiksemaid väärtusi.

Et vähendada autori subjektiivset arvamust otsinguraadiuse ehk silumislaiuse valikul, kasutati eelkõige optimeeritud tulipunktide analüüsi poolt leitud otsinguraadiust (400 ja 1100 meetrit). Olenevalt selle raadiusega saadud tulemustest, tehti ka teisi katsetusi autori subjektiivse arvamuse järgi valitud silumislaiusega, et võrrelda parameetri mõju tulemustele. Teelõigu pikkuse määramisel lähtuti põhimõttest, mis on kirjas SANETi juhendis ehk lõigu pikkus peaks olema ligikaudu kümme korda väiksem kui silumislaius (Okabe, 2015). Tallinn – Tartu – Võru – Luhamaa maantee puhul määrati lõigu pikkuseks 40 meetrit ning Harju maakonna andmestiku puhul mõningase lihtsustusena 100 meetrit. Kerneli funktsioonina kasutati SANETi KDE meetodi poolt vaikumisi seatud *equal-split continuous* funktsiooni. Katsetuste käigus ei olnud arvutusteks kuluv aeg suur ning selle funktsiooniga on ka ristmikualadele arvutatud väärtused pidevad.

Visualiseerimiseks ning koondumiskohta moodustavate väärtuste eristamiseks kasutati standardhälbe klassifikatsiooni (*standard deviation classification*) ning tulipunktiks määrati suurima standardhälbega klass. Selline valik tehti, kuna ka optimeeritud tulipunktide analüüs arvutab Z-väärtust, mis sisuliselt on standardhälve, mis näitab lokaalse punktiple keskmise erinevust üldisest keskmisest (Peet ja Padrik, 2015). Klassifitseerimise põhimõte seisneb selles, et esmalt leitakse väärtuste keskmine ning klassid moodustatakse keskmist väärtust ja standardhälvet summeerides. Koondumiskohaks määrati $>2,5$ standardhälbega klass.

3 Tulemused ja arutelu

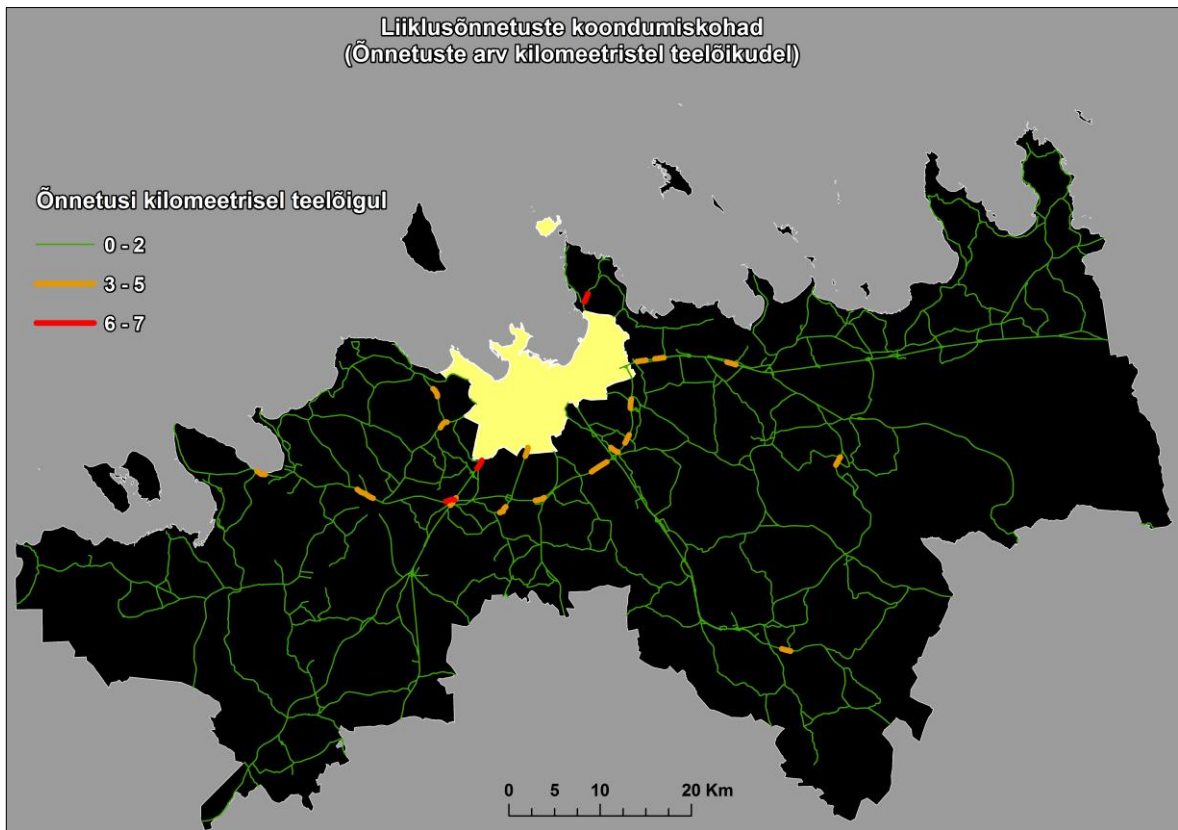
3.1 Liiklusõnnetuste arv kilomeetri pikkustel teelõikudel

Kilomeetrise lõikude meetodi järgi oli Tallinn – Tartu – Võru – Luhamaa maanteel seitse liiklusõnnetuste koondumiskohta (joonis 12). Tegelik arv võis olla veelgi suurem, kuna selles töö etapis kasutati kahe ja poole aasta andmeid (01.01.2013–29.06.2015). Neljal lõigul toimus kolm õnnetust ning ülejäänud kolmel vastavalt 4, 5 ja 8 õnnetust. Kõige ohtlikuma lõigu kolm õnnetust juhtus ristmikul.



Joonis 12. Koondumiskohad Tallinn – Tartu – Võru – Luhamaa maantee kilomeetristel teelõikudel

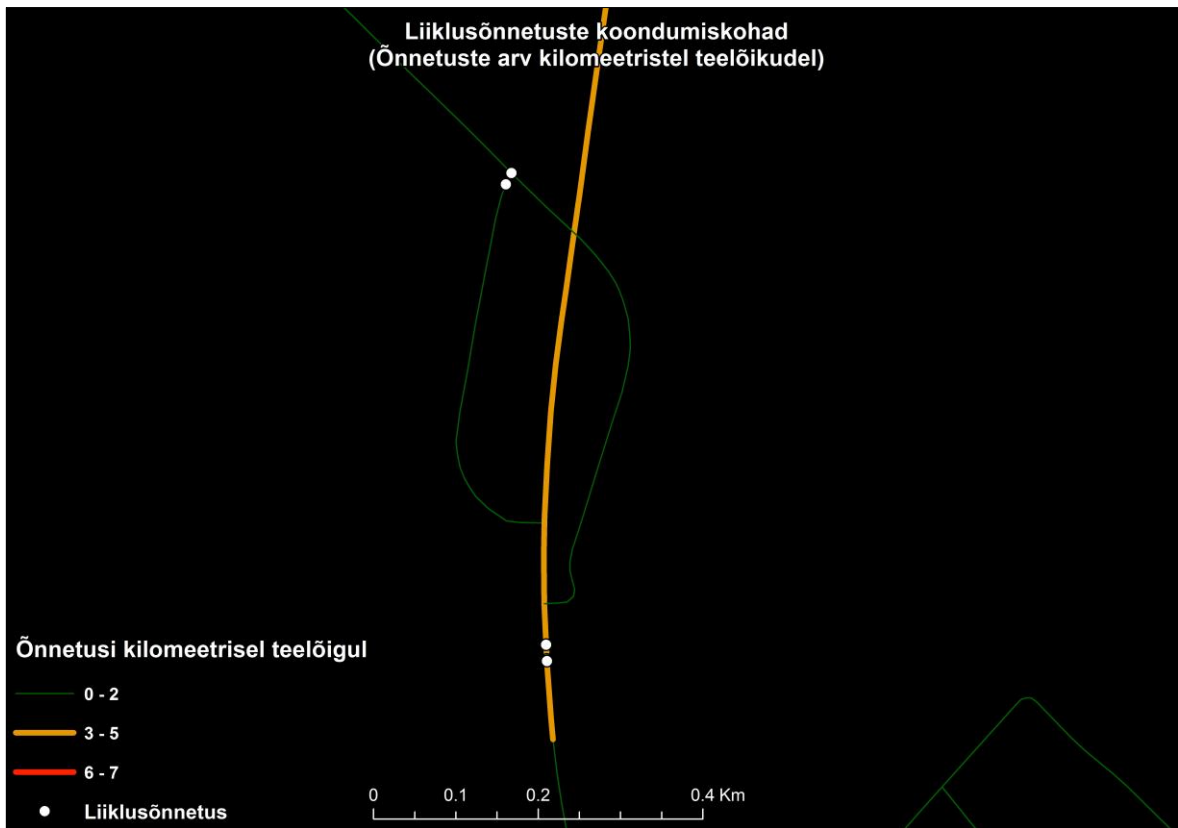
Harjumaal leiti sama meetodiga 22 koondumiskoha definitsioonile vastavat ohtlikku teelõiku (joonis 13). Viiel lõigul juhtus 4, kolmel 5, ühel 6 ning kahel 7 õnnetust. Suurem osa ohtlikest lõikudest paiknesid Tallinna ringteel (tee nr 11). Poolte tulipunktide hulka kuulus vähemalt üks õnnetus, mis juhtus kahe riigimaantee ristumiskohas.



Joonis 13. Koondumiskohad Harjumaa riigiteede kilomeetristel teelõikudel

Kuigi meetod on lihtne, kiiresti rakendatav ning annab ka arvestatavad tulemused, on sellel ka mõningaid miinuseid. Probleemsed on juhud, kus õnnetusepunktid paiknevad lähestikku, kuid satuvad segmenteerimise tõttu erinevatele kilomeetristele teelõikudele. Kui üks kilomeetri pikkune lõik lõpeb ning teine algab ja õnnetused on toimunud eri lõikudel, siis meetod õnnetuste koondumiskohta ei leia, kuigi vahemaa esimese ja kolmanda õnnetuse vahel on alla kilomeetri. Samasugust puudust märkis oma töös ka inseneribüroo Stratum (Stratum, 2013).

Lisaks on võimalus, et liiklusõnnetused juhtusid väga väikeste vahemaade tagant, kuid nad kõik jäävad kobarana kas kilomeetrise lõigu algusesse või päris lõppu (joonis 14). Joonisel 14 asuvad kolm õnnetust lõigu lõpust 100 meetri kaugusel. Ülejäänud 850 meetri pikkune teelõik, mis jääb ettepoole, võiks olla kõrgema õnnetuse riskiga ala, kuid õnnetustest edasi liikudes oleks selleks ohutsooniks vaid 100 meetrit. Kuna ohutsooni ulatus on õnnetuste paiknemisest ja segmenteerimise meetodikast tulenevalt väga erinev, siis ei ole kindlat reeglit, kuhu see ohtlik ala ühe või teise koondumiskoha puhul jääb (algusesse või lõppu).



Joonis 14. Liiklusõnnetuste paiknemine lõigu lõpuosas

Andmestiku puhul, mis sisaldab ka ristmikke, on reaalsed ka olukorrad, kus koondumiskoht leitakse vaid põhimaanteele, kuid sellega liituvale tugimaanteel toimunud õnnetused arvesse ei lähe. See on tingitud teede segmenteerimisest, kus ohtlik kilomeetrine lõik on tekkinud mööda põhimaanteed. Tugimaanteele on tekkinud aga eraldi lõik, mille kaks liiklusõnnetust ei ole koondumiskoha definitsiooni järgi ohtlikud. Sellest tulenevalt võib oluline osa koondumiskohast leidmata jääda, kuigi nii põhi- kui ka tugimaanteel toimunud õnnetusi võisid põhjustada samad faktorid. Lisaks eelnevalt mainitud kolmele puudusele ei ole selle meetodi tulemused statistiliselt kontrollitud.

Meetod on suhteliselt lihtne ning annab esmase ülevaate, kus on Eesti riigimaanteedel toimunud õnnetuste koondumiskohad. Kõige paremini töötab see meetod ühel konkreetsel teel. Andmestiku puhul, kus on palju mitmeid teid, keerulisi ristmikke ja ringristmikke, on seda raskem rakendada, kuna teede automaatselt kilomeetristeks lõikudeks segmenteerimine võib osutuda keeruliseks. Käesolevas magistritöös kasutati ühe andmestikuna Harjumaa riigimaanteed, kus juba algandmetes esines kilomeetrist lühemaid teelõike. Kui siia juurde arvestada keerulised ringteed ning asjaolu, et mitte kõik Harjumaa teed ei ole riigimaanteed, siis võib automaatne segmenteerimine osutuda arvatust palju keerulisemaks. Teoreetiliselt

võinuks kõik lõigud tulla võimalikult kilomeetri pikkused, kuid 1872st tekkinud lõigust 393 olid alla 900 meetri pikad. 100 lõiku olid pikkusega vähem kui 100 meetrit. Enamus nendest segmentidest olid teelt maha keeramise kohad või ringteed.

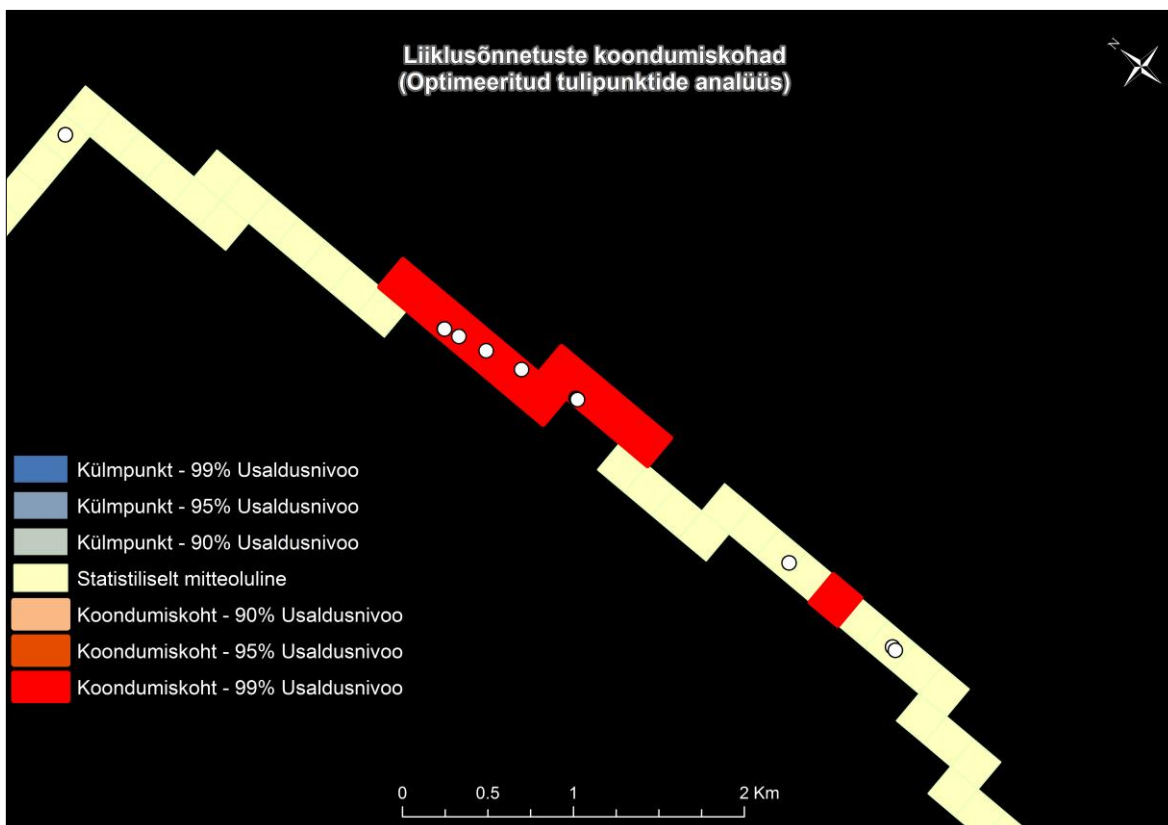
3.2 Optimeeritud tulipunktide analüüs

Tallinn – Tartu – Võru – Luhamaa andmestikul rakendatud optimeeritud tulipunktide analüüs tekitas uuritavale alale 1786 kalavõrgu polügoni. Neist 38 olid meetodi poolt leitud statistiliselt olulised liiklusõnnetuste koondumiskohad. 99%, 95% ja 90% usaldusnivooga polügone oli vastavalt 36, 1 ja 1. Tulemuseks olid samad koondumiskohad, mis leiti ka kilomeetrise lõikude meetodiga. Lisaks nendele leiti ka teisi tulipunkte, kus õnnetusi oli 3 ja rohkem. Kokku leiti 11 koondumiskohta. Mõned neist asuvad üksteise suhtes väga lähedal (joonis 15). Näiteks koondumiskoht, mis asub Tartust lõunapool, koosneb kahest ohtlikust teelõigust.



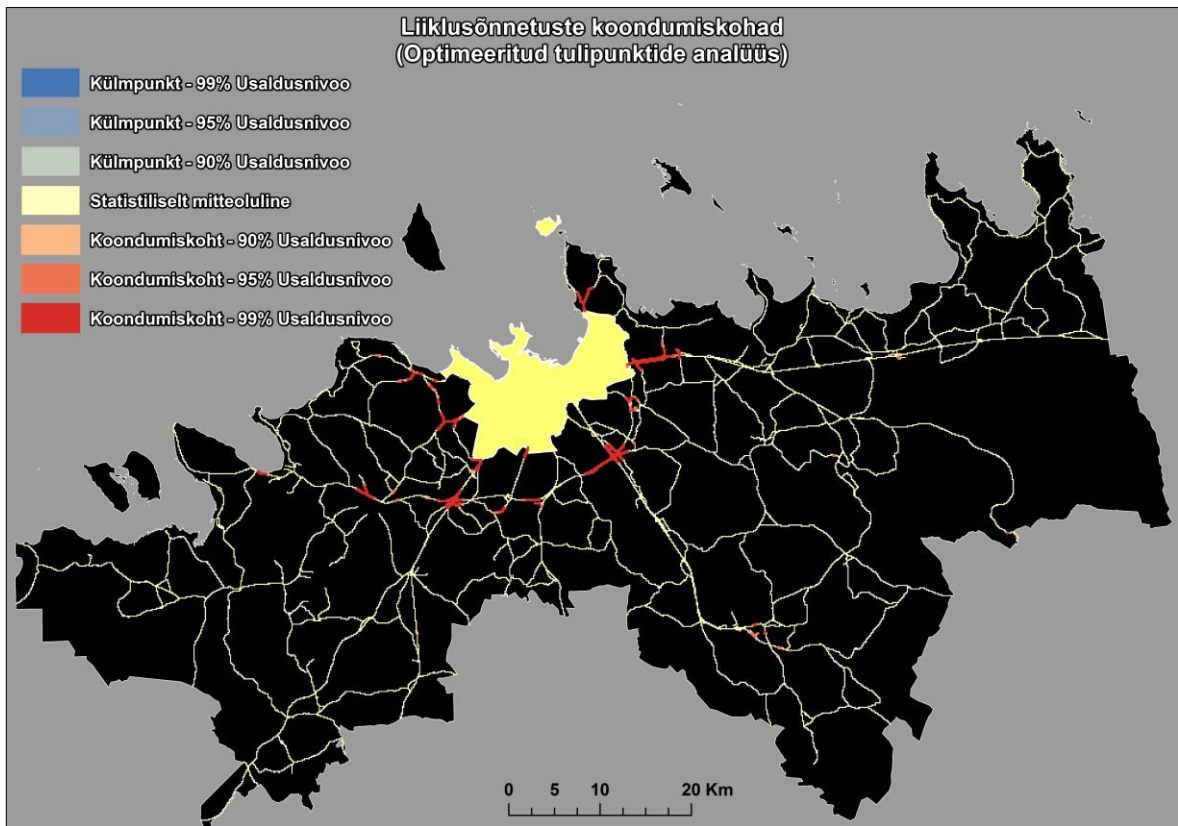
Joonis 15. Optimeeritud tulipunktide analüüsi leitud koondumiskohad

Teatud kohtade puhul leidis töövahend ka ohutsoonid, mis asusid väljaspool konkreetset koondumiskohta. Need ohutsoonid võivad olla kõrgema riskiga alad (joonis 16). Seetõttu on koondumiskohaks selle analüüsi järgi ka enne ja pärast õnnetusi paiknevad naaberalad. Tulipunktid ei olnud selle töövahendiga alati sellised. Väikese pikkusega teelõikudele, kus juhtus 3-4 õnnetust, tekitati statistiliselt usaldusväärsed kalavõrgu polügonid õnnetuste vahele (joonis 16 kaguosa). 11st koondumiskohast 5 olid just sellised. Need ei ole aga definitsioonile vastavad korrektsed tulipunktid.



Joonis 16. Optimeeritud tulipunktide analüüsi erinevad tulemused

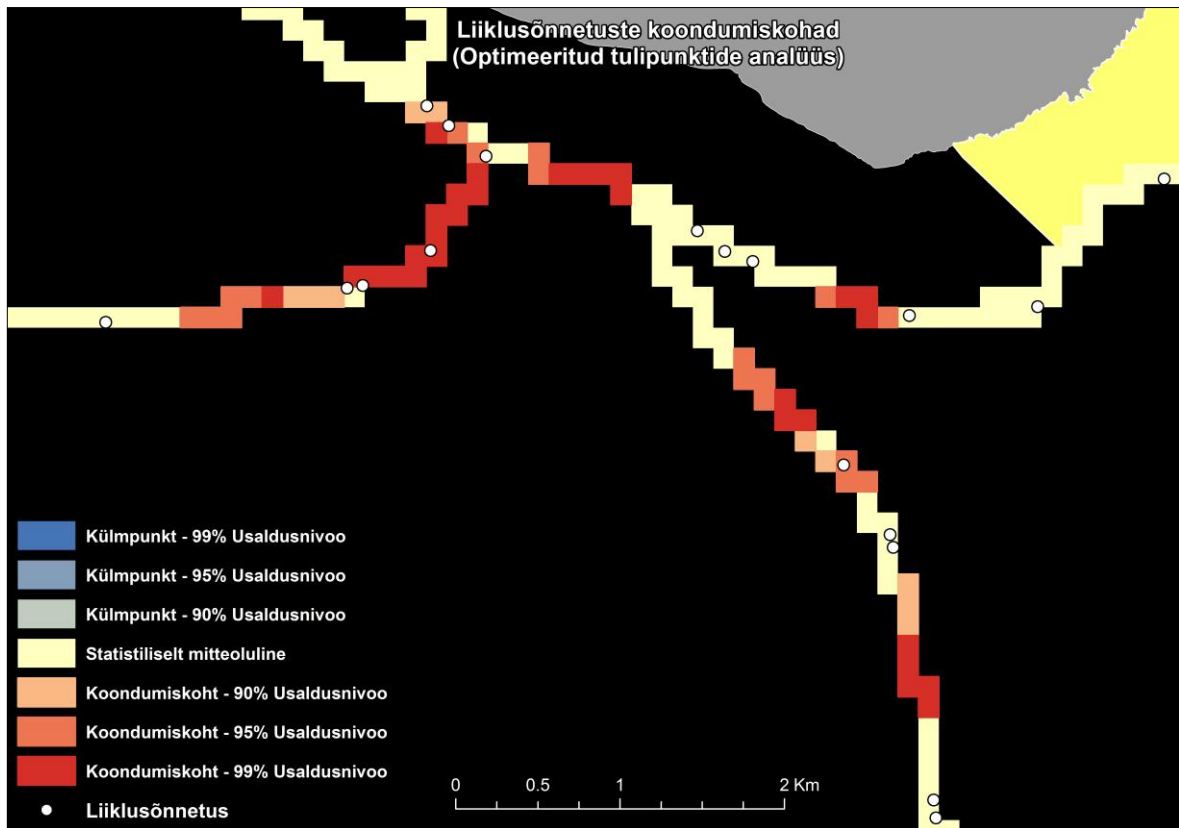
Harjumaa 15708st polügonist 784 olid meetodi poolt leitud statistiliselt olulised õnnetuste koondumiskohad (joonis 17). 99%, 95% ja 90% usaldusnivooga polügone oli vastavalt 578, 125 ja 81. Statistiliselt olulisi külmpunkte, kus õnnetusi toimub keskmisest vähem, meetod ei leidnud. Üldiselt olid tulipunktide asukohad sarnased kilomeetriste lõikude meetodiga leitud koondumiskohtadega. Lisaks nendele leiti üles ka ohtlikud ristmikud koos nende ohutsoonidega. Taas oli suurem osa tulipunktidest koondunud Tallinna ringteele. Joonisel 14 kujutatud kilomeetriste lõikude meetodiga leitud koondumiskohta optimeeritud tulipunktide analüüs ohtlikuks ei pidanud.



Joonis 17. Optimeeritud tulipunktide analüüsi leitud koondumiskohad

Väikeses mõõtkavas vaadatuna, nagu joonisel 17, tunduvad koondumiskohad koos oma ohutsoonidega piisavalt korrektsed. Seda on nad vaid piirkondades, kus on juhtunud palju õnnetusi. Vahel hindab meetod ohutsooni suure varuga üle, isegi kuni kilomeetri kaugusele äärmisest õnnetusest.

Kui suuritel ristmikel leitud koondumiskohtadega võib üldiselt rahule jääda, siis väiksemate riigiteede ristumiskohtade puhul on tulipunktid leitud õnnetuste vahele, nagu ka Tallinn – Tartu – Võru – Luhamaa andmestiku puhul (joonis 18). Joonise parempoolses osas on statistiliselt olulised polügonid leitud õnnetuste vahele ning kahe paralleelse tee puhul ka nende intsidentide raskuskeskmesse. Selliselt leitud nii-öelda tulipunkte oli Harjumaa andmestiku puhul mitmeid. Selle töö definitsiooni järgi neid aga koondumiskohtadeks ei tohiks lugeda.



Joonis 18. Optimeeritud tulipunktide analüüsi poolt leitud probleemsed koondumiskohad

Lisaks eelnevatele näidetele esines Harjumaa puhul statistiliselt olulisi tulipunkte kohtades, kus õnnetusi ei olnud ning lähim intsident koondumiskohale asus ligikaudu kolme kilomeetri kaugusel, kui mõõta kaugust mööda teed. Eukleidilise kauguse puhul oli sellele tulipunktile lähim õnnetus ligikaudu 400 meetri kaugusel.

Üldjoontes toimis ArcGISi automaatne optimeeritud tulipunktide analüüsi töövahend kahel andmestikul piisavalt hästi, et näidata ära umbkaudsed koondumiskohtade asukohad. Tööriista kasutamine on lihtne, kuna see vajab vaid kahte sisendandmestikku, millest üks on punktobjektide ning teine ümbritseva polügoni kiht. Kõigi parameetrite suurused valib töövahend vastavalt sisendandmestikule ise. Meetod on küll algselt välja töötatud pindadel kasutamiseks, kuid ümbritseva polügoni funktsioon võimaldas seda teatud mööndustega kasutada ka joontel.

Ümbritseva polügonina kasutati Harju maakonna puhul 1-meetrist ning Tallinn – Tartu – Võru – Luhamaa maantee puhul 5-meetrist puhvertsooni, mis määras ära, kus intsidendid üleüldse toimuda saavad (ArcGIS, 2015a). Ilma puhvertsoonita oli töövahendi tulemus väga üldistatud, sisaldades vaid üksikuid kalavõrgu polügone üle kogu uuritava ala. Mida kitsamat puhvrit ümbritseva polügonina kasutati, seda täpsemaks muutusid tulemused. Tallinn –

Tartu – Võru – Luhamaa maantee puhul andis töövahend vigu, kui kasutati liiga väikese puhvri suurusega polügoni. Tulemused tulid alles siis, kui kasutati 5-meetrist puhvrit. Ka arvutusteks kuluv aeg oli mõlema andmestiku puhul väga suur, ulatudes tundidesse. Tallinn – Tartu – Võru – Luhamaa maantee katsetus 1-meetrise puhvriga töötas 4 tundi ja 15 minutit ning alles siis andis veateate. Tööriist seiskus otsinguraadiuse määramise etapi juures.

Põhjus, miks väiksema puhvri kasutamine mõjutab kalavõrgu polügonide suuruse valikut, jääb selgusetuks. Huvipärast tehti katsetus ka ArcGIS 10.3 versiooniga, kus kasutati sama optimeeritud tulipunktide analüüsi töövahendit. Mitme erineva suurusega puhvri kasutamine ümbritseva polügonina ei muutnud tulemustes mitte midagi. Olenemata puhvrist valis tööriist kalavõrgu polügonide küljepikkuseks väga suured väärtused ning analüüsi tulemus ei olnud kasutuskõlblik. Selles töös kasutatud ArcGIS 10.2 versiooni puhul on puhvri suurusel aga väga suur mõju tulemuste täpsusele ja õigsusele.

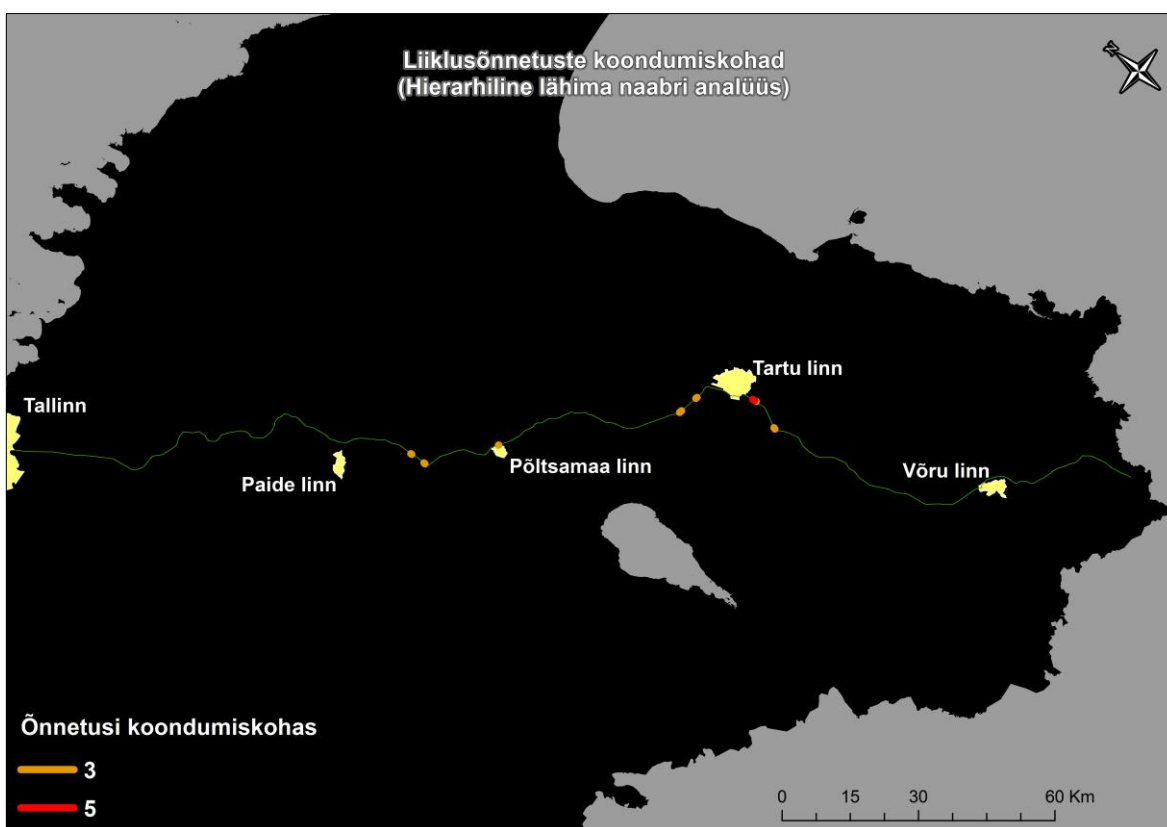
Harju maakonna riigiteede puhul leidis meetod hästi üles suuremad ohtlikud ristmikud. Sarnase tähelepaneku tegid ka Erdogan et al. (2015), kuid nende töös arvutati Getis-Ord statistik tasandilise KDE tulemuse põhjal. Optimeeritud tulipunktide analüüsi poolt automaatselt leitud analüüsiskaalat saab edukalt kasutada teiste meetodite otsinguraadiusena. Sellisel juhul vähendatakse analüüsija subjektiivset arvamust otsinguraadiuse kriteeriumi määramisel. Nii tehti ka selles magistritöös. Analüüsiskaala valimisel kasutab töövahend ruumilist astmelist autokorrelatsiooni ning selle leidmiseks globaalset Morani I statistikut (ArcGIS, 2016),

Kuna töövahendi tulemuseks on statistiliselt olulised kalavõrgu polügonid, mitte jooned, siis konkreetseid ohtlikke teelõike ei ole selle meetodiga kohe võimalik saada. Seetõttu tuleks huvipakkuvad segmendid eraldi GIS vahenditega joontele genereerida. Lisaks on leitud tulipunktid tihtipeale erinevad, kuna mõned neist on leitud koos ohutsoonidega ning teised ilma. Samuti on osad väiksemad koondumiskohad märgitud üksiku statistiliselt olulise polügonina õnnetuste raskuskeskmesse. Seetõttu on vajalik need manuaalselt üle vaadata ja teha otsus, millised kohad vastavad kasutaja poolt defineeritud koondumiskoha mõistele ning millised mitte.

Optimeeritud tulipunktide analüüsi meetodi üheks suureks plussiks on see, et seda on võimalik kasutada koos teiste koondumiskohtade leidmise meetoditega, nagu seda tegid Peet ja Padrik (2015) ning Nie et al. (2015). Sellisel juhul viiakse läbi teise meetodiga leitud väärtuste Hotspot-analüüs, et eristada statistiliselt olulised punktid.

3.3 Hierarhiline lähima naabri analüüs

Tallinn – Tartu – Võru – Luhamaa maanteel leidis HLNA 400-meetrise otsinguraadiusega kaheksa koondumiskohta (joonis 19). Üks neist oli viie õnnetusega ohtlik teelõik ning ülejäänud 7 kolme liiklusõnnetusega tulipunktid. Võrreldes optimeeritud tulipunktide analüüsiga leidis meetod kolm koondumiskohta vähem. Nendeks olid Tartu linna piirialal, Tallinna ja Paide vahel ning Paide ja Põltsamaa vahel paiknevad tulipunktid ehk need, kus optimeeritud tulipunktide analüüs leidis ühe statistiliselt olulise kalavõrgu polügoni. Kuigi nende koondumiskohtade kaks liiklusõnnetust paiknesid lähestikku, siis kolmas intsident toimus nendest rohkem kui 400 meetri kaugusel ning seetõttu ei mahtunud see otsinguraadiuse kriteeriumi sisse. Suurema otsinguraadiuse puhul oleks meetod ka need kolm koondumiskohta üles leidnud.

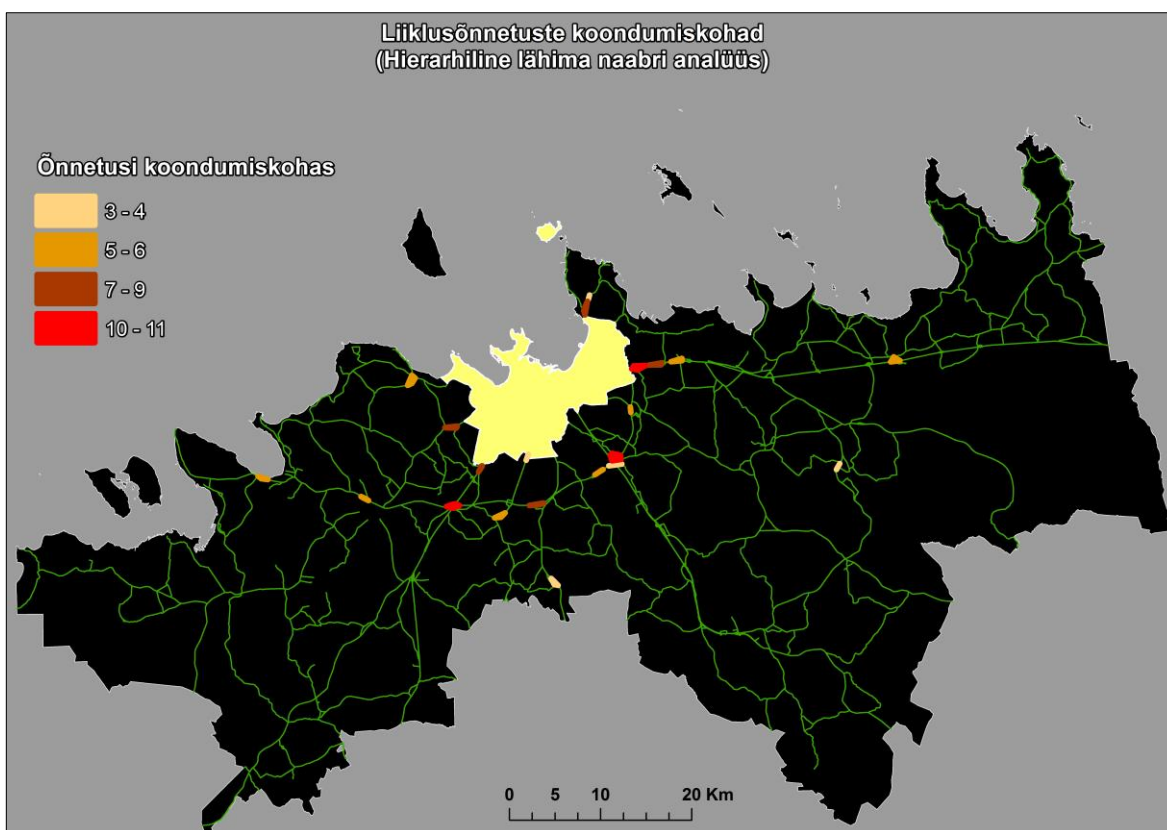


Joonis 19. HLNA leitud õnnetuste koondumiskohad (otsinguraadius 400 meetrit)

Monte Carlo simulatsiooni on üht kindlat maanteed analüüsides keeruline teostada. Kui lasta meetodil arvutada uuritava ala pindala ise (pindalast sõltub simulatsiooni tulemus), siis teeb ta seda sisendpunktide abil, kasutades nende miinimum- ja maksimumkoordinaate. Nende alusel leitakse moodustunud ristküliku pindala, mis enamasti on aga pindalalt suurem kui tegelik uuritav ala (Levine, 2013c). Rakendades Monte Carlo simulatsiooni sellel ristkülikul,

on tulemuseks, et kõik koondumiskohad on statistiliselt olulised. Valides aga näiteks optimeeritud tulipunktide analüüsis kasutatud 5-meetrise puhvertsooni pindala, on tulemuseks, et peaaegu kõik leitud klastrid on juhuslikud ning statistiliselt olulisi kobaraid ei ole. See on tingitud sellest, et meetod on arendatud töötama pinnal ning juhuslikud punktid genereeritakse kogu risküliku ulatuses. Antud juhul on tegu ainult joonobjektidel toimuvate intsidentidega.

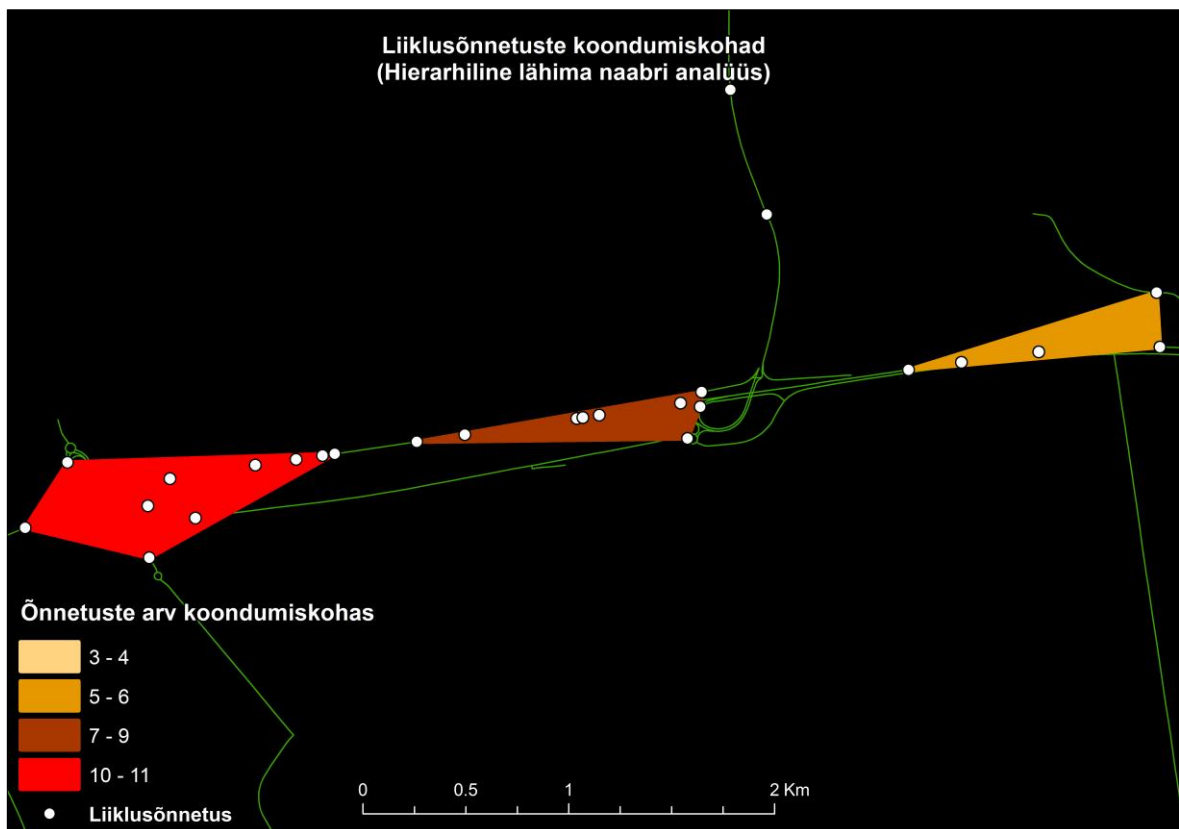
Harjumaal leidis HLNA 21 koondumiskohta, kasutades kriteeriumitena minimaalselt kolm õnnetusepunkti ning 1100-meetrise otsinguraadiust (joonis 20). Maksimalne õnnetuste arv ohtlikus kohas oli 11. Võrreldes kilomeetrise löikude meetodiga (22 koondumiskohta) paiknesid tulipunktid sarnaselt, kuid osa tulipunkti moodustavatest õnnetustest olid kahe meetodi puhul erinevad.



Joonis 20. HLNA leitud õnnetuste koondumiskohad (otsinguraadius 1100 m)

Väikeses mõõtkavas vaadatuna jääb mulje, et HLNA leitud koondumiskohad asuvad konkreetsetel teelõikudel, kuid tegelikkuses grupeeris meetod otsinguraadiuse, minimaalse punktide arvu ja olulisusenivoo alusel ka õnnetusi, mis vastavad küll nendele kriteeriumitele, kuid asuvad tegelikult hoopis teistel teelõikudel (joonis 21). Joonise 21 parempoolsema koondumiskoha puhul on grupeeritud ka põhjapoolsem õnnetus, mille kaugus sellele lähima

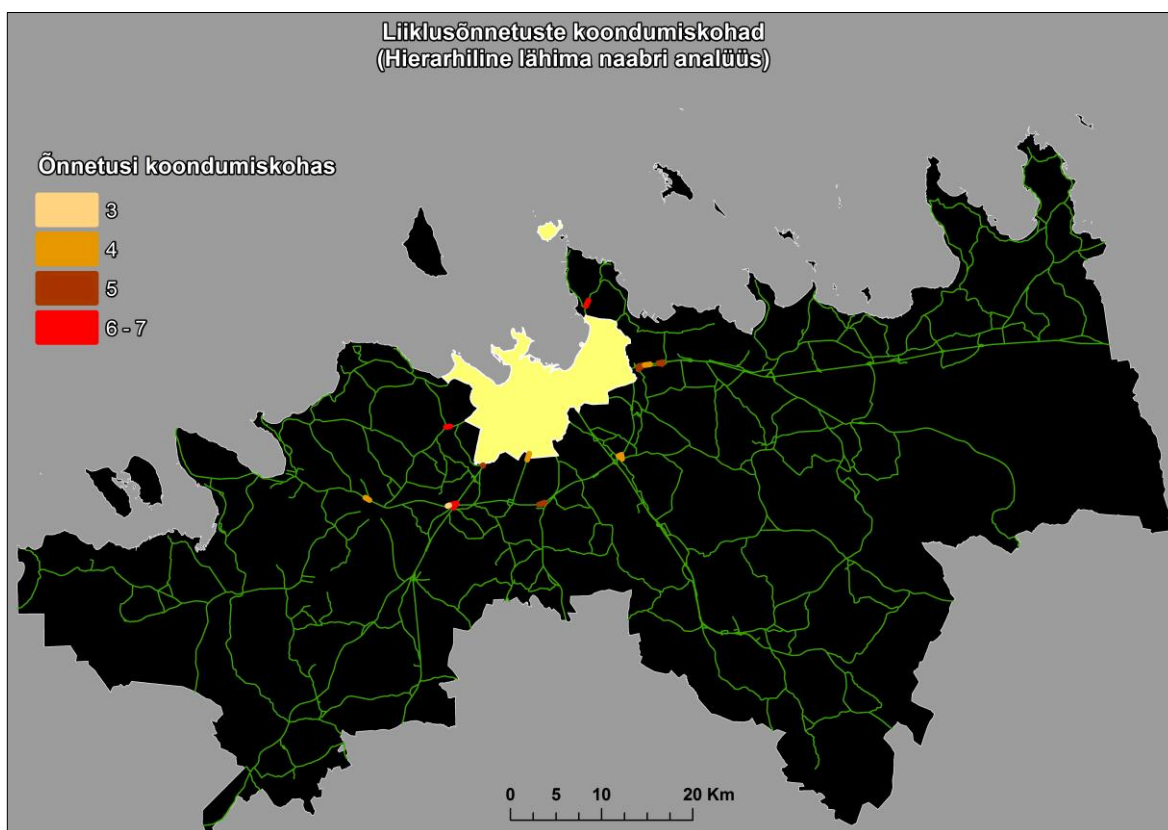
teise intsidendini mööda teed on ligikaudu 1700 meetrit. Manhattani kaugus selle õnnetuseni on kõigest 280 meetrit. Selline suurus on väiksem kui otsinguraadius ning seetõttu on see punkt ka grupeeritud. Sarnane olukord on ka keskmisel koondumiskohal (joonis 21), kus lääneosas on tegu keerukate teedestruktuuridega, kuhu kuuluvad ka ringteed ja viadukt. Kaugus mööda teed ühe õnnetuse juurest teiseni oleks olnud kordi suurem, kui 1100 meetrit, kuid Manhattani kaugus ühest punktist teiseni oli väiksem kui otsinguraadiuse kriteerium. Keskmise koondumiskoha läänepoolses osas on lausa kolm õnnetust, mis vastasid määratud kriteeriumitele, kuid ei asu tulipunkti põhiteljel.



Joonis 21. HLNA meetodi tulemused suurendatuna

Selliseid näiteid oli Harjumaal mitmeid, kus isegi paralleelselt kulgevatel teedel toimunud õnnetused grupeeriti klastrisse. Need grupid aga definitsiooni järgi koondumiskohad ei ole. Kuna võrgustikukaugust ei saanud HLNA puhul kasutada (põhjused on peatükis 2.2.4.1), siis Manhattani kaugus ei sobi hästi kasutamiseks teedel, mille läheduses võivad kulgeda ka teised teed. Leitud koondumiskohad võivad olla väga erinevate liiklustingimustega ning seetõttu ei tohiks neid ühtsete tulipunktidenähtude analüüsida (Moreira, 2011). Selline väide kehtib aga vaid suurte otsinguraadiuste puhul, nagu seda oli 1100 meetrit. Eukleidilise kauguse puhul oleks koondumiskohtadesse kaasatud õnnetuste arv olnud veelgi suurem.

Kasutades 500-meetrist otsinguraadiust oli tulemus mõnevõrra parem, kuid endiselt grupeeriti erinevate teede liiklusõnnetusi (joonis 22). Näiteks joonise 21 keskmine koondumiskoht sisaldas samu õnnetusi ka 500-meetrise otsinguraadiusega.



Joonis 22. HLNA tulemused Harjumaal (otsinguraadius 500 meetrit)

Ka Harjumaal koondumiskohtade statistilist olulisust hinnati Monte Carlo simulatsiooniga. Kui uuritava ala pindalaks määrati ümbritseva polügonina kasutatud 1-meetrise puhvri pindala, siis olid kõik leitud klastrid juhuslikud ning statistiliselt mitteolulised. Kasutades kogu Harjumaal pindala, leidis simulatsioon, et 21st klastrist ainult 2 (otsinguraadiusena kasutati 1100 meetrit) on juhuslikud ning ülejäänud on statistiliselt olulised. Kuna Monte Carlo on pindadele mõeldud meetod, siis paigutatakse punkte juhuslikult mitte teedele, vaid üle terve risküliku, millel on sama suur pindala kui Harju maakonnal.

Mõlema andmestiku puhul esines ka kilomeetrist lühemaid kolme õnnetusega teelõike, mida HLNA koondumiskohaks ei lugenud, kuigi otsinguraadiuse ja minimaalse punktide arvu kriteeriumile need lõigud vastasid. Siin toimis tõenäoliselt 95%line olulisusenivoo, mis määras tõenäosuseks 0,05, et kaks juhuslikult valitud punkti on üksteisele lähemal kui keskmine juhuslik kaugus (Levine, 2013c). Et seda tõestada, tehti katse, kus minimaalseks punktide arvuks oli 2, otsinguraadiuseks 1100 meetrit ning usaldusnivooks 95%. Sellisel

juhul oli tulemuseks 45 klastrit, millest 4 sisaldasid kaks õnnetusepunkti. Kui enne oli kolme õnnetusega klastreid ainult kaks, siis uue katsetuse puhul oli neid koguni 20. Kuigi Monte Carlo simulatsioon sellisel kujul, nagu ta on CrimeStatis, ei sobi liiklusõnnetuste koondumiskohtade statistiliseks kontrollimiseks, siis grupeeritakse punktid klastritesse ikkagi statistilisel alusel, kasutades olulisuse nivood.

Meetodi väljundiks on koondumiskohta moodustavate õnnetusepunktide kumer kate (*convex hull*), mis märgib konkreetset tulipunkti ilma ümbritseva ohutsoonita. See tähendab, et kõrgema riskiga ala piirneb koondumiskohta moodustavate äärmiste liiklusõnnetustega. Ehk teisisõnu ei kaasne tulemustega võimalikku ohutsooni nagu teiste meetodite puhul. Kuna kumer kate on pindobjekt, siis mõistlik oleks koondumiskohad joonobjektideks genereerida. Joonobjektidega on lihtsam läbi viia edasist analüüsi, et selgitada välja, miks just nendel teelõikudel on õnnetuste arv keskmisest suurem. Sellisel juhul on lihtsam seletada ka tulipunktide asukohta, näiteks Tallinn – Tartu – Võru – Luhamaa maantee esimene koondumiskoht algab 5,5. ja lõpeb 6,3. kilomeetril.

HLNA puhul sõltub tulemuste usaldusväärsus ja vastavus koondumiskoha definitsioonile väga palju otsinguraadiuse suuruselt. Tallinn-Tartu-Võru-Luhamaal jäi leidmata vähemalt 3 tulipunkti, kus juhtus minimaalselt kolm õnnetust vähem kui kilomeetri pikkusel teelõigul. Samas ei ole otstarbekas kasutada liiga suurt otsinguraadiust, kuna õnnetused võisid toimuda üksteisest näiteks 500 meetri kaugusel ning üle 500 meetri pikkust otsinguraadiust kasutades oleks tulemuseks teelõik pikkusega 1500 meetrit, kus juhtus kolme aasta jooksul 3 õnnetust. Sellisel juhul ei ole see teelõik selle töö definitsiooni järgi koondumiskoht. Kaugust õnnetuste vahel tuleks võimaluse korral määrata mööda teed liikudes. Ka Moreira (2011) sõnul võib suurt otsinguraadiust kasutades saada tulemuseks väga pikki teesegmente. Seetõttu tuleb andmestikku tundma õppida ning otsinguraadiust korrigeerida vastavalt analüüsi eesmärgile (Smith ja Bruce 2008). Kasutades suurt otsinguraadiust ja väikest minimaalset punktide arvu, on tulemuseks suurem juhuslikult leitud koondumiskohtade arv (Eilat, 2011).

Harju maakonna liiklusõnnetuste puhul oli HLNA suureks puuduseks võrgustikukauguse mitte töötamine (Levine, 2016; peatükk 2.2.4.1). Selle asemel tuli kasutada Manhattani kaugust. Võrgustikukaugus töötas vaid väikese otsinguraadiusega, mille maksimumväärtus, millega tulemusi saadi, oli Harjumaa puhul 250 meetrit. Maksimaalne otsinguraadius, millega tulemused saadakse, sõltub õnnetuspunktide arvust ning teedevõrgu keerukusest ja pikkusest.

Kasutades näiteks 250-meetrist otsinguraadiust, on võimalik võrgustikukaugusega leida eriti ohtlikud teelõigud, kus vähem kui 750 meetri pikkusel segmendil on toimunud minimaalselt 3 õnnetust. Selline katse ka tehti ning 95% ja 50% olulisuse nivood kasutades leiti vastavalt 7 ja 10 koondumiskohta. Üks 10st koondumiskohast leiti sarnaselt joonise 21 keskel kujutatud tulipunktiga. Selle alusel võib väita, et kaugus mööda teedevõrgustikku, kas teede keerukuse või muu põhjuse tõttu, korrektselt ei tööta. Sama mainis ka tarkvaralooja Ned Levine isiklikus kirjavahetuses (Levine, 2016) ning soovitas selle asemel kasutada eukleidilist või Manhattani kaugust. Moreira (2011) märkis oma töös HLNA põhiliseks miinuseks selle, et eukleidiline kaugus ei anna korrektseid tulemusi paralleelsete teede ja ristmike puhul. Viidete järgi võiks järeldada, et ta kasutas oma töös CrimeStati III versiooni, kus võrgustikukauguse funktsioon on juba olemas. Seetõttu jääb arusaamatuks, miks Moreira (2011) seda ei kasutatud. Käesoleva töö tulemusena selgus, et võrgustikukaugust saab kasutada vaid teatud tingimusteni ning ka sellisel juhul ei tööta see täiesti korrektselt.

CrimeStat tarkvara juhend ei soovita kasutada võrgustikukaugust põhjusel, et tihtipeale on teede andmed ja ruumikujud vigased ning arvutused võtavad mitu korda kauem aega võrreldes eukleidilise või Manhattani kaugusega (Levine, 2013b). Väidetakse ka, et sellest on saadav kasu võrreldes teiste tüüpidega (eukleidiline, Manhattani) vähene. Selle töö põhjal nii väita ei saa. Kui võrgustikukaugus töötaks korrektselt ning võimaldaks kasutada parameetreid olenemata nende suurusest, siis oleksid ka tulemused oluliselt täpsemad. Samas ei saa välistada, et teistsuguste analüüside puhul saadakse arvestatavaid tulemusi ka eukleidilist ja Manhattani kaugust kasutades. Nendes uurimustes võivad püstitatud uurimisküsimused ning koondumiskoha definitsioonid olla teistsugused.

Manhattani kauguse tõttu grupeeris HLNA Harju maakonnas ka naaberteedel toimunud õnnetusi, mis jäid küll otsinguraadiuse sisse, kuid toimusid hoopis teiste liiklustingimustega teelõikudel. Seetõttu ei ole soovituslik seda meetodit rakendada kogu teedeandmestiku peal. Üht konkreetset teed analüüsidest annab meetod usaldusväärsed tulemused ka eukleidilise ja Manhattani kaugusetüübiga.

3.4 KDE (*Kernel Density Estimation*)

KDE arvutas Tallinn – Tartu – Võru – Luhamaa teelõikude keskmiseks väärtuseks 0,23. Standardhälve oli 0,47. Sellest tulenevalt oli klassi väärtus, millest loeti lõiku koondumiskohaks, 1,4 ($0,23+0,47 \times 2,5$).

KDE leidis näitlikult kõige rohkem koondumiskohti (joonis 23). Lisaks eelmiste meetodite klastritele leiti lisaks veel üheksa ohtlikku kohta. Kahes leitud tulipunktis juhtus kolme aasta jooksul vaid üks õnnetus. Teistes eelmiste meetodite poolt leidmata kobarates toimus kaks õnnetust, kuigi need teelõigud olid keskmiselt 300 meetrit pikad. Need üheksa tulipunkti ei ole aga definitsiooni järgi liiklusõnnetuste koondumiskohad.



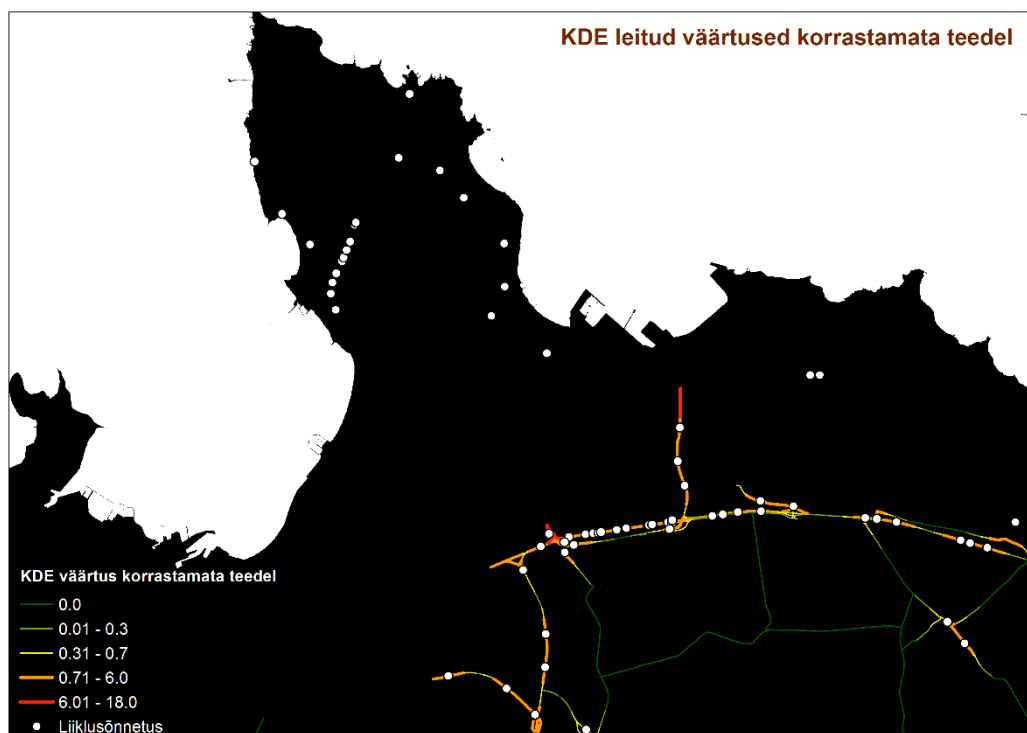
Joonis 23. KDE meetodiga leitud õnnetuste koondumiskohad (otsinguraadius 400 ning lõigu pikkus 40 meetrit)

Harjumaa katsetuste käigus selgusid ka mõningad probleemid seoses KDE meetodi ja Teeregistri andmestikuga. Kuna Teeregistri väljavõttes olid olemas ainult riigimaanteedel toimunud õnnetused ning sellest tulenevalt ruumikujudena ka ainult riigimaanteed, siis asula- ja kohalikud teed olid puudu. Sellest tingituna ei olnud teedevõrk sidus ning kõik teed ei olnud omavahel korrektselt ühendatud. Näiteks Viimsi poolsaare riigiteed Viimsi-Rohuneeme ning Viimsi-Randvere ei ole ühenduses Tallinn-Narva maanteega.

SANETi KDE meetodi suureks plussiks on see, et enne algoritmi käivitamist kontrollitakse, kas teede sisendandmestik on lõike, mis ei ole põhivõrgustikuga ühendatud. Kui esineb ühendamata teid, küsib töövahend, kas jätkata algoritmiga või lõpetada protsess. Lõpetamisel kirjutab meetod uue *shp* formaadis faili nimega *WARNING*, kus kindla

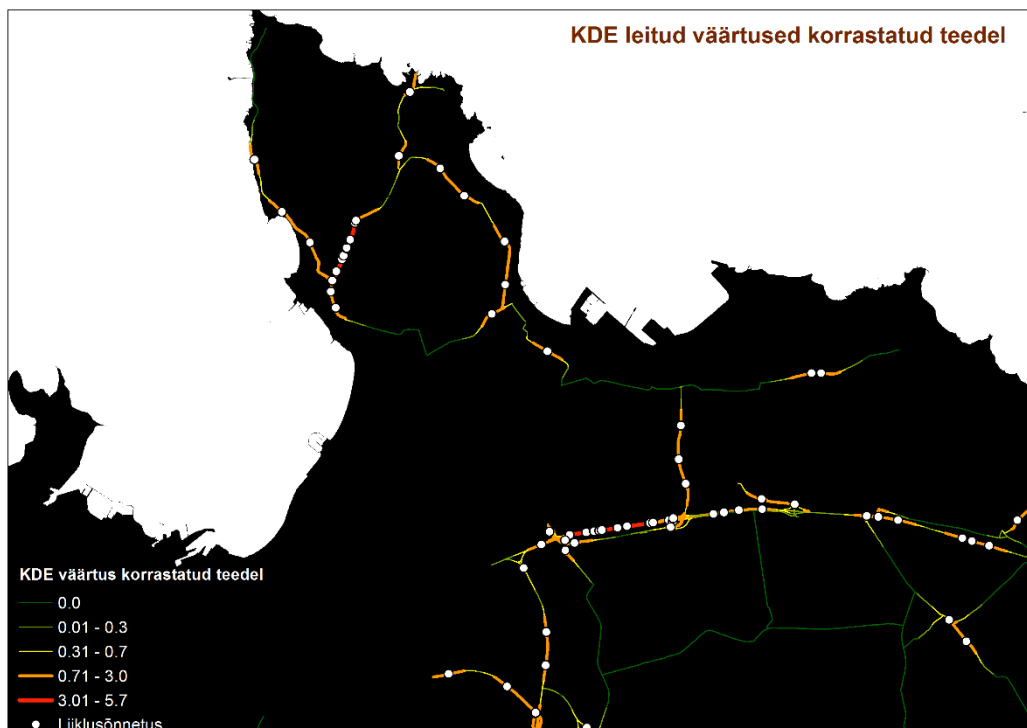
atribuudi järgi on võimalik leida teelõigud, mis on põhivõrgustikuga ühendamata. Kui jätkata algoritmiga, siis valib meetod suurima ühendatud võrgustiku ning teeb vajalikud arvutused.

Katsetades selgus, et põhivõrgustikuga ühendamata teelõigud kaotas KDE ära. Kaotatud lõikudel olnud õnnetused andsid lähimatele allesjäänud teesegmentidele põhjendamatult suure väärtuse (joonis 24). Muuga sadama tee, mis asub joonise 24 idaosas ning on väärtusega >6 , sai põhjapoolsemas osas kogu Harjumaa suurima KDE väärtuse, milleks oli 17,98. Loo tee, mis asub joonise 24 lääneosas ning on märgitud punase joonleppemärgiga, sai põhjapoolsemas osas väärtuse 9,4. Viimsi poolsaarest eemaldudes ehk lõuna suunas kahanesid mõlema suure kaaluga lõigu väärtused, kuid jäid sellegipoolest väga suureks kui võrrelda tulemust teiste lõikudega. Katsetustest võib järeldada, et meetodi korrektseks toimimiseks peab kogu võrgustik olema omavahel korrektselt ühendatud.



Joonis 24. KDE katsetus korrastamata teedel 1000-meetrise otsinguraadiuse ja 250 meetri pikkuse teelõiguga

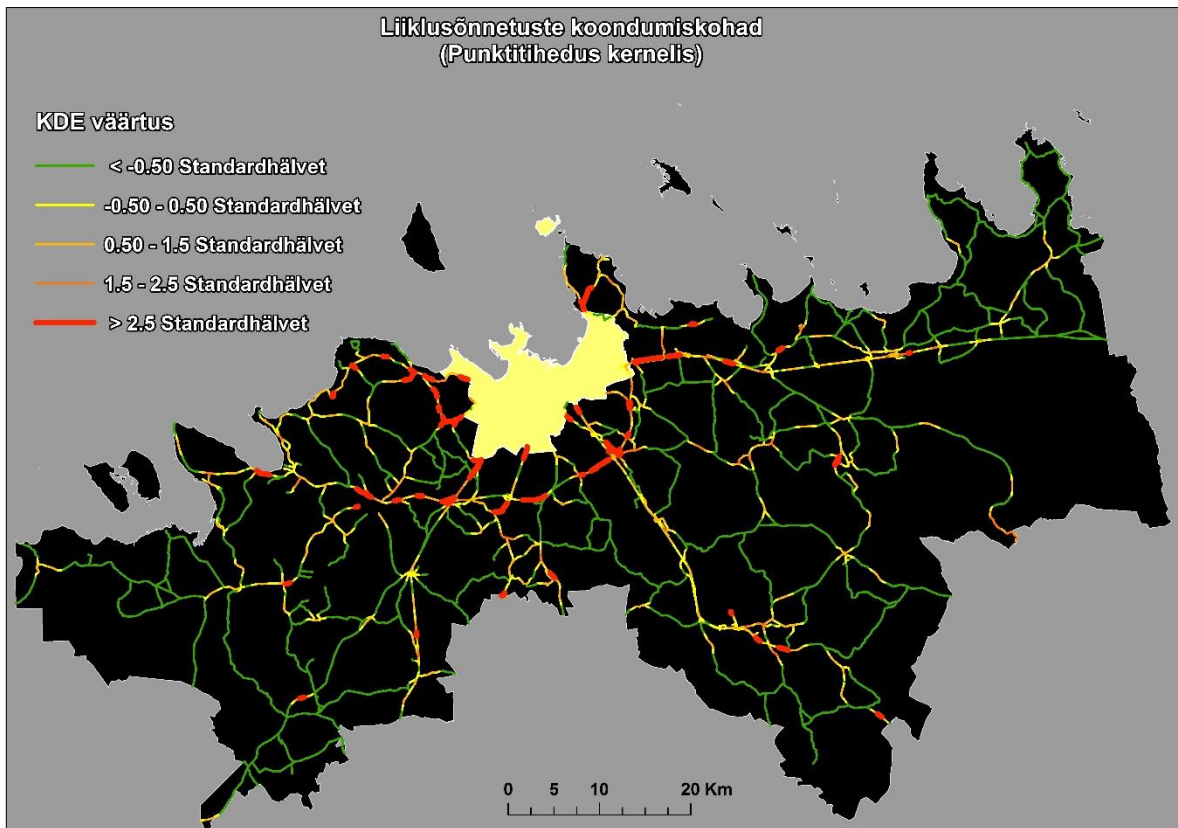
Eesti Topograafilise Andmekogu (ETAK) andmete abil ühendati maanteed omavahel sidusaks teedevõrguks seal, kus SANETi töövahend leidis katkestusi. Peale paranduste tegemist leiti KDE väärtus korrektselt (joonis 25).



Joonis 25. KDE katsetus korrastatud teedel 1000-meetrise otsinguraadiuse ja 250 meetri pikkuse teelõiguga

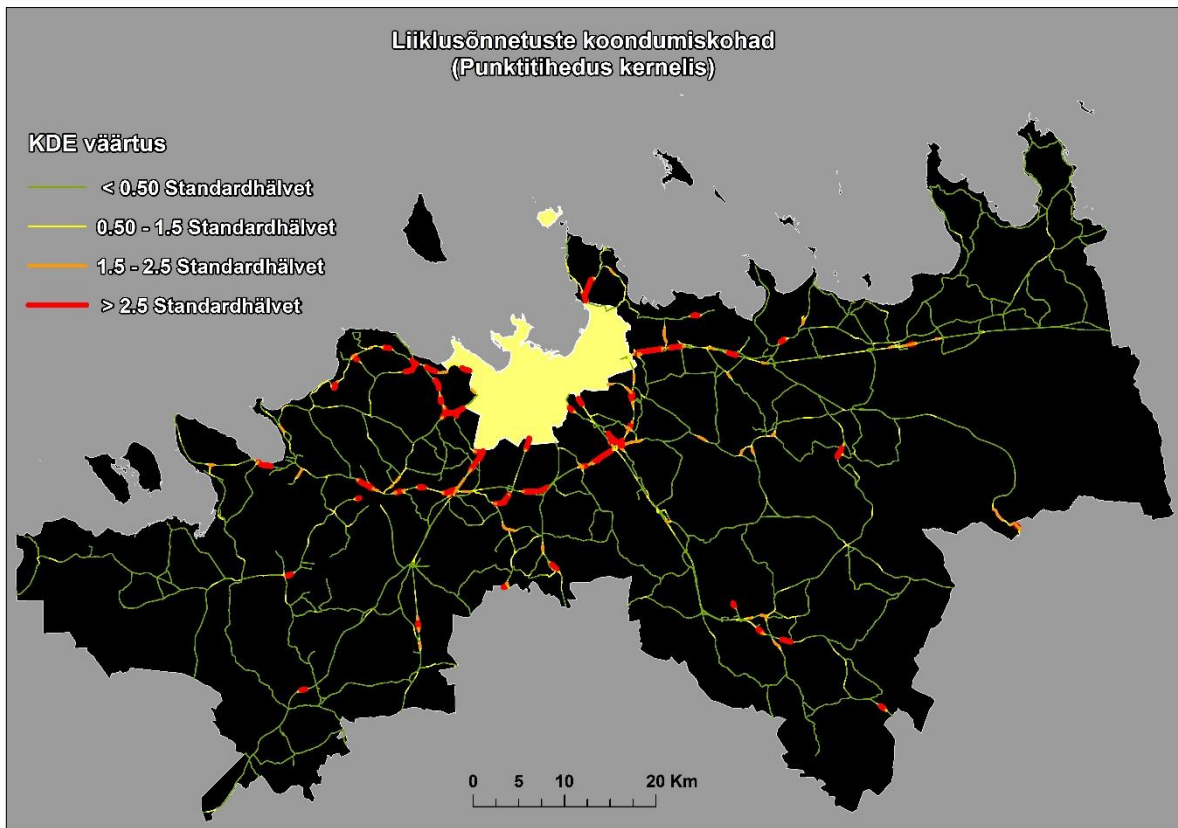
Tallinn – Tartu – Võru – Luhamaa maanteel töötas meetod korrektselt, kuna teelõigud olid omavahel korrektselt snäpitud ning lõikude vahel katkestusi ei esinenud. Sellest tulenevalt ei olnud ka parandused vajalikud.

Sidusat teedevõrku kasutades arvutas KDE Harjumaa lõikude keskmiseks väärtuseks 0,3. Standardhälve oli 0,6. Sellest tulenevalt oli klassi väärtus, millest loeti lõiku koondumiskohaks, 1,8 ($0,3+0,6\times 2,5$). Nagu Tallinn – Tartu – Võru – Luhamaa maantee puhulgi, leidis KDE Harju maakonnas teiste meetoditega võrreldes kõige rohkem koondumiskohti (joonis 26). Eelnevate meetodite poolt leidmata tulipunktid koosnesid enamasti ühest või kahest õnnetusest. Definitsioonile vastavad koondumiskohad leiti koos ohutsoonidega ehk kuni paarsada meetrit enne esimest ja peale viimast tulipunkti moodustavat õnnetust. Kõige rohkem tulipunkte oli Tallinna ringteel.



Joonis 26. KDE leitud õnnetuste koondumiskohad (otsinguraadius 1100 meetrit, lõigu pikkus 100 meetrit)

Et analüüsida lõigu pikkuse mõju tulemustele, tehti katsetus sama otsinguraadiuse, kuid lõiguga, mille pikkus on 500 meetrit (joonis 27). Koondumiskohtade arv ja asukoht jäid samaks, mistõttu kinnitab ka see töö Xie ja Yani (2008) väidet, et lõigu pikkus avaldab tulemustele väikest mõju.



Joonis 27. KDE leitud õnnetuste koondumiskohad (otsinguraadius 1100 meetrit, lõigu pikkus 500 meetrit)

Silumislaiuse mõju hindamiseks tehti katsetus ka 500-meetrise otsinguraadiusega. Selle tulemuseks, kasutades standardhälbe klassifikatsiooni, oli veelgi rohkem nn koondumiskohti, mis küll definitsiooni kohaselt sellele ei vasta. Rohkem leiti ühe õnnetusega lõike, mis näitab, et väiksemat otsinguraadiust kasutades on tulemused detailsemad. Selle töö analüüsi eesmärgist lähtuvalt need detailid kasutegurit ei oma, vaid suurendavad töömahtu, kus kõigist leitud ohtlikest lõikudest tuleb välja valida need, mis vastavad töös püstitatud koondumiskoha definitsioonile.

Kui välja arvata kasutatud teedevõrgustiku iseärasused (ainult riigimaanteed), siis SANETi KDE on lihtsasti kasutatav. Meetod eeldab, et kogu teedevõrgustik on omavahel korrektselt snäpitud. Vastasel juhul on tulemused moonutatud. Kui uurida kogu Eesti põhimaanteed liiklusõnnetuste koondumiskohti, siis tuleb meeles pidada seda, et Saaremaa ja Muhu saare tee nr 10 (Risti – Virtsu – Kuivastu – Kuressaare) on vaja praamiliini abil ühendada sama teega mandril.

Kuna meetod on välja töötatud spetsiaalselt joontele ning kaugusena kasutatakse kaugus mööda teed, siis olid ootused selle meetodi suhtes kõige suuremad. Mõlema andmestiku puhul leidis meetod kõige rohkem koondumiskohti, kuid osa neist olid ühe ja kahe õnnetusega tulipunktid. Kahe õnnetusega lõigud said tulipunkt staatuse, kuna toimunud õnnetused asusid üksteisele väga lähedal ning sellest tulenevalt arvutas KDE nendele lõikudele suured väärtused.

Palju esines sarnast nähtust, mis optimeeritud tulipunktide analüüsi puhulgi, kus ohtlik lõik asub näitlikult mitme õnnetuse vahel, kuid õnnetused ise koondumiskohta ei kuulu. Lisaks esines olukordi, kus tulipunkt leiti hoopiski enne või pärast õnnetusi. Selliste nähtuste üheks põhjuseks on meetodi arvutusalgoritm. Õnnetuste vahel asuvatele lõikudele arvutatakse suurem KDE väärtus siis, kui punktid paiknevad suhteliselt hajusalt ning otsinguraadiuse sisse jääb näiteks kaks õnnetusepunkti. Kui aga arvutada väärtust õnnetuse all olevale teelõigule, siis ei jää teine õnnetus otsinguraadiuse sisse. Põhjus võib olla ka selles, et väärtuse määramiseks, millest alates on tegu koondumiskohaga (*cutoff value*), kasutati ArcGISi standardhälbe klassifikatsiooni. Prooviti ka intervallide ja kvantiilide meetodeid, kuid sellisel juhul muutus koondumiskohtade arv veelgi suuremaks või hoopis liiga väikeseks. Standardhälbe klassifikatsioon andis kõige mõistlikumad tulemused.

Käesolev töö kinnitab Andersoni (2009) väidet, et KDE meetod omab teiste meetodite, kaasa arvatud ka siin töös kasutatud meetodite, ees eeliseid. Võimalik on leida konkreetsed ohtlikud teelõigud, nende ohutsoonid ja ka ohtlikud ristmikud. Suur eelis selles töös kasutatud KDE meetodil on see, et kaugusena kasutatakse võrgustikukaugust. Kahjuks ei ole Peet ja Padriku (2015) töös kirjas, millised on ArcGISi *Kernel Density* meetodi plussid või puudujäägid, kuid kindel on see, et võrgustikukaugust ArcGISis selle meetodiga kasutada ei saa. KDE väärtuste ülehindamise vältimiseks analüüsisid Peet ja Padrik (2015) iga maanteed eraldi. Selles töös kasutatud KDE arvutab korrektsed väärtused ka ristmikule ja paralleelselt kulgevatele teedele. Seetõttu on võimalik sisendandmestikuna kasutada kõiki maanteid ja nendel toimunud õnnetusi korraga.

Käesolevas töös analüüsiti kogu Harjumaa andmestikku korraga ning arvutusteks kulunud aeg oli suhteliselt väike (~3 minutit). Kuna käesolevas töös kasutati sülearvutit (lisa 1) ja arvutusteks kulunud aeg ei olnud suur, siis ei tohiks KDE hätta jääda ka mahukamate andmete puhul. Tarkvara autorid (Okabe et al., 2008) nendivad, et SANETi KDE üheks piiravaks teguriks mahuka andmestiku puhul on arvutuste kiirus. Seetõttu soovivad nad kasutada *equal-split discontinuous* kerneli funktsiooni. Selles töö andmestiku puhul see aga

piirav tegur ei olnud ja tulemused saadi mõistliku aja jooksul ka *equal-split continuous* funktsiooniga.

KDE puhul ei ole ohtlike kohtade joonobjektideks genereerimine hädavajalik, kuna tegu juba on teejoontega. Võimalik on teelõigud, mille standardhälve on suurem kui $>2,5$, eraldi välja eksportida ning seejärel üksteisega lõikuvad segmendid üheks tervikuks kokku liita. Viimase sammuna on vaja üle vaadata, millised leitud lõigud vastavad koondumiskoha definitsioonile. Tulemustest on võimalik eemaldada lõigud, kus vaadeldava ajaperioodi jooksul on juhtunud vaid üks või kaks õnnetust.

Kuna KDE kasutab kauguse tüübina võrgustikukaugust, siis on selle meetodi poolt leitud tulemused võrreldes optimeeritud tulipunktide analüüsi ja HLNaga kõige täpsemad ja usaldusväärsemad. Katsetuste tulemusena selgus, et lõigu pikkus mõjutab tulemusi oluliselt vähem kui otsinguraadius (Xie ja Yan, 2008; Anderson, 2009). Kasutades aga väiksemat otsinguraadiust (selles töös võrreldi 1100- ja 500-meetrist otsinguraadiust), on tulemused detailsemad ehk koondumiskohtadena leitakse veelgi rohkem ühe ja kahe õnnetusega teelõike. Selle töö seisukohalt on need aga mitteolulised. Seetõttu kinnitab ka see töö Xie ja Jani (2008) väidet, et väikese silumislaiusega leitakse lokaalsed erinevused, kuid suurt otsinguraadiust kasutades kerkivad esile eelkõige suure väärtusega segmendid. Optimaalset otsinguraadiust ja lõigu pikkust ei ole võimalik määrata (Xie ja Jan, 2008; Tang et al., 2015), kuna andmestikud võivad olla erinevad. Sellest tulenevalt sobivad eri andmestikele erinevad optimaalsed väärtused. Väiksemad lõigud ja otsinguraadiused võiksid sobida rohkem linnadele, kuid maanteede uurimisel tuleks kasutada suuremaid väärtusi.

KDE (SANETi tarkvaras) suureks miinuseks on see, et tulemusi ei saa statistiliselt kontrollida. Ehk kasutaja ei tea, millised meetodi poolt arvutatud väärtused on statistiliselt olulised. Sama puudjääki KDE puhul on oma töödes tõdenud ka Xie ja Jan (2008), Erdogan et al. (2015), Bills (2009) ja Nie et al. (2015). Selles töös kasutati standardhälbe klassifikatsiooni, kus KDE tulemused klassifitseeriti keskväärtuse ja standardhälbe summa alusel klassidesse. Koondumiskohaks valiti viimane klass. Selliselt püüti vähendada kasutaja subjektiivsust läviväärtuse valikul. Ka Getis-Ord statistik on sisuliselt standardhälve (Peet ja Padrik, 2015). Bil et al. (2013, 2015) on arendanud KDE+ tarkvara, mille abil on võimalik arvutatud KDE tulemust statistiliselt kontrollida juba tarkvarasiseselt. KDE+ meetodit tahtsid algselt oma uurimuses kasutada ka Peet ja Padrik (2015), kuid sellel oli mitmeid puudusi (põhjused peatükis 1.2.2).

3.5 Meetodite võrdlus

Kokkuvõttes on igal meetodil omad head ja halvad küljed, kuid päris ideaalseid tulemusi ei andnud neist ükski. Punktitihedus kernelis, optimeeritud tulipunktide analüüs ning kilomeetriste lõikude meetodid annavad enamasti tulemuseks ka teatud kõrgema riskiga alad enne ja pärast konkreetset koondumiskohta. Kui need alad oleksid alati leitavad kindla reegli järgi, näiteks 200 meetrit enne ja pärast koondumiskohta, siis annaksid need kindlasti analüüsile lisaväärtust juurde. Paraku ei saa eelmistes peatükkides olevate näidete põhjal alati kindel olla, mis suunas ja kui kaugele need ohtlikud alad ulatuvad. Antud juhul oleks KDE puhul abiks statistilise olulisuse kontroll.

Optimeeritud tulipunktide analüüsi tulemustes oli tulipunktina määratud kohti, kus ei toimunud ühtegi õnnetust ning lähim intsident sellele kohale oli ligikaudu kolme kilomeetri kaugusel, kui mõõta kaugust mööda teid. See tuleneb sellest, et meetod on välja töötatud nende punktide analüüsimiseks, mis saavad toimuda uuritava alal igal pool, mitte ainult joontel. Seepärast leitakse osa statistiliselt olulistest polügonidest õnnetuste raskuskeskmesse.

KDE ja optimeeritud tulipunktide analüüsi puhul oli olukordi, kus koondumiskohana oli defineeritud hoopis õnnetuste vahel olev teelõik. Seetõttu tuleb kõik leitud koondumiskohad visuaalselt üle vaadata ja leida sellised, mis vastavad püstitatud eesmärgile. Eelnevalt leitud alad ei peagi tegelikult vastama selles töös kasutatud liiklusõnnetuste koondumiskoha definitsioonile. Näiteks optimeeritud tulipunktide analüüs leiabki üles alad, kus on juhtunud konkreetse andmestiku puhul keskmisest rohkem õnnetusi.

Hierarhiline lähima naabri analüüs töötas Tallinn – Tartu – Võru – Luhamaa maanteel väga hästi, vaatamata isegi sellele, et kauguse tüübina oli kasutusel Manhattani, mitte võrgustikukaugus. HLNA leiab konkreetseid koondumiskohad, ilma ohtlike tsoonideta, ning nii võib kindel olla, et tulipunktid leitakse alati kindla reegli alusel. Teiste meetodite puhul, nagu selle peatüki eelmistes lõikudes kirjeldatud, konkreetset ühtemoodi toimivat reeglit ei ole. HLNA plussiks on see, et võimalik määrata usaldusnivood, et kaks juhuslikult valitud punkti on üksteisele lähemal kui keskmine juhuslik kaugus (Levine 2013b), mis tagab selle, et punktid grupeeritakse klastritesse ikkagi statistilisel alusel.

Harjumaa andmestiku puhul HLNA hästi ei töötanud, kuna otsinguraadiuse kriteeriumile vastasid tihti peale ka naaberteedel toimunud õnnetused. Ka võrgustikukaugus väikese otsinguraadiusega ei toiminud korrektselt. Endiselt grupeeriti õnnetusi, mis toimusid

erinevatel teedel, kuigi intsidentide vahemaa mööda teed oli suurem kui otsinguraadius. Seetõttu tasuks HLNA meetodit kasutada üksikute maanteede või nendega ristuvate teede puhul. Paralleelselt kulgevate teede kaasamine toob kaasa väärad tulemused.

Kõige vähem koondumiskohti leidis kilomeetriste lõikude meetod, kuna segmenteerimise metoodikast tulenevalt võisid õnnetused sattuda ühe segmendi lõppu ja teise algusesse. Samas on selle meetodi puhul kindel, et kusagil leitud lõigul on definitsioonile vastav koondumiskoht olemas. Olgu õnnetused lõigu alguses, lõpus või paiknevad nad hajusamalt kogu kilomeetrise segmendi ulatuses. Kilomeetriste lõikude meetodi veel üheks puuduseks on see, et tulemused ei ole statistiliselt kontrollitud. Meetodi plussiks on see, et segmenteerimisel kasutatakse kaugust mööda teed.

Ka HLNA puhul on konkreetne õnnetuste arv koondumiskohas kindel. Optimeeritud tulipunktide analüüsi ja KDE puhul aga ei saa koheselt öelda, et ühes või teises koondumiskohas on toimunud arvuliselt konkreetne arv õnnetusi. Need meetodid annavad tulemuseks ohtlikud alad, kuid õnnetuste arv nendel aladel tuleb kindlaks teha kasutajal.

3.6 Järeldused

Magistritöö eesmärgiks oli välja selgitada, kui hästi kogu pinnal töötavaid meetodeid on võimalik korrektselt tööle saada ka sündmuste jaoks, mis saavad toimuda vaid kitsamates kanalites. Üldise ülevaate tulipunktide asukohtadest annavad kasutajale kõik selles töös kasutatud meetodid.

Optimeeritud tulipunktide analüüs on ümbritseva polügoni funktsiooniga kasutatav teatud mööndustega. Ilma selle funktsioonita on tulemused kasutuskõlbmatud. Ümbritseva polügonina tasuks kasutada võimalikult väikese suurusega puhvrit, mis tagab täpsemad tulemused. Optimeeritud tulipunktide analüüsi ei saa väga hästi kasutada ühe konkreetse maantee analüüsimiseks, sest tihtipeale tööriist tulemusteni ei jõua ja tagastab veakoodi, kuigi valib ise kõik parameetrid analüüsi läbiviimiseks. Seetõttu on selle meetodiga otstarbekas analüüsida teedevõrgustikke, mitte üksikuid maanteid. Probleeme võib tulla ka mitme maantee uurimisel, mis ei paikne üksteise lähedal, näiteks Eesti põhimaanteed. Kõige olulisem selle meetodi puhul oleks aga funktsioon, mis võimaldaks kasutada võrgustikukaugust. Sellisel juhul oleksid liiklusõnnetuste koondumiskohtade asukohad veelgi täpsemad. Nie et al. (2015) kasutasid oma töös võrgustikule kohandatud Getis-Ord

statistikut. Selliselt kohandatud statistiku kasutamine optimeeritud tulipunktide analüüsis oleks väga oluline.

Hierarhilist lähima naabri analüüsi on võimalik kasutada just ühe maantee uurimisel, sest sellisel juhul ei ole võimalik grupeerida paralleelselt kulgevate teede õnnetusi. Kuigi tarkvara võimaldab kasutada ka võrgustikukaugust, siis korrektselt see ei toiminud ning parameetrite väärtusi sai tõsta kuni teatud suuruseni. Kui võrgustikukaugus töötaks korrektselt, võimaldaks see korraga analüüsida ka kogu andmestikku. Statistilise olulisuse kontroll ehk Monte Carlo simulatsioon on HLNA puhul hetkel kasutatav vaid tasandil. Koondumiskohtade statistilise olulisuse kontrollimiseks võiks edukalt sobida ka eelnevalt mainitud simulatsioon, kuid see peaks olema kohandatud võrgustikule ja kasutama võrgustikukaugust. Sellisel juhul genereeritakse punktid ainult teedele ja leitakse juhuslikud klastrid nende punktide alusel.

KDE võimaldab analüüsida kogu andmestikku korraga. Kuigi teedel toimuvate intsidentide analüüsiks on kõik vajalikud eeldused olemas, siis SANETi tarkvara KDE suurimaks puuduseks on statistilise olulisuse kontrolli puudumine. Viimastel aastatel on püütud seda puudust kõrvaldada. Bil et al. (2013, 2015) on arendanud KDE+ tarkvara, mille abil on võimalik arvatud KDE tulemust statistiliselt kontrollida juba tarkvarasiseselt. Mõistlik on KDE tulemust kombineerida ka teiste meetoditega, mis arvutavad lokaalstatistikuid. Morani I lokaalstatistikut kasutasid näiteks Xie ja Jan (2013) ning Getis-Ord statistikut Nie et al. (2015) ning Peet ja Padrik (2015).

3.7 Eesti põhimaanteedel 2013–2015 toimunud õnnetuste koondumiskohad

3.7.1 Metoodika

Kuna selles töös kasutatud meetoditest oli parim KDE, siis kasutati põhimaanteedel toimunud õnnetuste koondumiskohtade leidmiseks just seda meetodit. Nagu eelnevalt kirjeldatud, vajab KDE korrektseks toimimiseks sidusat teedevõrgustikku. Seetõttu kasutati taaskord ETAKi teid, mille abil ühendati riigimaanteed sidusaks võrgustikuks kohtades, kus meetod leidis katkestusi. Andmestikuna kasutati ajavahemikul 01.01.2013 kuni 31.12.2015 riigimaanteedel toimunud õnnetusi.

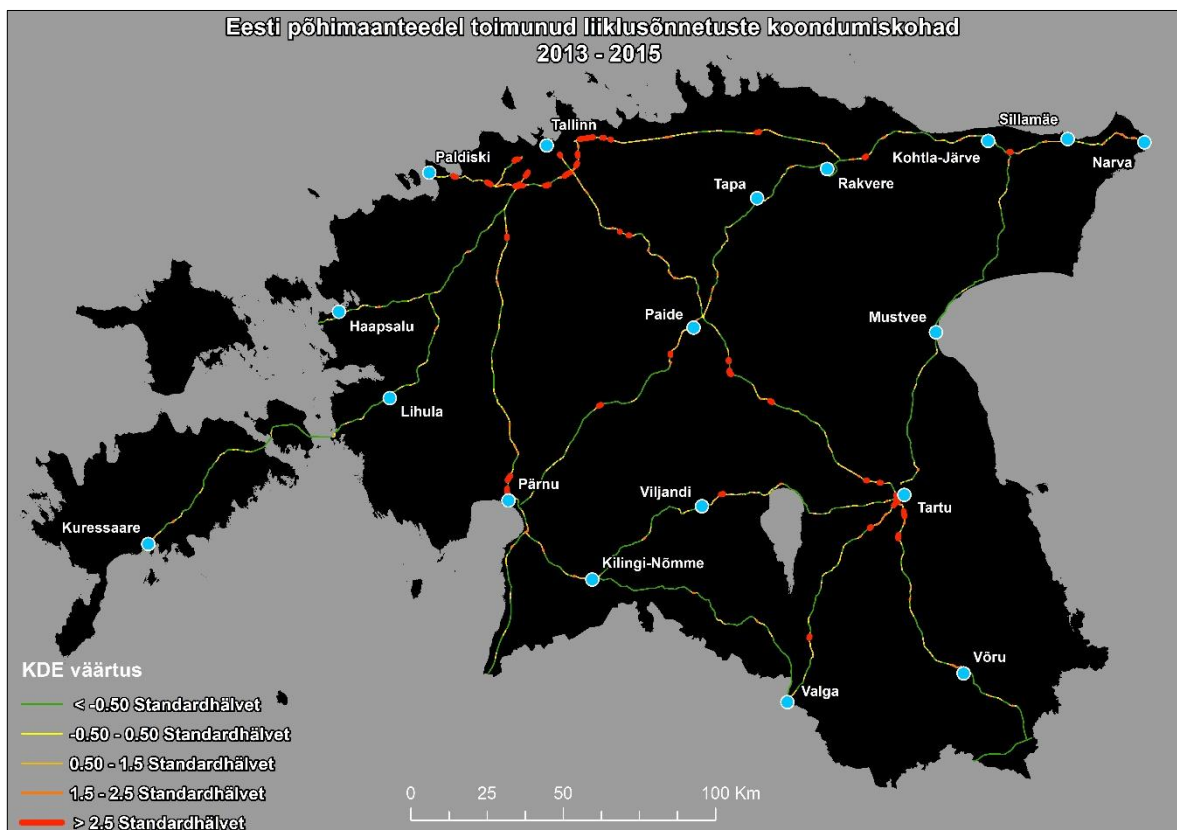
KDE otsinguraadiuseks määrati 1000 meetrit, mis Harju maakonna puhul andis häid tulemusi. 500 meetrise raadiusega leiti liialt ühe ja kahe õnnetusega koondumiskohti. Lõigu pikkuseks määrati 100 meetrit. Selles töös kasutatud KDE suurimaks miinuseks oli see, et statistilist olulisust ei olnud koheselt võimalik kontrollida. Optimeeritud tulipunktide analüüsiga statistiliselt olulisi KDE väärtusi leida ei õnnestunud, kuna tulemuseks oli weakood. Seetõttu kasutati selleks ArcGISi *Hotspot*-analüüsi tööriista (*Hot Spot Analysis Getis-Ord Gi**). Sama meetodit kasutasid oma töös tõenäoliselt ka Peet ja Padrik (2015).

Hotspot-analüüsi (*Getis-Ord Gi**) puhul kasutati sisendina KDE tulemust ning kaaludena meetodi poolt arvatud väärtusi. SANETi KDE suureks plussiks on ka see, et tulemused väljastatakse nii joonte kui ka punktidenä. ArcGIS aga arvutab punktutihedust rastrina ning Peet ja Padrik (2015) pidid punktid tekitama ise ja tegid seda iga 100 meetri tagant. Selles töös kasutatud KDE annab tulemuseks aga juba punktidest koosneva tiheduspinna koos nende KDE väärtustega. 1000 meetrist otsinguraadiust ja 100 meetrist lõigu pikkust kasutades oli tulemuseks 34727 punkti. Põhimaanteede kogupikkus selle töö andmestiku järgi oli 1 626 760 meetrit (koos praamiliini ja ETAKi teelõikudega). Kauguse tüübina kasutati *Hotspot*-analüüsi puhul Manhattani kaugust ning otsinguraadiuseks määrati 500 meetrit. See suurus valiti seetõttu, et punkte võrreldaks omavahel vaid kuni ühe kilomeetri pikkusel teelõigul (500 meetrit ühele ja teisele poole). Lisaks kasutati kauguse puhul pöördfunktsiooni ruutu, mille tulemusena üksteisele lähedal juhtunud õnnetused avaldavad arvutuste tulemusele suuremat mõju kui kaugemal toimunud õnnetused.

Statistiliselt oluliseks koondumiskohaks arvestati punkte, mille usaldusnivoo oli 90% ja rohkem. Tulipunkti algus- ja lõpupunktid määrati selle järgi, kus oli esimene või viimane vähemalt 90%lise usaldusnivooga *Hotspot*-analüüsiga saadud punkt. Lõikude m-koordinaadid saadi *Identify Route Locations* tööriista kasutades ning aluskihina kasutati Tanel Jairusel algandmestikuna saadud riigimaanteede kihti, millel olid m-koordinaadid olemas. Joonistel esitatakse kõik statistiliselt olulised punktid, kuid tabelis kirjeldatakse vaid neid teelõike, kus on toimunud kolm ja rohkem õnnetust, mis jäävad statistiliselt oluliste punktide vahele. Kuna koondumiskoht ulatub statistiliselt oluliste punktide alusel toimunud õnnetustest kaugemale, siis hõlmavad need teelõigud ka ohtlikke kõrgema riskiga alasid. Ohutsooniks ei ole esimese ja viimase õnnetuse vahele jääv ala, kuna tulipunktide mõju ulatub ka kaugemale.

3.7.2 Tulemused

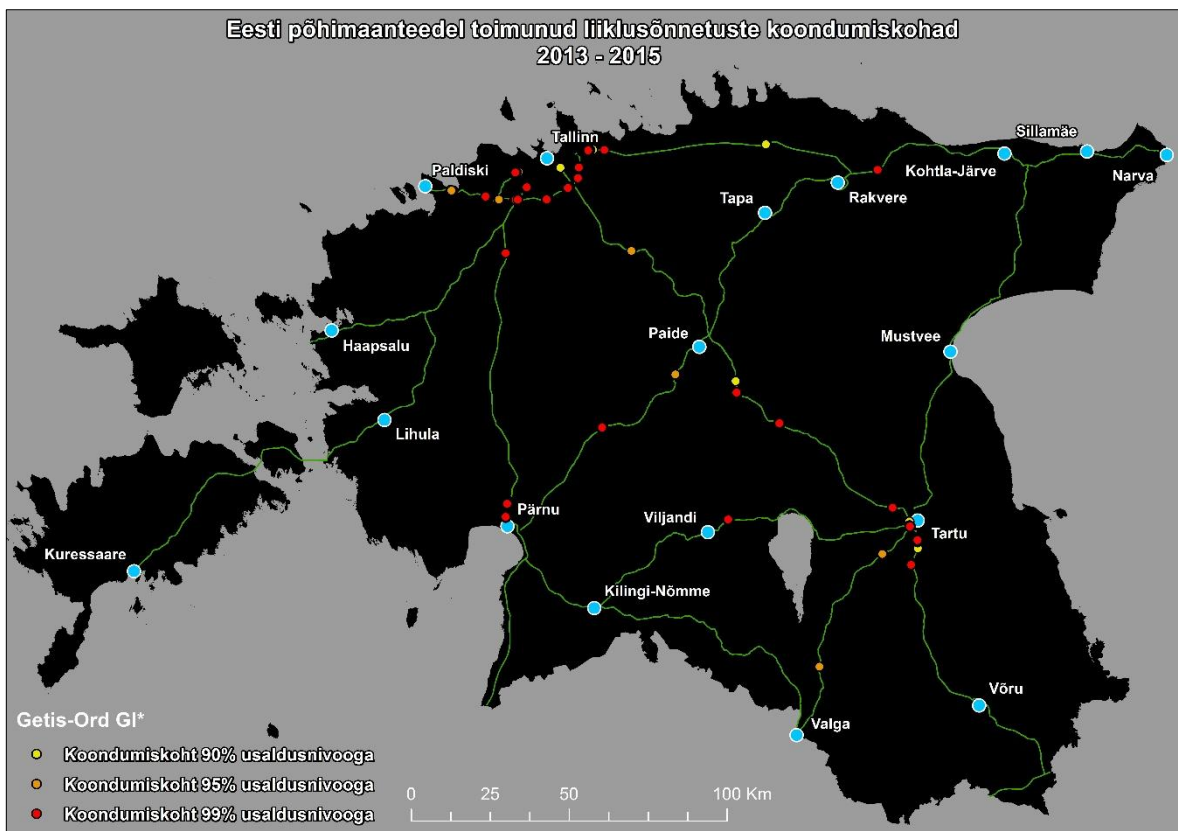
Kokku juhtus aastatel 2013–2015 Eesti põhimaanteedel 633 liiklusõnnetust. Nendes sai vigastada 1032 ning hukkus 77 inimest. KDE esmane tulemus (väärtuste standardhälbe klassifikatsiooniga) leidis kõige rohkem koondumiskohti Tallinna ringtee ja Tartu linna lähiümbruses (joonis 28).



Joonis 28. Eesti põhimaanteedel toimunud õnnetuste koondumiskohad aastatel 2013–2015 (KDE)

KDE tulemus oli sisendiks *Hotspot*-analüüsile, et leida statistiliselt olulised punktid. Tulemus oli väga sarnane standardhälbe klassifikatsiooni tulemusele (joonis 29). Statistiliselt oluliseks ei osutunud vaid Tallinn – Narva maantee Kohtla-Järve lähistel olev teelõik ning Tallinn – Tartu – Võru – Luhamaa maantee Paide ja Tallinna vahel olev teelõik.

Ajavahemikul 2013–2015 põhimaanteedel toimunud õnnetuste koondumiskohtades, mille arv käesoleva töö järgi on 21, toimus kokku 101 õnnetust. Kui üleüldse hukkus põhimaanteedel 77 inimest, siis koondumiskohtades vaid kolm. Põhimaanteedel vigastada saanute üldarv oli 1032, kuid tulipunktides sai vigastada 110 inimest. Samas tasub meele pida, et vaid 101 õnnetust 633st moodustavad koondumiskohti. Kõige rohkem tulipunkte oli Tallinna ringteel ja Tallinn – Tartu – Võru – Luhamaa maanteel, kummalgi viis ohtlikku teelõiku (joonis 29).

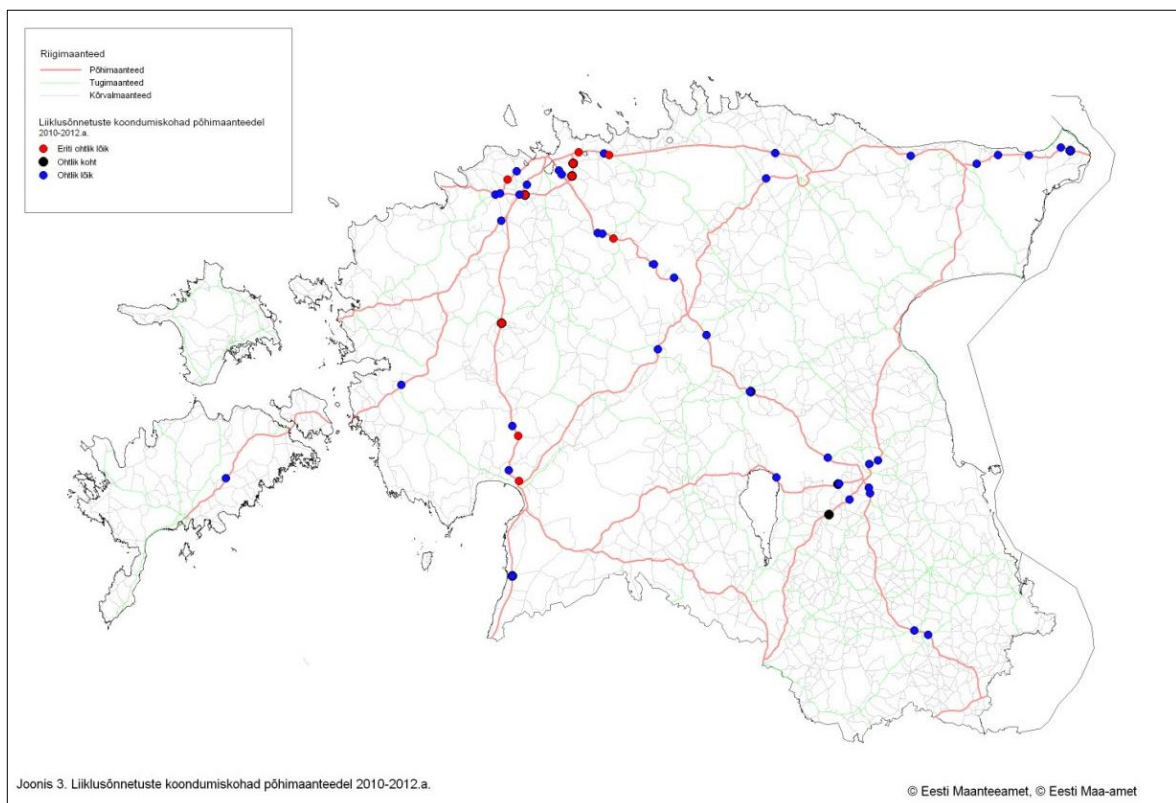


Joonis 29. Eesti põhimaanteedel toimunud õnnetuste koondumiskohad aastatel 2013–2015 (KDE ja Getis-Ord Gi*)

Kombineeritud meetodi abil leiti üles vaid statistiliselt olulised koondumiskohad. Andmestikus oli tegelikult veelgi selliseid teelõike, mis olid alla ühe kilomeetri pikad ning kus toimus vähemalt kolm õnnetust. Nendel segmentidel toimunud õnnetused paiknesid üksteisest suhteliselt kaugel, mistõttu nende KDE väärtus jäi suhteliselt väikeseks. Sellest

tulenevalt ei pidanud *Hotspot*-analüüs neid löike statistiliselt oluliseks. Stratum (2013) kasutas koondumiskohtade leidmiseks liikuva akna meetodit, mille tulemused tõenäoliselt statistiliselt kontrollitud ei ole. Kui võrrelda ajavahemikke 2010–2012 ning 2013–2015, siis Stratum (2013) on leidnud oluliselt rohkem koondumiskohti (joonis 30). Põhjus seisneb eelkõige andmetes, kuna kahes töös kasutati erineval ajavahemikul toimunud õnnetusi. Samas on oluline roll ka erinevatel meetoditel. Tasub meeles pidada, et 2013-2015 ajavahemikul toimunud õnnetuste koondumiskohad on statistiliselt olulised.

Mõlema tulemuse puhul on kõige ohtlikumad Tallinna ringtee ja Tallinn – Tartu – Võru – Luhamaa maantee. Suur osa Tallinna ringtee kahe ajavahemiku koondumiskohtadest langevad kokku (joonis 30). Osa Tallinn – Tartu – Võru – Luhamaa tulipunktidest on samuti samades kohtades, näiteks Lõuna-Tartu, Põltsamaa ja Ravila aleviku piirkonnad. Väiksemas määras on tõenäoliselt sarnaste tulipunktide asukohad veidi nihkes. Võrreldes Stratumi (2013) tulemusega, leiti käesolevas töös kõige vähem koondumiskohti Tallinn – Narva maanteel.



Joonis 30. Liiklusõnnetuste koondumiskohad põhimaanteedel 2010-2012.a (Stratum, 2013)

3.7.2.1 Teepõhised tulemused ajavahemikul 2013–2015

Tallinn – Narva maanteel leiti kaks statistiliselt olulist koondumiskohta (tabel 3). Üks teelõik oli küll 2,6 kilomeetrit pikk, kuid seal toimus ka 10 õnnetust. Vaid ühes õnnetuses sai vigastada inimene. Peamiseks õnnetuse tüübiks oli kokkupõrge teevälise takistusega. Teise ohtliku teelõigu pikkus oli vaid 225 meetrit ning seal juhtus kolm õnnetust, milles ei hukkunud ega saanud vigastada ükski inimene.

Tabel 3. Tallinn – Narva maanteel toimunud liiklusõnnetuste koondumiskohad

Tee nr	Tee nimetus	Algus (m)	Lõpp (m)	Pikkus	LÕ	H	V	Valdav õnnetuse tüüp
1	Tallinn – Narva	11483	14092	2609	10	0	1	Kokkupõrge teevälise takistusega
1	Tallinn – Narva	70491	70716	225	3	0	0	Sõiduki teelt väljasõit
Kokku					13	0	1	

Tallinn – Tartu – Võru – Luhamaa maantee koondumiskohtades juhtus kolme aasta jooksul 23 liiklusõnnetust, milles hukkus üks ning vigastada sai 29 inimest (tabel 4). 1496 meetri pikkusel teelõigul juhtus üheksa õnnetust ning seal sai ka kõige rohkem inimesi vigastada (kümme inimest). Tabelis kolmel viimasel ohtlikul lõigul juhtusid õnnetused ka ristmike läheduses.

Tabel 4. Tallinn – Tartu – Võru – Luhamaa maanteel toimunud liiklusõnnetuste koondumiskohad

Tee nr	Tee nimetus	Algus (m)	Lõpp (m)	Pikkus	LÕ	H	V	Valdav õnnetuse tüüp
2	Tallinn – Tartu – Võru – Luhamaa	105017	105274	257	3	0	5	Sõiduki teelt väljasõit
2	Tallinn – Tartu – Võru – Luhamaa	108650	109618	968	4	1	7	Sõiduki teelt väljasõit ja Kokkupõrge vastutuleva sõidukiga
2	Tallinn – Tartu – Võru – Luhamaa	126678	127288	610	3	0	3	Kokkupõrge vastutuleva sõidukiga
2	Tallinn – Tartu – Võru – Luhamaa	189352	190848	1496	9	0	10	Kokkupõrge ees liikuva sõidukiga
2	Tallinn – Tartu – Võru – Luhamaa	197743	198600	857	4	0	4	Erinevad
Kokku					23	1	29	

Jõhvi – Tartu – Valga maanteel oli üks õnnetuste koondumiskoht, pikkusega 284 meetrit (tabel 5). Sellel lõigul toimus kolm õnnetust ning vigastada sai viis inimest. Kõik õnnetused toimusid ristmikul.

Tabel 5. Jõhvi – Tartu – Valga maanteel toimunud liiklusõnnetuste koondumiskohad

Tee nr	Tee nimetus	Algus (m)	Lõpp (m)	Pikkus	LÕ	H	V	Valdav õnnetuse tüüp
3	Jõhvi – Tartu – Valga	150543	150827	284	3	0	5	Kokkupõrge sõidukiga küljelt
Kokku					3	0	5	

Tallinn – Pärnu – Ikla maanteel oli neli koondumiskohta (tabel 6). Nendest kolm olid pikkusega üle ühe kilomeetri, kuid ka õnnetuste arv oli suurem kui kolm. Kokku toimus koondumiskohtades 20 liiklusõnnetust. Vigastada sai 15 ning hukkus üks inimene.

Tabel 6. Tallinn – Pärnu – Ikla maanteel toimunud liiklusõnnetuste koondumiskohad

Tee nr	Tee nimetus	Algus (m)	Lõpp (m)	Pikkus	LÕ	H	V	Valdav õnnetuse tüüp
4	Tallinn – Pärnu – Ikla	13040	14171	1131	7	0	4	Erinevad
4	Tallinn – Pärnu – Ikla	36606	36825	219	3	0	2	Erinevad
4	Tallinn – Pärnu – Ikla	119626	120981	1355	4	0	4	Erinevad
4	Tallinn – Pärnu – Ikla	123991	125246	1255	6	1	5	Erinevad
Kokku					20	1	15	

Pärnu – Rakvere – Sõmeru maanteel oli üks koondumiskoht pikkusega 717 meetrit (tabel 7). Sellel teelõigul toimus neli liiklusõnnetust, milles sai vigastada seitse inimest. Üks õnnetus toimus ristmikul ning valdavaks õnnetuste tüübiks oli kokkupõrge sõidukiga küljelt.

Tabel 7. Pärnu – Rakvere – Sõmeru maanteel toimunud liiklusõnnetuste koondumiskohad

Tee nr	Tee nimetus	Algus (m)	Lõpp (m)	Pikkus	LÕ	H	V	Valdav õnnetuse tüüp
5	Pärnu – Rakvere – Sõmeru	43844	44561	717	4	0	7	Kokkupõrge sõidukiga küljelt
Kokku					4	0	7	

Tallinn – Paldiski maanteel leiti kaks koondumiskohta (tabel 8). Kumbki oli kuue õnnetusega tulipunkt. Kui ühes sai vigastada viis inimest, siis teises kümme inimest. Esimese koondumiskoha viis ja teise tulipunkti kaks õnnetust toimusid ristmikul.

Tabel 8. Tallinn – Paldiski maanteel toimunud liiklusõnnetuste koondumiskohad

Tee nr	Tee nimetus	Algus (m)	Lõpp (m)	Pikkus	LÕ	H	V	Valdav õnnetuse tüüp
8	Tallinn – Paldiski	12950	13985	1035	6	0	5	Kokkupõrge vastutuleva sõidukiga
8	Tallinn – Paldiski	26788	27984	1196	6	0	10	Erinevad
Kokku					12	0	15	

Tallinna ringtee oli koos Tallinn – Tartu – Võru – Luhamaa maanteega kõige suurema koondumiskohtade arvuga põhimaantee. Ringteel leiti viis koondumiskohta, milles hukkus kaks ning vigastada sai 34 inimest (tabel 9). Tabelis teise koondumiskoha kõik kolm õnnetust ning tabelis viimase koondumiskoha neli õnnetust toimusid ristmikul.

Tabel 9. Tallinna ringteel toimunud liiklusõnnetuste koondumiskohad

Tee nr	Tee nimetus	Algus (m)	Lõpp (m)	Pikkus	LÕ	H	V	Valdav õnnetuse tüüp
11	Tallinna ringtee	4744	5120	371	3	0	4	Erinevad
11	Tallinna ringtee	8083	8548	465	3	0	3	Kokkupõrge sõidukiga küljelt
11	Tallinna ringtee	12271	13633	1362	5	0	9	Erinevad
11	Tallinna ringtee	20010	21075	1065	5	2	8	Erinevad
11	Tallinna ringtee	30021	31313	1292	7	0	10	Kokkupõrge ees liikuva sõidukiga
Kokku					23	2	34	

Tartu – Viljandi – Kilingi-Nõmme maanteel oli üks ohtlik teelõik, pikkusega 573 meetrit (tabel 10). Kolmes õnnetuses sai vigastada neli inimest.

Tabel 10. Tartu – Viljandi – Kilingi-Nõmme maanteel toimunud liiklusõnnetuste koondumiskohad

Tee nr	Tee nimetus	Algus (m)	Lõpp (m)	Pikkus	LÕ	H	V	Valdav õnnetuse tüüp
92	Tartu – Viljandi – Kilingi-Nõmme	65265	65838	573	3	0	4	Sõiduki teelt väljasõit
Kokku					3	0	4	

Põhimaanteedel nr 6, 7, 9 ja 10 ei leitud ühtegi statistiliselt olulist õnnetuste koondumiskohta.

Kokkuvõte

Käesoleva magistritöö eesmärgiks oli leida Eesti põhimaanteedel ajavahemikus 2013 kuni 2015 toimunud liiklusõnnetuste tulipunktid. Koondumiskohtade leidmiseks on arendatud mitmeid tööriistu, millest suur osa on välja töötatud pindade analüüsimiseks, kus kõik suunad on avatud. Seetõttu püüti selles töös ka välja selgitada, kuidas on pindadele välja töötatud meetodeid võimalik tööle saada ka nende sündmuste jaoks, mis saavad toimuda ainult kitsamates kanalites. Selliste intsidentidena kasutati selles töös riigimaanteedel liiklusõnnetusi. Töö käigus katsetati esmalt nelja erinevat meetodit, et võrrelda tulemusi ja selgitada välja nende head ja halvad küljed. Nendeks meetoditeks olid õnnetuste arv kilomeetristel lõikudel, optimeeritud tulipunktide analüüs, hierarhiline lähima naabri analüüs ja kerneli punktiheduse meetod. Esimene ja viimane meetod arvestavad punktidevahelist kaugust mööda teed ning teise ja kolmanda meetodi puhul kasutatakse eukleidilist või Manhattani kaugust. Parima siin töös kasutatud meetodi abil leiti Eesti põhimaanteedel toimunud liiklusõnnetuste koondumiskohad.

Töö käigus selgus, et kilomeetriste lõikude meetod leiab kõige vähem koondumiskohti. Põhjus seisneb segmenteerimise meetodikas, mis põhjustab selle, et õnnetused võisid toimuda kahel erineval kilomeetril lõigul. Samuti on võimalus, et õnnetused paiknevad kas lõigu päris alguses või lõpus ühe kobarana. Selline meetod sobib kõige paremini esmase ülevaate saamiseks.

Optimeeritud tulipunktide analüüsil on olemas ümbritseva polügoni funktsioon, mille abil on võimalik ära määrata alad, kus uuritavad intsidendid üleüldse toimuda saavad. Selle funktsiooni abil on töövahendi tulemused ka oluliselt täpsemad. Funktsiooni mittekasutamisel on tulemused kasutuskõlbmatud. Väga hästi leidis meetod ülesse ohtlikud ristmikud. Kuna meetod on välja töötatud tasandil toimuvate intsidentide uurimiseks, siis suur osa statistiliselt olulistest ohtlikest kohtadest leitakse hoopiski õnnetuste raskuskeskmesse. Seda aga paralleelsete teede puhul ja kui õnnetused on eukleidilise kauguse poolest lähedal. Samuti ei ole selge, kuidas mõjutab ümbritseva polügoni suurus töövahendi parameetrite valikuid, kuna nende valik peaks lähtuma lähima naabri kaugusest. Ka ArcGIS versioonide 10.2 ja 10.3 optimeeritud tulipunktide analüüsi töövahendid andsid sama andmestikuga erinevaid tulemusi. Meetod sobib paremini tiheda teedevõrgustikuga alade uurimiseks.

Hierarhilise lähima naabri analüüsi tulemusi mõjutasid paralleelselt kulgevatel teedel toimunud õnnetused. Erinevatel teedel juhtunud intsidente tuleks aga analüüsida eraldi ning seetõttu ei sobi HLNA tiheda ja keeruka teedevõrgu analüüsimiseks. Kui aga analüüsida ühte konkreetset teed, on tulemused usaldusväärsed. Ettevaatlik tuleb olla otsinguraadiuse parameetri määramisel, kuna liiga suurt väärtust kasutades on tulemuseks väga suured klastrid, mis ei pruugi vastata kasutaja poolt määratud koondumiskoha definitsioonile.

Kerneli punktiheduse meetod andis töös kõige täpsemaid tulemusi. Põhjus seisneb nii meetodis endas kui ka kauguse tüübis, milleks selles töös oli võrgustikukaugus. Probleemaatiliseks võib osutuda väärtuse määramine, millest alates on tegu õnnetuste koondumiskohaga. Selles töös kasutati ArcGISi standardhälbe klassifikatsiooni ning tulipunktiks määrati suurima väärtusega klass. Kuna KDE suurimaks puudujäägiks on statistilise olulisuse kontrolli puudumine, siis põhimaanteed koondumiskohtade leidmiseks viidi läbi *Hotspot*-analüüs, kasutades sisendandmestikuna KDE poolt arvatud väärtusi. Tulemused olid väga sarnased standardhälbe klassifikatsioonile, mistõttu võib öelda, et ArcGISi pakutavatest klassifikatsioonidest annab parima tulemuse just standardhälbe rühmitamise meetod. Statistilise olulisuse kontroll on aga kohustuslik, kuna tulemused sõltuvad oluliselt otsinguraadiuse parameetrist. Väikese otsinguraadiusega tulevad esile lokaalsed erinevused, kuid suurema raadiusega juba eelkõige suurte väärtustega teesegmendid. Võrgustikule kohandatud KDE puhul tasub meeles pidada, et teedevõrgustik peab olema sidus ja teed omavahel ettenähtud kohtades ühendatud. Vastasel juhul on tulemused moonutatud.

Käesoleva töö koondumiskoha definitsiooniks oli üldjuhul kuni ühe kilomeetri pikkune teelõik, kus kolme järjestikuse aasta jooksul on registreeritud kolm või enam liiklusõnnetust. Sellest definitsioonist lähtuvalt leiti Eesti põhimaanteedel aastatel 2013 kuni 2015 toimunud liiklusõnnetuste koondumiskohad. Neid oli kokku 21. Kõige rohkem oli neid Tallinna ringteel ja Tallinn – Tartu – Võru – Luhamaa maanteel, kummalgi viis ohtlikku lõiku. Õnnetuste koguarv kõigi põhimaanteed tulipunktides oli 101. Neis hukkus kolm ja vigastada sai 110 inimest. Põhimaanteedel nr 6, 7, 9 ja 10 ei leitud ühtegi statistiliselt olulist õnnetuste koondumiskohta.

Methods for Identification of Accident Black Spots on National Highways

Anton Štšeglakov

Summary

The aim of this thesis was to find traffic accident black spots on the main roads of Estonia using accident data from 2013 to 2015. Several different methods have been developed to locate the black spots but most of them are made to analyse incidents in planar space. In this thesis different traffic accident black spot detection techniques were tested to find out if methods, which were developed for analysing events in planar space, would give accurate results for events, which can only happen in network space. Data of highway traffic accidents from period of 2013 to 2015 were used. Four different methods were used: road segments with length of one km; Optimized Hot Spot Analysis; Nearest Neighbour Hierarchical Clustering; Network Constrained Kernel Density Estimation. The results were compared and the pros and cons of each method were described. Traffic accident black spots on main roads of Estonia were detected using the best of previously named methods and data from 2013 to 2015.

Road segmentation method found the least number of accident black spots. The result depends on segmentation method which in some cases might cause some accidents to fall on different one kilometre segments, which are located side by side. Furthermore there is a possibility that black spot accidents are located in the very beginning or in the end of the segment. This method only suits for a brief preview of black spot locations.

Optimized Hot Spot Analysis included a bounding polygon function which allows user to define locations where incidents can take place. The results were unusable without this function. Hazardous intersections were well found with this method. Due to the planar space and Euclidean distance, statistically significant fishnet polygons were found between traffic accidents. Especially in cases, where two or more parallel roads with accidents were present. The influence of bounding polygon size remains unknown because parameters of Optimized Hot Spot Analysis are based on nearest neighbour distances. ArcGIS versions 10.2 and 10.3 give different results when using identical data and the reason behind this is also unknown.

Nearest Neighbour Hierarchical Clustering method results were influenced by accidents on parallel roads. Roads with different driving conditions should be analysed separately and this is the reason why this method is not suitable for analysing dense and difficult road

network. It is suitable for one road analysis. Search radius parameter should be defined with caution. Too big values can cause large clusters, which do not suit the definition of black spot.

Results of network constrained Kernel Density Estimation were the most precise. The reason behind these results is the method itself and usage of network distance instead of Euclidean and Manhattan distance. Defining black spot cutoff value of KDE result is difficult. ArcGIS standard deviation classification was used in this work.

Statistical significance of KDE is not determined in SANET software. This is the biggest con of network constrained KDE method. This is the reason why Getis-Ord statistic was calculated from the result of KDE points. It was calculated to locate statistically significant black spots of Estonian main roads. The results were very similar to standard deviation classification. Test for statistical significance is necessary because the results strongly depend on bandwidth size. Local differences can be found with small bandwidths, but with greater bandwidths values, segments with the highest point density can be located. KDE user must keep in mind that road network must be topologically correct. Otherwise the results are wrong.

In this work a black spot is defined as a road segment less than one kilometre in length, where at least three traffic accidents have occurred during period of three sequential years. Traffic accident black spots on main roads of Estonia were detected using a combination of two methods - Network Constrained Kernel Density Estimation and ArcGIS Hot Spot Analysis (Getis-Ord G_i^*). Data of highway traffic accidents from period of 2013 to 2015 were used. 21 statistically significant black spots were located. Two of Estonian main roads had five black spots and this was the biggest count per certain road. Overall number of accidents in all black spots was 101. Number of fatalities was three and 110 people got injuries. No black spots were found on four main roads of Estonia.

Tänuavaldused

Suur tänu minu juhendajatele, Evelynile ja Raivole, kes jagasid asjakohaseid soovitusi töö kirjutamisel ning pühendasid oma aega töö lugemisele. Tänan ka Tanel Jairust, kes edastas Riikliku Teeregistri andmed ning vastas mitmetele andmestikku puudutavatele küsimustele. Tänan ka Sirje Lilleorgu, kes edastas täiendavad liiklusõnnetuste andmed. Lõpetuseks soovin tänada ka oma perekonda ja lähedasi inimesi, kes on mulle toeks olnud.

Kasutatud kirjanduse loetelu

Anderson, T. K. (2009). Kernel density estimation and k-means clustering to profile road accident hotspots. *Accident Analysis and Prevention*, 41 (3), 359-364.

AS Teede Tehnokeskus. (2008). Liiklusõnnetuste koondumiskohtade väljaselgitamine ja nende ohutustamise programmi koostamine, 2008-10.

Bíl, M., Andrášik, R., Janoska, Z. (2013). Identification of hazardous road locations of traffic accidents by means of kernel density estimation and cluster significance evaluation. *Accident Analysis and Prevention* 55 (2013) 265–273.

Bíl, M., Andrášik, R., Svoboda, T., Sedoník, J. (2015). The KDE+ software: a tool for effective identification and ranking of animal-vehicle collision hotspots along networks. *Landscape Ecol* (2016) 31:231–237.

Bills, S. D. (2009). Examining Hotspots of Traffic Collisions in Redlands, California, University of Redlands. (Magistritöö)

Cheng, W., and Washington, S. P. (2005). Experimental evaluation of hotspot identification methods. *Accident Analysis & Prevention*, 37(5), 870-881.

Eilat, T. (2011). Suurulukitega Eesti maanteedel toimuvad liiklusõnnetused, Tallinna Ülikool. (Magistritöö)

Elvik, R. (2007). Prospects for improving road safety in Norway. Report 897. Oslo: Institute of Transport Economics.

Elvik, R., Høy, A., Vaa, T., Sørensen, M. (2009). *The Handbook of Road Safety Measures*. London: Emerald Group.

Erdogan, S., İlçi, V., Soysal, O., Korkmaz, A. (2015). A model suggestion for the determination of the traffic accident hotspots on the turkish highway road network: A pilot study. *Boletim de Ciências Geodésicas Journal*.

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 (2008) 174–181.

- Ernits, E., Koppel, O. (2012). Liiklusõnnetustest ühiskonnale põhjustatud kahjude määramise meetodika täiustamine, kahjude suuruse hindamine ja prognoosimine. Tallinna Tehnikaülikool: Logistikainstituut.
- Euroopa komisjon. (2011). Valge raamat. Euroopa ühtse transpordipiirkonna tegevuskava – Konkurentsivõimelise ja ressursitõhusa transpordisüsteemi suunas.
- Jairus, T. (2015). Maanteede parameetrite mõju liiklusohutusele, Tallinna Tehnikaülikool. (Magistritöö).
- Koppel, K. (2015). PheWAS ja selle praktiline läbiviimine Tartu Ülikooli Eesti Geenivaramu andmete põhjal. Tartu Ülikool. (Bakalaureusetöö)
- Kundakci, E. (2014). Identification of Urban Traffic Accident Hot Spots and Their Characteristics by Using GIS. (Magistritöö).
- Levine, N. (2013a). CrimeStat IV documentation - Chapter 2: Quickguide to CrimeStat IV.
- Levine, N. (2013b). CrimeStat IV documentation - Chapter 3: Entering Data into CrimeStat.
- Levine, N. (2013c). CrimeStat IV documentation - Chapter 7: Hot Spot Analysis of Points: I, Part III: Hot Spot Analysis.
- Maailma Terviseorganisatsioon. (2015). Global Status Report On Road Safety 2015.
- Majandus- ja Kommunikatsiooniministeerium. (2013). Transpordi arengukava 2014-2020.
- May, J.F., (1964). A Determination of Accident Prone Location. Traffic Eng. 34 (5), 21–27.
- Merchant J.W. (1984). Using spatial logic in classification of Landsat TM data. Proceedings of the 9th Annual Pecora Symposium, Sioux Falls, South Dakota, 378–385.
- Metsvahi, T. (2002). Liiklusõnnetuste koondumiskohtade muutused maanteedel 1996 – 2001. aasta andmetel. TTÜ: Teedeinstituut.
- Metsvahi, T. (2005). Liiklusõnnetuste koondumiskohtade muutused maanteedel 2000 – 2004. aasta andmetel. TTÜ: Teedeinstituut.

- Moreira, I. (2011) Identification of Hazardous Road Locations in Urban Areas through Spatial analysis. Lissaboni Ülikool. (Magistritöö)
- Mungnimit, S., Jierranaitanakit, K., Chayanan, S. (2009). Sequential Data Analysis for Black Spot Identification. 4th IRTAD Conference, Seoul, Korea.
- Nie, K., Wang, Z., Du, Q., Ren, F., Tian, Q. (2015). A Network-Constrained Integrated Method for Detecting Spatial Cluster and Risk Location of Traffic Crash: A Case Study from Wuhan, China. Sustainability, Vol 7, Iss 3, Pp 2662-2677.
- Okabe, A. (2015). SANET: Spatial Analysis along Networks. User Guide/Manual for SANET Standalone Beta.
- Okabe, A., Satoh, T., Sugihara, K. (2008). A Kernel Density Estimation Method for Networks, its Computational Method and a GIS-based Tool. International Journal of Geographical Information Science, Volume 23 Issue 1.
- Peet, K., Padrik, J. (2015). Eesti riigimaanteedevõrgu loomaõnnetuste registri loomine ning liiklusohlike lõikude selgitamine. OÜ Hendrikson & Ko.
- Prasannakumar, V., Vijitha, H., Charuthaa, R., Geethaa, N. (2011). Spatio-Temporal Clustering of Road Accidents: GIS Based Analysis and Assessment. Procedia Social and Behavioral Sciences 21 (2011) 317–325.
- Produit, T., Lachance, N., Strano, E., Porta, S., Joost, S., & Timothy, P. (2010). A Network Based Kernel Density Estimator Applied to Barcelona Economic Activities. In Computational Science and Its Applications–ICCSA 2010 (pp. 32–45). Springer Berlin Heidelberg.
- Smith, S., Bruce, C. (2008). CrimeStat III User Workbook and Data.
- Stratum. (2013). Liiklusõnnetuste koondumiskohtade väljaselgitamine.
- Swedish National Road Consulting AB (SweRoad). (2001). Black Spot manual, Road Improvement And Traffic Safety, Final Report Volume I.

Štšeglakov, A. (2014). Maanteede kurvilisuse ja liiklusõnnetuste esinemissageduse vahelised seosed. Tartu Ülikool. (Bakalaureusetöö)

Tang, L., Kan, Z., Zhang, X., Sun, F., Yang, X., Li, X. (2015). A network Kernel Density Estimation for linear features in space–time analysis of big trace data. *International Journal of Geographical Information Science*.

Tang, Y., Knodler, M., Park, M. (2013). A comparative study of the application of the standard kernel density estimation and network kernel density estimation in crash hotspot identification. 16th Road Safety on Four Continents Conference Beijing, China 15-17 May 2013.

Trasberg, T. (2014). Sissejuhatus geoinformaatikasse, 3. loeng: Ruumianalüüs ja kaardi päringud. Ettekanne Euroakadeemias 20. märtsil 2014.

Äär, H. (2014). Eesti maanteedele paigutatud kiiruskaamerate mõju liiklusõnnetuste vähenemisele. Tartu Ülikool. (Magistritöö)

Xie, Z., Yan, J. (2013). Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: an integrated approach. *Journal of Transport Geography* 31 (2013) 64–71.

Xie, Z., Yan, J. (2008). Kernel Density Estimation of traffic accidents in a network space. *Computers, Environment and Urban Systems*, 32(5), 396-406.

Yu, H., Liu, P., Chen, C., Wang, H. (2014). Comparative analysis of the spatial analysis methods for hotspot identification. *Accident Analysis and Prevention* 66 (2014) 80–88.

Internetiallikad:

ArcGIS. (2015a). How Optimized Hot Spot Analysis Works.

<https://desktop.arcgis.com/en/desktop/latest/tools/spatial-statistics-toolbox/how-optimized-hot-spot-analysis-works.htm> (Kasutatud 20.12.2015).

ArcGIS. (2015b). How Hot Spot Analysis (Getis-Ord G_i^*) works. <http://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-how-hot-spot-analysis-getis-ord-gi-spatial-stati.htm> (Kasutatud 21.12.2015).

ArcGIS. (2016). Incremental Spatial Autocorrelation. <http://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/incremental-spatial-autocorrelation.htm> (Kasutatud 10.03.2016).

ArcGIS. (2015c). Optimized Hot Spot Analysis. <https://desktop.arcgis.com/en/desktop/latest/tools/spatial-statistics-toolbox/optimized-hot-spot-analysis.htm> (Kasutatud 20.12.2015).

ArcGIS. (2015d). What is a z-score? What is a p-value? <http://desktop.arcgis.com/en/desktop/latest/tools/spatial-statistics-toolbox/what-is-a-z-score-what-is-a-p-value.htm> (Kasutatud 20.12.2015).

Eesti ohtlikumad teelõigud. (2009). Eesti ohtlikumad teelõigud 2003-2007, veebikaart.

Leitav aadressilt:

<https://www.google.com/maps/d/viewer?mid=zY7Y8tcFViHg.keMYDMFw2kdo&hl=en>

Levine, N. (2016). Isiklik kirjavahetus.

Lilleorg, S. (2011). Liiklusõnnetuste koondumiskohad Eesti riigimaanteedel perioodil 2008-2010. <http://www.mnt.ee/?id=15953> (Viimati kasutatud 10.01.2016).

Maanteeamet. (2015). Liiklusõnnetused ning neid puudutavate mõistete seletused. <http://www.mnt.ee/index.php?id=12948> (Viimati kasutatud 20.02.2016).

Remm, K., Remm, J., Kaasik, A. (2012) Ruumiliste looduandmete statistiline analüüs. Õpik-käsiraamat. (http://kalleremm.ee/RASA/RASA_teorია.pdf).

Spatial Analysis. (2015). http://help.cadcorp.com/en/7.1/sis/help/Spatial_Analysis.html (Kasutatud 24.01.2016).

Wikibooks. (2016). Transportation Geography and Network Science/Circuitry. https://en.wikibooks.org/wiki/Transportation_Geography_and_Network_Science/Circuitry (Kasutatud 23.04.2016).

Lisad

Lisa 1. Arvutusteks kasutatud sülearvuti üldandmed

Protsessor: Intel Core I5-2410m

Graafikakaart: integreeritud Intel HD 3000

Kõvaketas: Crucial MX200 SSD

Mälu: 8 GB DDR3

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Anton Štšeglakov,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose
“Riigimaanteedel toimunud liiklusõnnetuste koondumiskohtade leidmise meetodid”,
mille juhendajad on Evelyn Uemaa ja Raivo Aunap,

- 1.1.reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil,
sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse
tähtaja lõppemiseni;
 - 1.2.üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas
digitaalarhiivi DSpace´i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.
3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega
isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, **20.05.2016**