



RITA

Rahvastikuregistri elukoha andmete õigsuse kontroll ja täpsustamise võimalused mobiilpositsioneerimise andmetel

TP 2 aruanne

Autorid:
Anto Aasa
Martin Haamer

Tartu 2021

Sisukord

| | |
|---|----|
| Sissejuhatus..... | 3 |
| Mobiilpositsioneerimise meetodika rahvastiku paiknemise ja mobiilsuse uurimiseks..... | 4 |
| Mobiilpositsioneerimine ja selle alamtüübid..... | 6 |
| Passiivne mobiilpositsioneerimine..... | 6 |
| Passiivse mobiilpositsioneerimise andmete ruumiline interpoleerimine | 9 |
| Ankurpunktide määramine | 12 |
| Telefonikasutajad kui valim | 13 |
| Tegevusruumide määratlemine..... | 15 |
| Isikustatud mobiilandmete sidumine rahvastikuregistriga..... | 16 |
| Andmed ja meetodika..... | 16 |
| Elisa ja Telia mobiilimastide paiknemise võrdlus..... | 16 |
| Ankurpunktide määramine | 19 |
| Elukohaandmed | 19 |
| Andmete terviklikus | 19 |
| Töövoo kirjeldus..... | 19 |
| Kodu ankurpunktide ja elukoha aadressi kattumise uurimine | 20 |
| Tulemused..... | 22 |
| Kodu ankurpunktide kattumine RR aadressidega..... | 22 |
| Mobiilsidevõrgu kärgede alusel | 22 |
| Haldusjaotuse alusel | 24 |
| Kodu ankurpunktide kattumine eneseraporteeritud aadressiga | 25 |
| Mobiilsidevõrgu kärgede alusel | 25 |
| Haldusjaotuse alusel | 26 |
| Rahvastikuregistri (RR) elukoha ja mobiilpositsioneerimise andmete kattuvus | 26 |
| Eneseraporteeritud elukoha ja mobiilpositsioneerimise andmete vastavus | 27 |
| Meetodika rakendamine..... | 28 |
| Mobiilpositsioneerimise rakendatavus rahvastiku paiknemise ja liikuvuse statistika tootmiseks... 28 | |
| Eesti Töötukassa tööturuteenuste ja -toetuste infosüsteemi andmete kasutamise võimalused inimeste elukohaandmete täpsustamiseks | 29 |
| Kinnistusraamatu andmete kasutatavus inimeste elukohaandmete täpsustamiseks | 31 |
| Mobiilpositsioneerimise andmete registritega võrdluse kokkuvõte | 33 |
| Kasutatud kirjandus | 35 |

Sissejuhatus

Mobiilpositsioneerimise andmeid tunnustatakse täna kui kõige suurema potentsiaaliga uut andmeallikat, mida kasutada sisendina rahvastiku paiknemise ja liikuvus analüüsiks ning statistika tootmiseks. Seejuures on võimalik kasutuse spekter väga lai nii uuritavate teemade kui ka mõõtkava suhtes. Mobiilpositsioneerimise andmed on kasutatavad nii üldise riikliku statistika tootmiseks (Kamenjuk, Aasa and Sellin, 2017; Saluveer *et al.*, 2020a; Aasa *et al.*, 2021) kui ka väiksemate sotsiaalsete gruppide kirjeldamiseks (Ahas, Aasa, *et al.*, 2010; Mooses *et al.*, 2020).

Mobiilpositsioneerimist käsitletakse sageli kui ühetaolist andmekogumise meetodit. Reaalsus on ka selles osa väga mitmekesine ja ulatub aktiivse positsioneerimise A-GPS'i mõnemeetrise täpsusest kuni passiivse mobiilpositsioneerimise mobiilantenni teeninduspiirkonna suuruseni. Väga erinev on positsioneerimismeetodite hind (valimi suurus, arvutusjõudlus, raha jm), riskid inimeste privaatsusele jm. Tänapäevase seisuga on üheks levinumaks mobiilpositsioneerimise andmeallikaks passiivne mobiilpositsioneerimine, mille puhul kasutatakse asukohaandmeid kõnetoimingute/mobiilseinterneti toimingute asukohainfot mobiilantenni täpsusega. See on odav, võimaldab katta suhteliselt kergesti väga suurt osa populatsioonist, lihtsasti anonümiseeritav ning seeläbi väga madala ohuga inimeste privaatsusele.

Traditsioonilisele andmekogumisele on väljakutse esitanud rahvastiku kasvav mobiilsus, mis on muutnud tavapärasemaks elukoha sagedasema vahetamise või lausa mitme elukoha vahel liikumise (nt sesoonne või nädalane rütm). Elukoha andmete korrektsus on vajalik paljude teenuste osutamiseks, kuid paraku on sageli probleemiks vananenud elukohaandmed erinevates registrites. Elukohta vahetades ei pöörata tihti tähelepanu elukohaandmete uuendamisele Rahvastikuregistris ning endine elukoht säilitatakse teadlikult omavalitsuselt teatud soodustuse saamiseks või siis missioonitundest, et omavalitsust oma maksudega aidata. Samuti ei teatata alati teisest elukohast juhul, kui mitme elukoha vahel regulaarselt liigutakse.

Traditsiooniliste rahvaloenduste korraldamine on väga ressursinõudlik: kulub palju aega ning raha. Seetõttu on rahvaloendusi korraldatud vaid iga 10 aasta tagant. Ühe lahendusena on nähtud registripõhise rahvaloenduse juurutamist. Eesti puhul on probleemiks aga rahvastikuregistri elukoha-aadresside suur lahknevus tegelikest elukohtadest. Erinevate hinnangute kohaselt ulatub rahvastikuregistris valesti registreeritud elukohtade osakaal 15 – 25 protsendini. Suure nihke kujunemist on põhjustanud erinevad faktorid, kuid igal juhul on selge, et selline täpsus ei vasta rahvaloenduse ootustele.

Probleemi lahendamiseks on tegeldud erinevate alternatiivsete andmeallikate kasutamise uurimisega, et nende abil rahvastikuregistri aadresside täpsust parendada. Nii on selleks kasutatud erinevate aadressandmeid sisaldavate riiklike registrite ristkasutust ning selle alusel erinevaid tegeliku elukoha paiknemise skoori arvutada (Eesti Statistika tegeleb residentsuse, partnerluse ja paiknemisindeksite arendamisega). Osades riikides on rahvastikuregistri elukohaga seotud inimese postiaadress (nt Soome). Uuritud on ka elektri- ja internetitarbimise andmestiku kasutatavust.

Käesoleva peatüki eesmärk on uurida passiivse mobiilpositsioneerimise andmete kasutamise võimalusi rahvastikuregistri elukohaandmete kvaliteedi tõstmiseks. Peaõhk on pööratud passiivse mobiilpositsioneerimise andmetele, mille analüüsi abil tuuakse välja kaks peamist rakendusvõimalust rahvastikustatistika ja elukohaandmete täpsustamiseks. Selleks võrreldakse mobiilpositsioneerimise andmeid eelkõige Rahvastikuregistri elukoha-aadressidega, aga ka kahe teise registri (Eesti Töötukassa tööturuteenuste ja -toetuste infosüsteem [EMPIS], Kinnistusraamat) aadressandmetega.

Rahvastikuregistrile teiste registre andmete lisamise eesmärgiks on selgitada, kas on alust üha laiemalt levival hüpoteesil, et mida rohkemate registre elukohad on samal aadressil, seda suurem on tõenäosus, et just seal on ka uuritava inimese tegelik elukoht.

Mobiilpositsioneerimise meetodika rahvastiku paiknemise ja mobiilsuse uurimiseks

Mobiilpositsioneerimine seisneb mobiiltelefonide koordinaatide määramises raadiolainete kaudu. Meetodi peamiseks eeliseks on mobiiltelefonide laialdane kasutus (Eestis on 2021 aastal 135 aktiivset SIM-i 100 inimese kohta¹). Mobiiltelefon on peaaegu kõigil olemas ning enamasti alati omanikuga kaasas, olles seega väga väärtuslik andmeallikas liikumismustrite uurimisel: pole ühtegi teist elektroonilist sensorit, mille kandja oleks niivõrd huvitatud selle pidevast kaasas kandmisest ja aku laetuna hoidmisest. Siiski on meetodil mitmeid kitsaskohti, näiteks on üks põhilisi probleeme tundlikele andmetele ligipääsemine rangete privaatsusnõuete tõttu ning saadud andmete töötlemine on ajamahukas väärtusahel kõnelogide tähenduslikeks kohtadeks konverteerimisel.

Positsioneerimiseks on mitu võimalikku lähenemist. Antud analüüsis kasutatakse laialt levinud passiivse mobiilpositsioneerimise andmeid, kus telefoni asukoht on teada masti identifikaatori alusel ehk määratakse sidemasti asukoht, millega mobiiltelefon ühenduses oli.

Mobiilpositsioneerimise meetodid võib jagada aktiivseks ja passiivseks positsioneerimiseks. Aktiivse positsioneerimise korral tehakse operaatori poolt eraldiseisev päring mobiiltelefoni asukohta määramiseks raadiolaine abil. Passiivse positsioneerimise korral rakendatakse andmestikku, mida automaatselt salvestatakse kasutaja logifailidesse igal kõnetoimingul. Teisisõnu, iga kõne, sõnum või andmekasutus loetakse kõnetoiminguks, mille aeg ja asukoht fikseeritakse sidemasti täpsusega. Igal mastil on kindel geograafiline leviala ning unikaalne identifikaator. Meetodi nimetuseks on kärje-ID (inglise keeles "*Cell Global Identity*"). Teadlasteni jõuab mobiilsideoperaatorite kogutud asukohateave anonüümselt, mis võimaldab selle kasutamise uurimistöodes (Ahas, Silm, *et al.*, 2010).

Mobiilsidevõrk koosneb paljudest sidemastidest, igal ajal on enamasti mitu erinevasse suunda orienteeritud antenni. Iga antenn moodustab eraldi kärje, millest moodustub omakorda kärjevõrgustik. Kõnetoimingul määratakse, millises kärjes mobiiltelefon parasjagu asub. Kärgede suurus ja kujud varieeruvad, enamasti kasutab mobiiltelefon parasjagu lähimal asuvat või tugevaimat antenni. Antenni koormuse suurenemisel või kättesaadavuse halvenemisel võivad mobiiltelefonid luua ühenduse ümbritsevate antennidega. Erandlikel juhtudel võib suurim vahemaa telefoni ja antenni vahel ulatuda paarikümne kilomeetrit.

Linnalistes piirkondades ning peamiste liikumiskoridoride ümber on antennid paigaldatud tihedamalt ning seega on ka mobiilsidevõrgu kärjestik palju tihedam (Joonis 5). Hõredalt asustatud piirkondades on antennid hajutatud suuremale alale. Seetõttu on mobiilpositsioneerimine linnades tunduvalt täpsem kui maa- piirkondades.

Enamasti pole teada uurijatel mobiilantennide täpset suunda, leviala sektori laiust ning signaali võimsust. See tähendab, et mobiilimasti levipiirkondade täpne piiritlemine on aeganõudev ja väga keeruline tegevus. Mobiilivõrk on pidevas muutumises (tihendamine uutega, vanade antennide demonteerimine jmt). Lisaks on signaali ulatus ja seeläbi ka antenni leviala suurus sesoonse iseloomuga – näiteks kevadine kasvuperioodi alguses toimuv puude lehtimine mõjutab mobiililevi

¹ <https://datareportal.com/reports/digital-2021-estonia>

oluliselt nii aedlinnades kui ka metsataguses taludes. Välitööd levilade kaardistamiseks on ebareaalselt kallid.

Kuna antenni tegeliku leviala määramine on väga keeruline, et mitte öelda võimatu, siis kasutatakse sageli (ka käesolevas analüüsis) teoreetilisi levialasid. Ligikaudse leviala määramiseks kindlal geograafilisel alal kasutatakse tessellatsiooni – pind jaotatakse lõplikuks hulgaks polügonideks, ilma et tekkiks kattumisi või tühikuid. Selle tulemusena tekib igale sidemasti ümber polügon, mille iga punkti jaoks antud mast on lähim, teisisõnu leitakse iga masti lähikond. Tasub aga meeles pidada, et selline meetod ei vasta masti tegelikule levialale, vaid annab matemaatiliselt optimaalsema ligikaudse levikuala. Antud töös kasutatakse terminit mobiilsidevõrgu kärp nimetatud meetodi kaudu leitud ala kohta.

Passiivse mobiilpositsioneerimise andmestik (*Call Detail Record*) koosneb kõnetoimingutest, nende asukohast, kellaajast ja isiku mobiiltelefonile mobiilioperaatori poolt juhuslikult määratud identifikaatorist, mis ei ole seotud mobiilnumbriga, kuid püsib iga telefoni kohta ajas muutumatuna. See loob eeldused isiku tähenduslike kohtade ehk ankurpunktide tuvastamiseks, tagades ühtlasi isiku privaatsuse. Ankurpunktide modelleerimine on üks väärtuslikemaid väljundeid anonüümse passiivse mobiilpositsioneerimise andmestiku rakendamisel (Ahas, Silm, *et al.*, 2010).

Mobiilpositsioneerimine hakkas kiiresti arenema üsna pea pärast mobiiltelefonide masskasutusse jõudmist (Eestis 1990. aastate keskel). Üheks oluliseks arengu kiirendajaks oli hädaabi teenuste vajadus määrata abivajaja asukohta. Esimesed asukohapõhised teenused läksid Eesti tööle 2001. aastal (WAP sõnumiteenused), iga asukohamäärang võttis aega üle kümne sekundi ja maksis 13 EEK. Üsna pea saadi aru, et asukoha infot kannab ka mobiilantenni leviala, milles telefon asetseb. Sellise täpsusega asukoha määramiseks polnud vaja mobiilioperaatori poolt teha muud, kui hoida mobiilivõrk toimivana, sest positsioneerimine toimus automaatselt iga kõnetoimingu (helistamine, SMS) sooritamisega. Nii määratud asukoha täpsus on ruumis väga varieeruv ja otseselt seotud mobiilivõrgu ruumilise paiknemisega: mida tihedamalt mobiilivõrgu antennid ruumis paiknevad, seda suurem on ka asukoha määramise täpsus. Kuna tegemist on inimestele pakutava teenusega, siis on mobiilivõrk tihedam ka suurema asustihedusega piirkondades: seal, kus on palju inimesi on palju mobiilimaste ja väiksed mastide teeninduspiirkonnad. Hõredalt asustatud aladel on antennide tihedus väiksem. Iga uue mobiiltehnoloogia põlvkonna saabumisega on muutunud ka mobiilivõrgu ruumiline paiknemine tihedamaks ja koos sellega positsioneerimine täpsemaks.

Tänaseks elame asukohapõhiste teenuste maailmas, kus on vaja mõnemeetrist asukohatäpsust. Seda pakub nutitelefonide satelliit positsioneerimise võimekus (A-GPS). Selline asukoha määramine on varasemate meetoditega võrreldes mobiilioperaatorist oluliselt sõltumatum. Kõrge täpsus tõstab hüppeliselt privaatsuse ja andmekaitsega seotud probleeme ning enamasti pole inimesed valmis sellist infot enda kohta teistega jagama. Nii ongi nii Eestis kui ka mujal maailmas ühiskonna hüvanguks valdavalt kasutatud oluliselt madalama ruumilise täpsuse ja ajalise tihedusega passiivne mobiilpositsioneerimine.

Passiivse mobiilpositsioneerimine on olnud võimalik tänases tähenduses mobiiltelefonide jaoks algusest peale: mobiilioperaator registreerib oma andmebaasis iga kõnetoimingu, mille käigus salvestub kõnetoimingu seotud seadme ID, aeg ning toiminguks kasutatava mobiilimasti ID. Tegemist pole sisu jälgimisega, salvestub vaid dünaamika, st sündmuse toimumise aeg ja koht. Kõige laiemas üldistuses kasutab mobiilioperaator seda infot oma klientidele igakuiste arvetega väljastamiseks. Sellise andmestiku kogumine ei eelda lisakulutusi, on rakendatav masspositsioneerimise ja võib olla ka väga kiire (reaalajalähedane). Lisades siia juurde ka juba varem mainitud suhteliselt madala privaatsuse häirimise ohu, jõuamegi tõdemuseni, et tegemist on ühe

suurema potentsiaaliga andmestikuga inimeste paiknemise ja mobiilsuse uurimiseks. Seda kinnitavad paljud teadustööd ning tänaseks on jõudnud mobiilpositsioneerimise andmestik riikliku statistika tootmisprotsessi Eestis ja Indoneesias. Selle poole on püüdnud mitmeid teisi riike ja organisatsioone (nt Eurostat).

Eestis on mobiilpositsioneerimise abil inimeste ajalis-ruumilist käitumist aktiivselt uuritud juba 2004. aastast (Ahas and Mark, 2005). Kiire ja suurt hulka alamteemasid tabas tagasilöök EL-i andmekaitse üldmääruse (GDPR) rakendumise järel (2018. aasta kevad). Mure andmekaitseaduste rikkumise suured trahvide oht ja üldine selgusetus viis teadus- ja arendustöök kasutatavate andmete teadastele edastamise seiskumiseni. Lisaks on mobiilioperaatoritel endil tekkinud ärihuvid seoses oma mobiiliandmete põhiste toodete arendamisega (nt Telia *Crowd Insights*) Tänapäevaks on andmevood osaliselt taastunud (COVID-19-ga seotud inimeste mobiilsuse monitoorimine, Eesti Panga turismistatistika, TÜ turismiuuringud).

Mobiilpositsioneerimine ja selle alamtüübid

Aktiivne vs passiivne positsioneerimine

Tihti ei teadvustata, et universaalset ühetaolist mobiilpositsioneerimist pole olemas. Nii nagu on kiiresti vaheldunud erinevad mobiiltelefonide põlvkonnad (GSM, 3G, 4, [peagi ka 5G]) nii on muutunud ka mobiilpositsioneerimise meetodid. Muutus pole seejuures olnud lineaarne, vaid täna on meil paralleelselt kasutusel kõik eeltoodud tehnoloogia põlvkonnad ning vastavalt sellele ka mobiilpositsioneerimise meetodid. Tehnoloogiast sõltumata võib omakorda eristada aktiivset ja passiivset mobiilpositsioneerimist. Aktiivse positsioneerimise puhul teeb mobiilioperaator spetsiifilisi päringuid ning andmete töötlust (vaja on ka vastavat lisainfrastruktuuri), passiivne positsioneerimine on aga mobiilivõrgu nõ kõrvalprodukt, mis genereerub erinevate sisuteenuste tarbimise tulemusena või kogutakse mobiilivõrgu üldise toimimise huvides (näiteks nn *signaling data*).

Omaette täpsusega klassist on nutitelefonide A-GPSiga asukoha määramise võimalus. Käesolevas uuringus on viimane vaatluse alt välja jäetud, kuna A-GPSi väga kõrge positsioneerimise täpsusega (kuni 1 m) kasvavad kvalitatiivselt ka privaatsusega seotud küsimused. On ilmne, et sellise täpsusega inimeste asukoha määramine riigi poolt pole ühiskonnale vastuvõetav.

Passiivne mobiilpositsioneerimine

Temaatilises teadustöös võib tihti kohata lakoonilist avaldust, et „töös kasutati mobiiltelefonide kõnetoimingute andmebaasi (*CDR, Call Detail Record*). Nii mõnigi kord sellega ka andmete kirjeldus puudub. Tegelikult võib CDR-andmestik olla väga erineva sisu ja struktuuriga. Üheks levinumaks CDR-andmete tüübiks nõ kõnetoimingute andmebaas, kus on fikseeritud kõnetoimingu (telefonikõne, lühisõnum) alustaja, algusaeg ning toimingu alguses kasutatud mobiilantenni identifikaator (Joonis 1). See, nn andmetabel, ühendatakse antenniinfo tabeliga, millega antakse kõnetoimingutele teatud täpsusega (leviala) asukoht ruumis.

data table

| parameter | value |
|-----------|---------------------|
| ID | 246513389 |
| event | call |
| timestamp | 12:15:11 07/04/2014 |
| cell ID | 6547 |

cell location table

| parameter | value |
|-----------|----------|
| cell ID | 6547 |
| longitude | 25.80527 |
| latitude | 58.34998 |



Joonis 1. Passiivse mobiilpositsioneerimise andmebaasi struktuuri näidis.

Kõige rohkem võib kohata CDR-andmeid, kus on fikseeritud ainult kõnetoimingu algusaeg, sel hetkel ühenduseks kasutatava masti identifikaator ja algataja identifikaator. Rikkalikuma andmestiku puhul sisaldab see ka kõikide teiste sama toimingu jooksul kasutatud mastide identifikaatoreid ning ideaalsel juhul kogu seda infot ka kõnetoimingu vastuvõtja kohta.

Vaatleme järgnevalt ühe juhusikult valitud mobiiltelefoni kliendi CDR-andmeid 2015. aasta märtsist (Joonis 2). Kokku on respondent olnud ühe kuu jooksul seotud 51 kõnetoiminguuga. Ise algatas ta 21 kõnetoimingu ja temale suunatud oli 30 kõnetoimingu. Kui siia lisada veel telefoni asukoht kõnetoimingu lõpetamise ajal, saame kokku 102 kõnetoimingu.

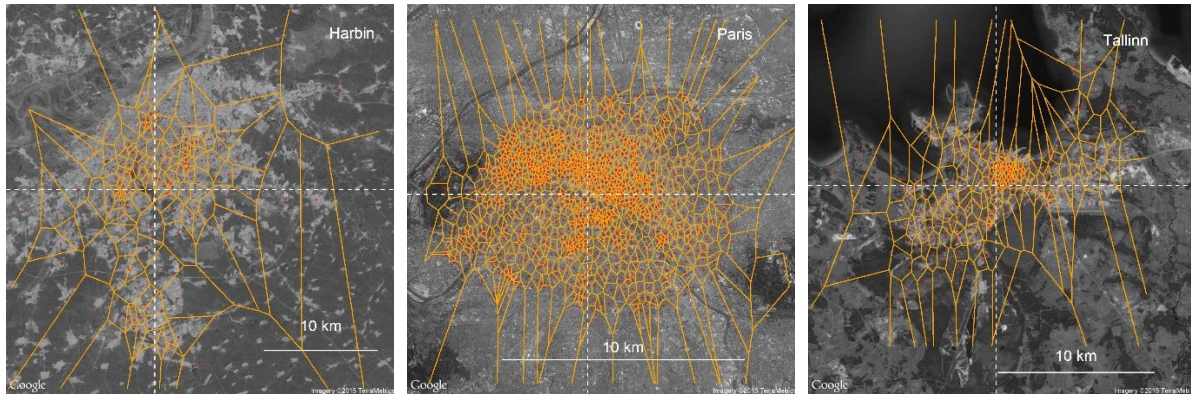


Joonis 2. Kõnetoimingute asukohad, kui on teada kõnetoimingu algataja, vastaja ning toimingu algus ja lõpp Tallinna piirkonnas.

Kuna kõnetoimingute alguse ja lõpu vahele jääb enamasti mõni minut (SMSi puhul ei midagi), siis see lisainfo märgatavat kasu andme paranemisse ei too, küll aga aitab kõnetoimingu adressaadi lisandumine andmeid oluliselt sisukamaks muuta. Kui käesoleva näite puhul oleks andmestik sisaldanud ainult kõnetoimingu algataja kirjeid, siis poleks tema Tallinna kesklinna külastust õnnestunud registreerida.

Väga oluline on arvestada ka mobiilivõrgu ruumilise struktuuriga. Lisaks asustustiheduse mõjule tuleb arvestada ka mobiilside võrgu tehnoloogiaga: erinevate põlvkondade ja ka erinevate operaatorite eri tehnoloogiaid kasutades rajatud võrgus võib mastide leviala suurus olla väga erinev (Ahas *et al.*, 2015).

Teiste piirkondade jaoks arendatud meetodikate matkimisel tuleb arvestada ka mobiilivõrkude erineva arhitektuuriga ning kasutajate erineva sotsiaal-majandusliku taustaga (Joonis 3).



Harbin

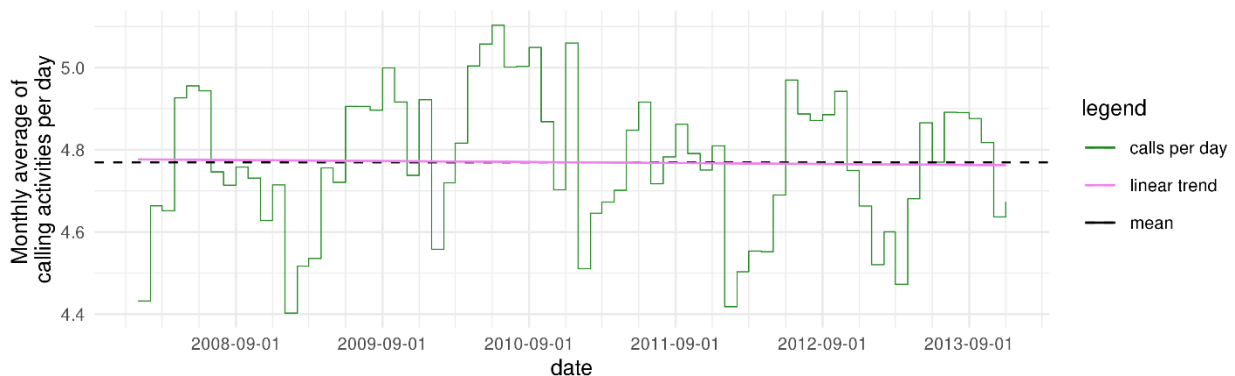
Pariis

Tallinn

Joonis 3. Mobiilimastide tiheduse erinevus kolmes maailma linnas (Ahas et al., 2015).

Mitme mobiiloperaatori andmete samaaegne kasutamine võib tekitada samuti probleeme. Peale selle, et mastide hulk ja ruumis paiknemise tihedus võib olla erinev, võib erinev olla ka kõnetoimingute ja internetikasutuse salvestumine (*a la* ühe operaatori puhul on CDR ja teise puhul CDRi ning DDRi kombinatsioon).

Lisaks mobiilside kiirest arengust tingitud tehnoloogiapõlvkondade vahetumisele on arenenud väga kiiresti ka erinevad suhtlusplatvormid. See on tõstatanud sageli küsimuse, kas telefoniga helistamine ja sõnumineerimine on ajas kokku kuivanud ning passiivse mobiilpositsioneerimise abil on inimeste asukoha tuvastamise ajaline sagedus langemas. Eesti andmetele tuginedes võib väita, et see pole nii ja kõnetoimingute keskmine hulk päevas inimese kohta on olnud ajas stabiilne (Joonis 4).



Joonis 4. Kõnetoimingute keskmine päevane hulk inimese kohta kuude lõikes (Aasa et al., 2021).

Lisaks CDR-andmetele on sageli analüüsitud inimeste liikuvuse uurimiseks ka mobiilse interneti kasutamise andmeid, mida inglisekeelses teemakirjanduses tuntakse *Data Detail record (DDR)* nime all. See tähendab, et lisaks klassikalisele helistamisele ja sõnumineerimisele salvestub ka mobiilse interneti kasutuslogi. Parematel juhtudel on teadlastel kasutada CDRi ja DDRi kombinatsioon. See tähendab, et andmepunktide hulk võib kasuda mitu korda või isegi suurusjärgu võrra. Lisaks on DDRi puhul väga suur andmetiheduse varieeruvus, kuna paljud kas ei kasuta mobiilset interneti või teevad seda väga harva. See viib meid nõ „mida rohkem ma kasutan, seda rohkem olen olemas“ efektini ning väheaktiivsete mobiilse interneti kasutajate mobiilsusinfo on väga tugevalt alaesindatud.

Paraku on mobiilpositsioneerimise andmestiku kättesaadavusega olukord nii halb, et päris andmete puuduses tegeldakse hoopis sünteetiliste andmete tootmisega, et nende abil matkida tegelikkust ning võimaldada mobiilpositsioneerimisega kogutud andmete analüüsimeetodite arendamist ning tõestada, et sellisel kujul kogutud ja töödeldud andmed ei kujuta inimeste privaatsusele mingit ohtu. Ka on mobiilpositsioneerimise andmetega seonduvalt tekkinud peaaegu paranoiline privaatsusmure. Eriti kummaline on see kui võrrelda mobiilpositsioneerimisega seonduvaid privaatsusohete internetihidude (Google, Facebook jt) kogutava andmestikuga. Kusjuures internetikäitumise puhul on inimeste privaatsusele suurimaks ohuks kasutajate mõõdutundetu eneseeksponeerimine (*self exposure*). Rõhutada tuleb, et kui mobiilpositsioneerimise andmed peegeldavad dünaamikat (Kes? Millal? Kus?), siis internetti jäetakse kergekäeliselt maha ka andmeid, mis vastavad eriliiki isikuandmete definitsioonile (nagu rassiline või etniline päritolu, poliitilised vaated, usulised või filosoofilised veendumused või ametiühingusse kuulumine, terviseandmed või andmed füüsilise isiku seksuaalelu ja seksuaalse sättumuse kohta) ning mida tuleks eriti hoolikalt kaitsta.

Kui ligipääs võimalikult rikkalikele andmetele on olemas, siis tegelik peavalu mobiilsuse uurijatele alles algab. Väärtusahel CDR ja DDR-andmetest mobiilsufoni on väga pikk: kõnetoimingud on vaja konverteerida paiknemise ja liikuvuse infoks. Seejärel tuleb mobiilimastide kaupa kogutud info ruumis tähenduslikule kujule (ankurpunktid nagu kodu, töö; turismireisid; transiit jmt) modelleerida ja interpoleerida.

Passiivse mobiilpositsioneerimise andmete ruumiline interpoleerimine

Nii CDR kui ka DDR- andmed kogutakse masti täpsusega (on teada, millise mastiga oli telefon teenuse tarbimiseks ühenduses). Masti asukoht on määratud ruumis diskreetse koordinaadipaariga. See tähendab, et me ei tea telefoni asukohta vaid mobiilimasti asukoht, mille levialas telefon viibis. Selline masti täpsusega saadud info on statistika tegijatele, planeerijatele ja ka teistele huvitatud pooltele väheväärtuslik. On väga oluline, et mastide kaupa kogud info oleks adekvaatselt ruumis nõ „laiali“ interpoleeritud. Kui mobiilimastide või nende otsas paiknevate üksikute antennide leviala ulatus oleks teada, oleks interpoleerimine suhteliselt lihtne. Paraku ei tea mobiilimastide tegelikke levialasid mitte keegi, ka mitte mobiilside teenuse pakkujad.

Lisaks sellele, et mobiilivõrk pidevalt areneb ning antenne lisandub juurde või võetakse maha, muutub antennide levimaastik pidevalt erinevatel põhjustel ka ise. Nii näiteks võib suurürituste ajal ühe või mitme mobiilantenni teenindusvõime maht ületatud saada ning osa kõnetoiminguid jääb sooritamata või kasutatakse selleks nõrgema signaaliga (ja kaugemal asetsevaid) antenne.

Oma mõju avaldab ka ilm ning aastaajalised muutused. Näiteks võib puude lehte mineku tõttu teatud piirkondades levi kvaliteet halveneda ning telefonid hakkavad sidepidamiseks kasutama muid antenne. Levi võib halvendada ka ehitatavate hoonete varjutav efekt, veekogude läheduses võib ilmnedu signaali peegelduse mõju. Tehnoloogiliste probleemide alla liigituvad häired ja katkestused mobiilivõrgu töös.

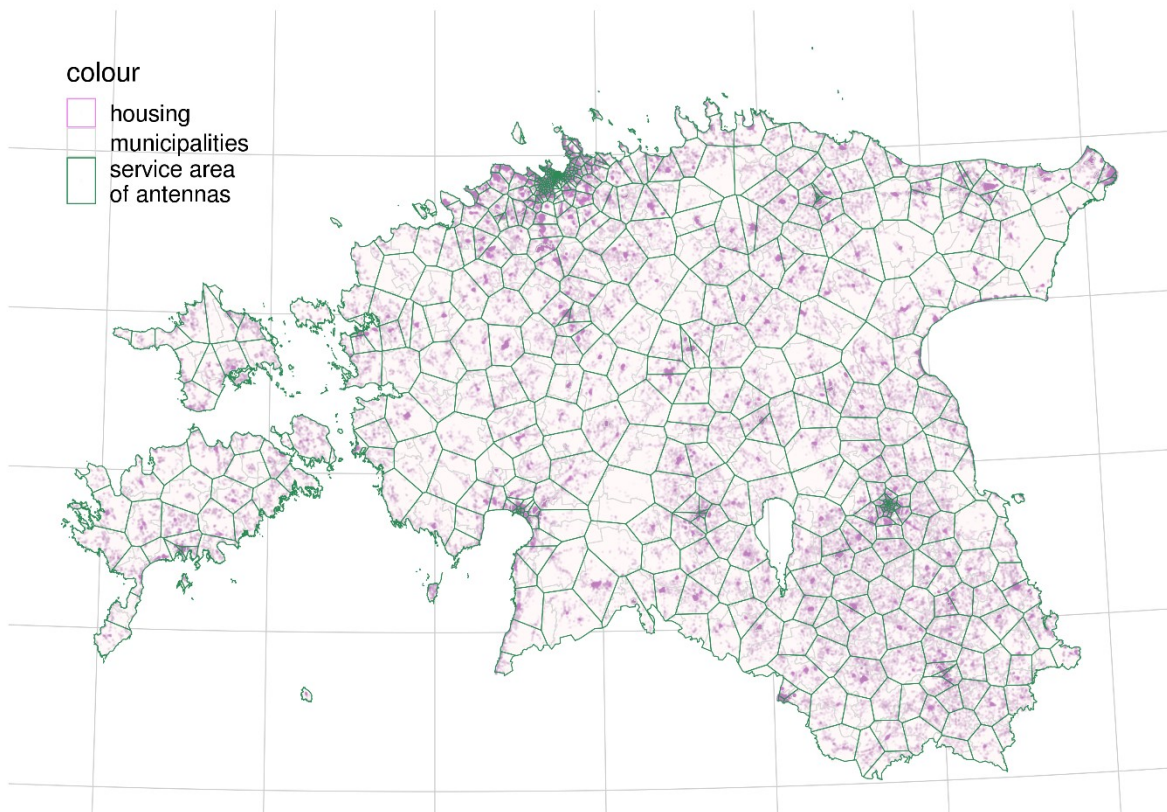
Kuigi mobiilioperaatoritel on nii ärihuvid kui ka kohustus eluliselt tähtsaid teenuseid toimivana hoida võib võrgu töö olla aeg-ajalt häiritud. Kasutades sellesse ajaperioodi jäävaid andmeid võib ka inimeste liikuvuse ja paiknemise hinnangud olla nihkega. Järgnev kokkuvõte mobiilpositsioneerimise andmete ruumilise interpoleerimise meetodikast on avaldatud käesoleva projekti toel 2021. aastal (Aasa *et al.*, 2021).

Esimestes mobiiliandmeid analüüsides kasutatud töödes ei pööratud punktandmete ruumilisele interpoleerimisele suurt tähelepanu ning kõnetoimingute hulgad konverteeriti huvi pakkuvate ruumiliste üksuste (nt omavalitsus) tasemele nn punkt-polügonis (point-in-polygon) meetodil. See

tähendab, et üksuse külastajateks määrati kõik need mobiiloperaatori kliendid, kes vastava üksuse piiridesse jääva mobiilmasti levialas oma kõnetoiminguid tegid.

Teiseks lihtsamaks ja levinumaks mobiilmastide teeninduspiirkondade määramise viisiks on thiesseni polügonide kasutamine (Joonis 5). See tähendab, et algoritm jagab mobiilmastide vahele jääva pinna nii, et iga masti ümber moodustub ala, mille iga koha jaoks on lähimaks just see mast. Seejuures toetub nii määratud masti levialadele mitu erinevat interpoleerimismeetodit. Kõige algelisem on nn pindalaga kaalumise meetod (*area weighted*). See tähendab, et thiesseni-põhisele mastide levialade kihile asetatakse sobivate haldusüksuste kiht ning arvutatakse, kui suur osakaal iga masti teeninduspiirkonnast mingi haldusüksuse piiridesse jääb. Vastavalt sellele osakaalule omistatakse ka konkreetse masti kõnetoimingute hulk vastavale ruumiüksusele. Kuigi Thiesseni polügonide puhul on tegemist lihtsustusega, mis ei arvesta mastide paiknemist, signaali tugevust ning tehnoloogilisi põlvkondi, on see tänaseni üks levinumaid meetodeid mastide levialade määramiseks. Mõlemad eespool toodud meetodid on siiski ebatäpsed ja suudavad kirjeldada vaid ca 40% rahvaarvu tegelikust ruumilisest varieeruvusest.

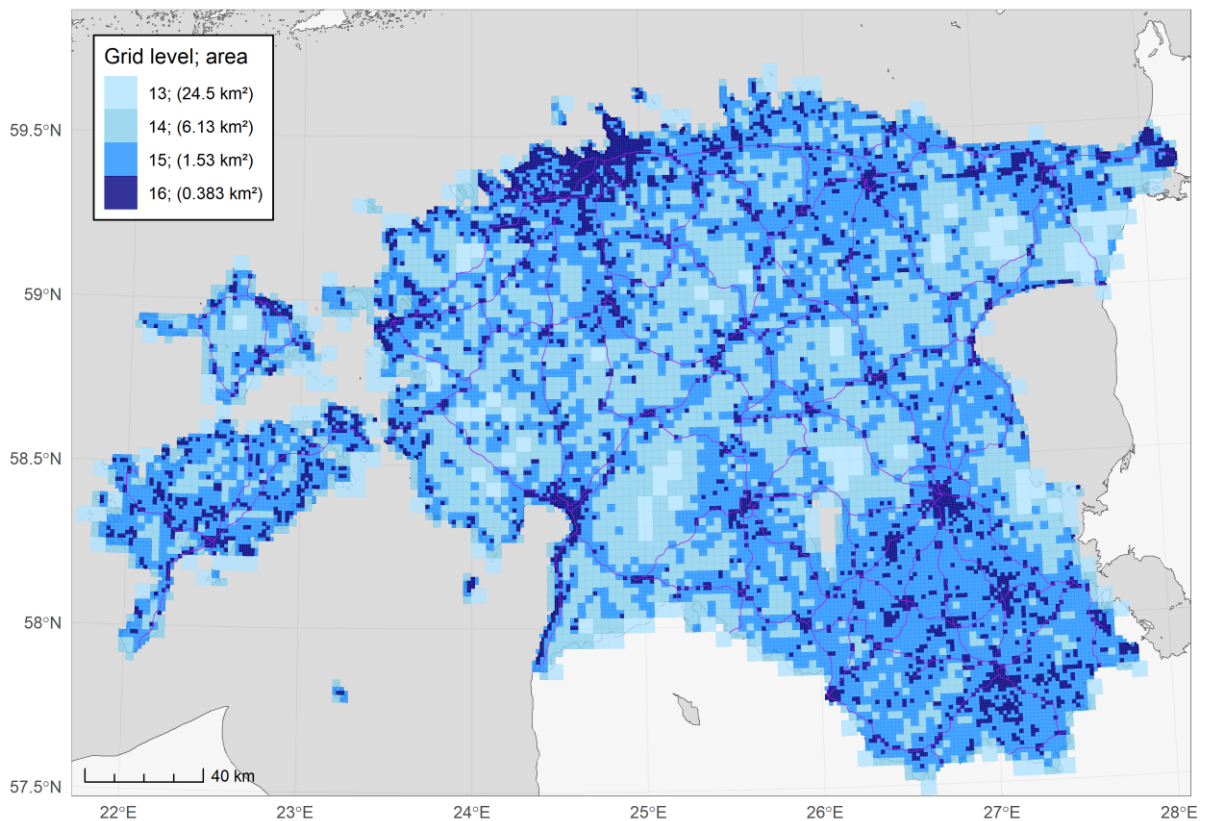
Oluliselt paremaid interpoleerimistulemusi annab abistavate andmekihti kasutamine. See tugineb lihtsale eeldusele, et inimeste elukohti ja liikumisi on interpoleerimise käigus mõistlik paigutada kohtadesse, kus on inimegevuseks sobilik infrastruktuur. Käesolevas metoodika kirjelduses on kasutatud selleks maa-ameti topograafilise andmekogu (ETAK) majade ja teede paiknemise ruumiinfot. Selle põhjal arvutatakse iga soovitud ruumiüksuse kohta inimeste olemasolu tõenäosuse skoorid. Leitud skooride alusel jaotatakse mobiilmastide kõnetoimingute info ruumiüksuste vahel.



Joonis 5. Mobiilmastide teoreetilised teeninduspiirkonnad ühe mobiiloperaatori mastide paiknemise alusel (Aasa et al., 2021).

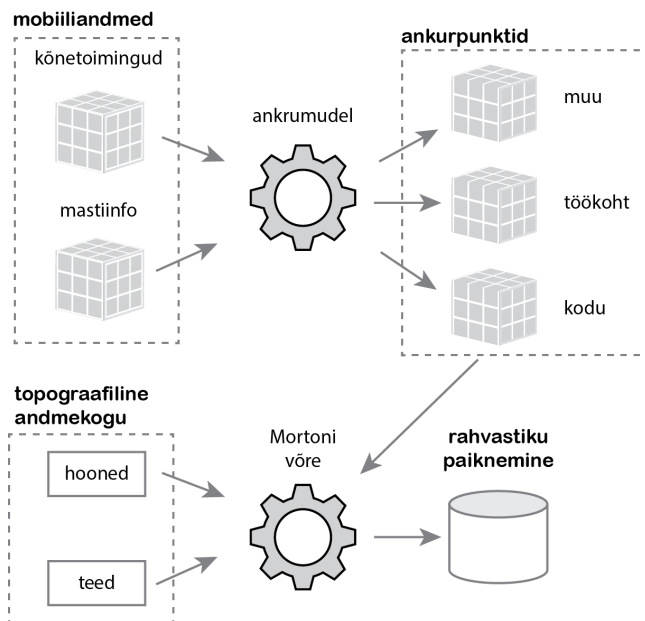
Eesti andmete tuginevas metoodikatöös (Aasa et al., 2021) on leitud, et inimeste paiknemise tõenäosuse skoorid pole mõistlik arvutada iga hierarhiataseme kohta eraldi ([asum,] asustusüksus,

[kant,] omavalitsus, maakond), vaid esmase tasemenä tuleks kasutada rahvastiku paiknemise tõenäosuse võret. Kuna rahvastik (ja ka mobiilantennid) pole ruumis paigutatud ühtlaselt, siis on soovitatud kasutada adaptiivset (kohanduva suurusega võrgusilmaga) võret. Aladel, kus inimeste tihedus on väiksem ja mobiilimastide vähem, on võrgusilm suurem ning suurema asustustihedusega aktiivse inimtegevuse piirkonnas väiksem. Ühest küljest vähendatakse sellega nõudeid arvutusvõimsusele, teisalt aitab see hoida k-anonüümsuse põhimõtetest lähtuvalt ka andmekaitse taset. Selleks on valitud nn Morton'i adaptiivne võre (erialakirjanduses *Z-order*, *Morton space filling curve*), mida on lihtne siduda globaalse süsteemiga ning on kasutusel ka muudes sarnastes rakendustes (Joonis 6). Arvestades mobiilimastide paiknemise tihedust on Eesti jaoks kasutatud Morton'i võre tasemeid 13 – 16, pindalad vastavalt 0,383km² ja 24,5km². See tähendab, et võrgusilma külje pikkus ja interpoleerimise eeldatav täpsus on vastavalt ca 0,6 ja ca 5 km.



Joonis 6. Mobiilandmete ruumiliseks interpoleerimiseks kasutatav Morton'i adaptiivne võre (Aasa et al., 2021).

Morton'i võre loomise aluseks on Eesti topograafilise andmekogu ehitiste ja teede andmekihid. Nende kihtide alusel on koostatud adaptiivne võre, millel suuremad võrgusilmad on aladel, kus on hoonestuse ja teede tihedus on madalam ja vastupidi (Joonis 7).



Joonis 7. Töövoogu elukohaandmete statistika tootmiseks passiivse mobiilpositsioneerimise andmestikust.

Ankurpunktide määramine

Inimestele meeldib mõelda, et oleme erilised ja meil on mitmekesine ning huvitav elu. Tegelikult on inimeste elu enamasti väga rutiinne, kindla ajagraafikuga ja koondunud vähestesse ruumipunktidest. Seda kinnitab näiteks ka GPS-iga läbi viidud uuring (Joonis 8). Peale selle, et valdava osa ajast (enamikul juhtudel > 80% ajast) on inimesed paigal ning ei liigu, veedetakse sellest paigal oldud ajast peamise osa ajast kodus, aga ka tööl. Muude kohtade külastuste peale kulub oluliselt vähem aega. Sageli on needki aga pelgalt hõredama rütmiga rutiinsed külastused.

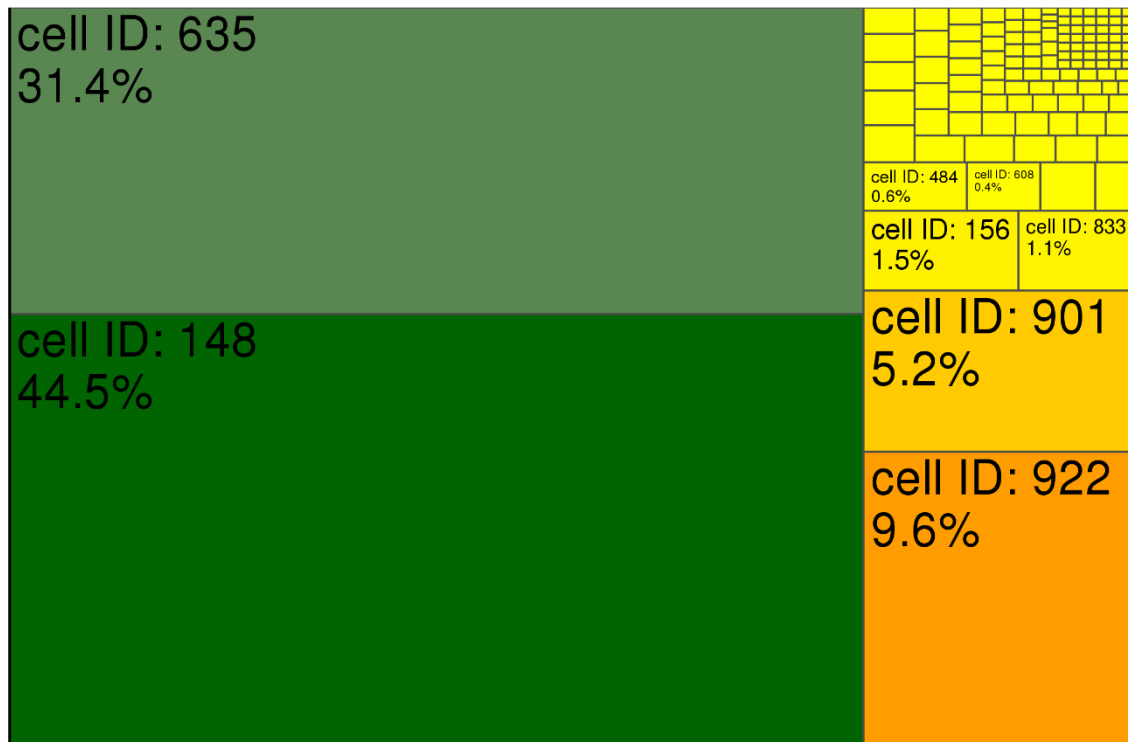
Inimeste paiknemine
GPS-andmed



A. Aasa

Joonis 8. Inimese paiknemise tähenduslikes kohtades vastavalt GPS-iga kogutud andmestikule.

GPS-i abil kogutud inimeste paiknemise ajamustrite loogikat kinnitavad ka passiivse mobiilpositsioneerimise andmed. Nii nagu GPSi asukohad on koondunud paari tähtsamasse kohta, nii on ka suurem osa mobiiltelefonide kõnedest koondunud vaid mõne mobiilmasti levialasse (Joonis 9). Kõnetoimingute koonduvuse esmapilgul paistev suurem varieeruvus on tingitud asjaolust, et joonis pole koostatud mastide (ühes kohas mitu antenni) kaupa, vaid iga mastis asuva antenni kohta eraldi.



Joonis 9. Kõnetoimingute ruumilist jaotust iseloomustab tugev toimingute koondumine samasse kohta.

Suuremahulistest mobiilpositsioneerimise logidest on vähe kasu juhul, kui nende põhjal ei oleks võimalik tuvastada isikule tähenduslikke kohti, näiteks kodu- ja töökoht. Neid regulaarselt külastatavaid kohtasid võib pidada isiku ankurpunktideks. Siiski ei ole passiivse mobiilpositsioneerimise täpsus ankurpunktide tuvastamisel perfektne, kuna erinevate inimeste liikumismustrid varieeruvad, mis raskendab kodu- ja töökoha määramist. Näiteks, juhul kui inimene teeb valdava osa kõnetoimingutest teel koju või tööle, ei vasta tema peamine ankurpunkt ei kodu- ega töökohale. Sellepärast kasutatakse paralleelselt kodu ja töö ankurpunktidega mõistet tähenduslik koht, mis viitab regulaarselt külastatavale asukohale (Ahas, Silm, *et al.*, 2010).

Kodu- ja töökoha ankurpunkt määratakse tuvastades sidemast, millega ühenduses olles on tehtud enim kõnetoiminguid uuritava perioodi jooksul. Igale mastile leitakse kõnetoimingute keskmine kellaeg ning selle standardhälve. Kodu ankurpunkti määramiseks kasutatakse masti, mille kõnetoimingute keskmine kellaeg on pärast kella 17. Juhul kui standardhälve on suurem kui 0,175, määratakse kodu ankurpunkt ka teistele mastidele, mille kõnetoimingute keskmine kellaeg on enne kella 17. Kui standardhälve on alla 0,175 ja keskmine kõnetoimingute algusaeg enne kella 17, määratakse kõrg töökoha ankurpunktina (Ahas, Silm, *et al.*, 2010).

On ka võimalus, et isikul on rohkem kui üks kodu või töö ankurpunkt. Üks enim levinud põhjuseid selleks on mobiilside masti vahetumine asukohta muutmata. Linnades on mastid palju tihedamalt ja seda probleemi eriti märgata. Juhul kui ankurpunktid asuvad kõrvuti olevates kärgedes, võrreldakse kõnetoimingutega päevade arvu ja valitakse kahest suurema viibitud päevade arvuga kõrg.

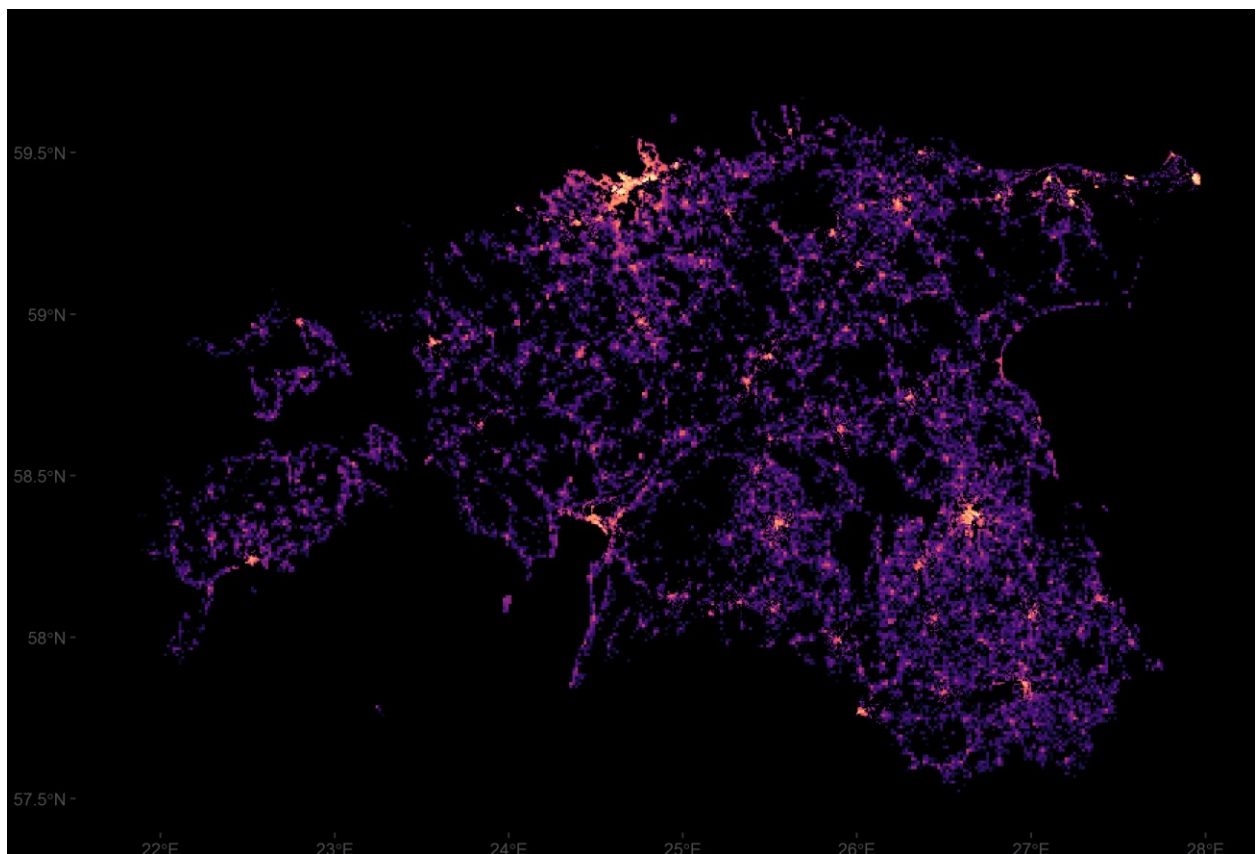
Telefonikasutajad kui valim

Kui ankurpunktid on mastidest adekvaatselt soovitud ruumiüksustesse interpoleeritud, on vaja lahendada järgmine probleem: kuidas viia ühe või mitme mobiiloperaatori kõnetoimingute alusel arvutatud ankurpunktid populatsiooni tasemele? Ka selleks etapi täitmiseks on võimalik kasutada erinevaid meetodeid. Kõige lihtsamal viisil on arvutatud tegeliku rahvaarvu (rahvaloendus,

rahvastikuregister) ja mobiilipõhiste kodukohtade hulga suhe. Saadud koefitsiendiga korrutatakse läbi kõikide ruumiüksuste ankurpunktide hulk. Nii saadud kogu populatsiooni iseloomustavad väärtused võivad olla tegelikkusega võrreldes tugevas nihkes. Selle põhjuseks võib olla konkreetse mobiiloperaatori turuosa varieeruvus piirkondade ja sotsiaalsete gruppide lõikes. Sellise arvutuse tulemusena on suur oht, et eelkõige väiksema rahvaarvuga ruumiüksuste jaoks prognoositud rahvaarv on tegelikkusest väga erinev.

Oluliselt paremaid tulemusi on andnud regressioonimudeli kasutamine. Et rahvaarvude jaotus administratiivüksustes vastab ka Eestis Zipfi mudelile, siis on rahvaarvud mudelis transformeeritud 10-logaritmi abil. Saadud lineaarne mudel annab prognoosväärtused iga uuritava ruumiüksuse kohta. Kui saadud mudelis pole erindeid (käesolevas näites pole), siis võib prognoosväärtust omakorda parandada vastava prognoosväärtuse mudeli jäägiga. Käesoleva meetodika jaoks on mudel loodud 2011. aasta seisuga. Seda on tehtud eeldusel, et mudeli argumenttunnusena kasutatud rahvaarv ja rahvastiku paiknemine on vahetult pärast rahvaloendust kõige täpsem (Aasa *et al.*, 2021).

Nii arvutatud mobiilpositsioneerimise rahvaarv kirjeldab rahvaloenduse aegset rahvastiku paiknemise varieeruvust 95% täpsusega (Joonis 10). Saadud mudeli ajaline püsivus (statsionaarsus), tugineb eeldusel, et kõnetoimingute hulk inimese kohta pole ajas muutunud. Võime isegi arutleda, et mobiilide alusel arvutatud rahvastikumuster on tegelikkusele lähemal.



Joonis 10. Mobiilpositsioneerimise abil modelleeritud rahvastiku paiknemine (tiheasustusaladel 100 m, mujal 1000 m võrel).

Kuigi käesoleva projekti eesmärk on parandada elukohaandmete kvaliteeti, siis tegelikult võimaldab ankurpunktide mudeli rakendamine välja selgitada ka teisi olulisi kohti, kus telefonikasutaja viibib (töökoht, teine kodu jm).

Tegevusruumide määratlemine

Võimalus mobiilandmete abil määrata telefonikasutaja mitte ainult ühte kodu piirkonda vaid ka teisi olulisi paiku, tähendab, et ühe diskreetse asukoha (kodu) asemel saame kirjeldada inimeste liikuvust ja regulaarseid tegevusruume. Siia võib lisada ka turismireisid, mille statistikat kasutab Eesti Pank (Saluveer *et al.*, 2020b).

Tegevusruumide kontseptsiooni juured ulatuvad mitmekümne aasta taha (Miller, 2005). Täna räägitakse tegevusruumist kui ruumilisest konstruktsioonist, milles toimub inimese mobiilsus ning seda võib defineerida 6-lemendiliseks: kodu asukoht, seal elamise aeg, tegevuskohad kodu lähedal, liikumine naabruskonnas, teiste sagedasti külastatavate tegevuskohtadega seotud liikumised, igapäevase elu sõlmpunktide vaheline liikumine (Schönfelder and Axhausen, 2010). Passiivne mobiilpositsioneerimine on tõestanud end väga hea andmesisendina tegevusruumide uurimisel. Ilmunud on arvukalt teadustöid, kus kirjeldatakse mobiilpositsioneerimise andmetel tegevusruumide erinevate aspektide analüüsimetoodikaid ning paiknemise ja mobiilsusega seotud statistika tootmise võimalusi: pendelränne (Novak *et al.*, 2013), elukohavahetused (Kamenjuk, Aasa and Sellin, 2017), rahvastiku sesoonsus (Silm and Ahas, 2010), hargmaisus (Mooses *et al.*, 2020).

Kuni viimase ajani kiiresti kasvanud inimeste liikuvus, nii igapäevane, maakodudega seotud kui ka turismiga seotud, on käivitanud diskussiooni, et inimesel võiks seaduse silmis olla mitu elukohta.

Tegevusruumide põhimõte on lihtne: määrata inimese igapäevane tegevusruum nii et sellesse oleks hõlmatud kõik olulised tegevuskohad (Newsome, Walcott and Smith, 1998). Meetodi tegelik rakendamine on aga märksa keerukam. Üheks levinumaks tegevusruumise kirjeldamise viisiks on olnud aktiivsuseellipsite arvutamine, milleks on kasutatud standardhälbe ellipsit (Järv, Ahas and Witlox, 2014). Sel juhul konstrueeritakse asukohapunktide ümber vastavalt nende hajuvusele ellips. Enamasti kasutatakse 1 standardhälbe ellipsit ning tegevusruumi kvantifitseerimiseks arvutatakse ellipsi pindala, telgede suhe (väljavenitatus) ning orientatsioon. Peale selle, et tegemist on väga abstraktse ja üldistava meetodiga on teatud juhtudel probleemiks ka ellipsi arvutamine ise. Ellipsi arvutamiseks on vaja vähemalt kolme üksteisest eemal asetsevat punkti. Kui kõikide kõnetoimingute puhul on kõnetoimingute asukohti rohkem, siis ankurpunkte võibki paljudel inimestel olla vaid üks või kaks. Seejuures tuleb meele pida, et nimi ankurpunkt on eksitav: punkti asemel on meil teada masti teeninduspiirkond või selle tuletis. Ja kui meenutame, et need teeninduspiirkonnad on suuremad hõredalt asustatud aladel, siis võib juhtuda, et kogu igapäevane tegevusruum mahub ühe või kahe masti teeninduspiirkonda. See aga tähendab, et ellipsit ei saa konstrueerida. Probleemi lahendamiseks võib ellipseid kombineerida puhvrite ja mastide teoreetiliste teeninduspiirkondadega. Puhvri (asukoha ebatäpsust arvestades nt 2...5 km) saab konstrueerida näiteks ümber kahe ankurpunkti vahele tõmmatud joone või siis ümber ühe masti. Ellipsite ja puhvrite kombineerimisega kaasneb probleem, et nende jaoks pole võimalik arvutada samu parameetreid ning nad pole omavahel hästi võrreldavad.

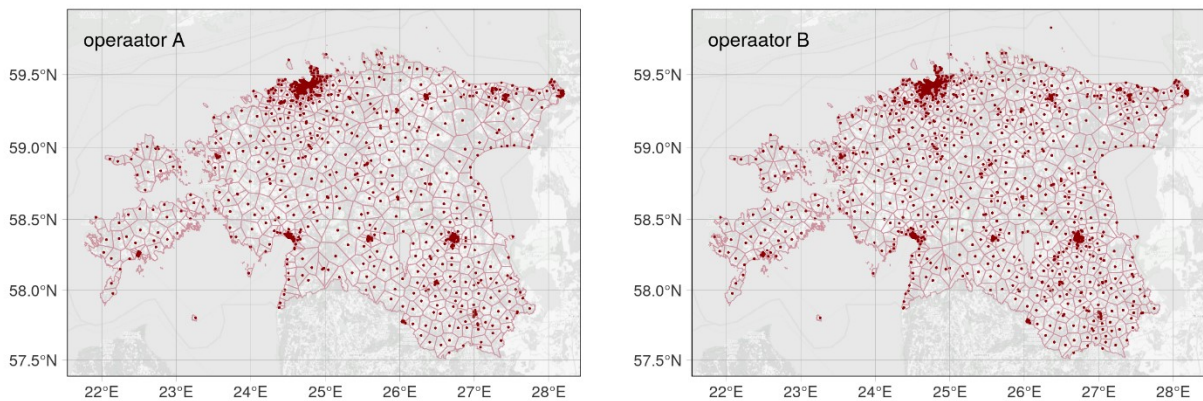
Üks võimalus on lausalise tegevusruumi konstrueerimise asemel arvutada tegevusruum ainult kohtadesse, kus paigal ollakse ning transpordilingid jäetakse arvutustest välja. Passiivse mobiilpositsioneerimise puhul on neiks kohtadeks ankurpunktid. Aga ka siin tuleb arvestada, et mastide teeninduspiirkonnad on erineva suurusega: nii võib näiteks arvutuste tulemuseks olla, et Tallinnas on inimestel palju väikse ulatusega tegevusruume Soomaal aga üks ja suur. Siin võib aidata arvutustesse maakasutusinfo kaasamine, mille abil inimeste tegevuseks vähetöenäolised kohad välja filtreeritakse.

Isikustatud mobiiliandmete sidumine rahvastikuregistriga

Kui anonüümse, statistikaks vajaliku, andmestiku töötlemine on tänaseks arenenud väga heale tasemele, siis konkreetse isikule täpseks asukoha määramiseks CDR-tüüpi andmestiku abil on kiiresti piirid ees. Nii on üldise statistika tootmiseks võimalik inimesi tõenäosuse alusel ruumi paigutada ning tulemuseks on väga täpne üldine rahvastiku paiknemise andmestik, siis üksikindiviidi puhul on sageli ootuseks fikseerida asukoht konkreetse aadressi täpsusega. Seda aga passiivne positsioneerimine ei võimalda. Passiivse asukohamääramise abil võime öelda vaid seda, millise antenni levialas telefon oli. Suurema täpsuse puhul tuleb hakata opereerima tõenäosustega. Osad kitsendused on seejuures suurema tõenäosusega, kui teised. Elukohtade määramisel võime ilmselt üsna kindlalt väita, et see asub mõnes elumajas, liikumine toimub teedel jne; konkreetse maja välja selgitamine pole võimalik.

Positiivse poole pealt võib lisada, et mitmed uurimisrühmad tegelevad mastide teeninduspiirkondade täpsema piiritlemise meetodika arendamisega. Näiteks arvestatakse, mitte ainult seda mis antenni levialas õige rohkem kõnetoiminguid tehakse, vaid arvutustustesse kaasatakse ka sama mobiilimasti külge installeeritud naaberantennide levialas tehtud kõnetoiminguid.

Kui üldise statistika tootmise jaoks pole probleemiks, et mõne mobiilioperaatori andmeid pole saada ning tulemused saab kergesti üldpopulatsioonile modelleerida, siis isikupõhise elukoha täpsustamise jaoks on kriitilise tähtsusega, et kõikide andmed oleks saadaval. Selle 100-protsendiline saavutamine on aga võimatu juba ainuüksi digitaalse lõhe ja asjade interneti tingimustes. Ühel juhul on märkimisväärset osal inimestest kas telefon üldse puudu (inimese kohta pole andmebaasis infot), kasutab seda harva (ankurpunktide määramine on ebatäpne), telefon suudab kasutada vaid varasema tehnoloogiapõlvkonna (nt GSM) võrku (suurem ebatäpsus).



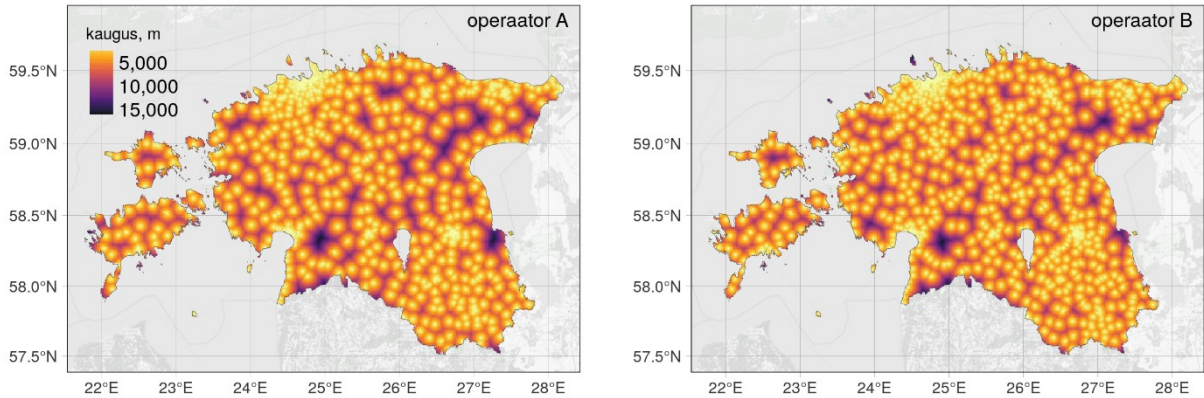
Joonis 11. Uuringusse kaasatud mobiilioperaatorite mastide paiknemine ja teoreetilised levialad (thiesseni polügonidena).

Andmed ja meetodika

Uuringusse on kaasatud 261 respondendi andmed, neist 100 on Elisa ja 157 Telia kliendid. Mobiilioperaatorite ärisaladuste ja muude huvide paljastamise vältimiseks pole mobiilioperaatoreid nimepidi mainitud. Tähistusena kasutatakse "Operaator A" ja "Operaator B".

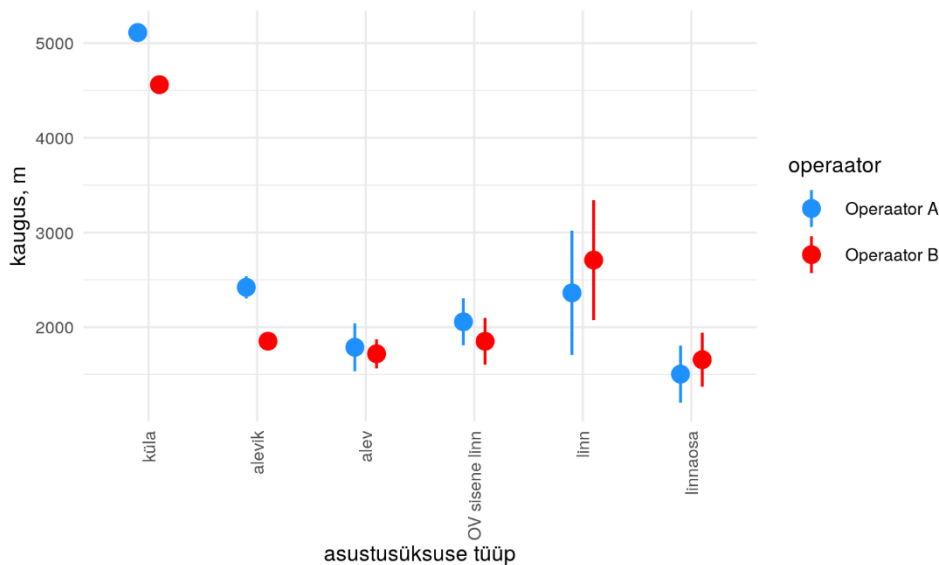
Elisa ja Telia mobiilimastide paiknemise võrdlus

Kuna passiivse mobiilpositsioneerimise (CGI) täpsus on otseses sõltuvuses positsioneeritava terminali ja mobiilimasti vahelisest kaugusest, siis on sellest tingitud operaatorite vaheliste erinevuste selgitamiseks arvutatud üle-eestiline kaugusmaatriks, kus mõlema operaatori jaoks on arvutatud iga ruumipunkti kaugus mobiilimastist.



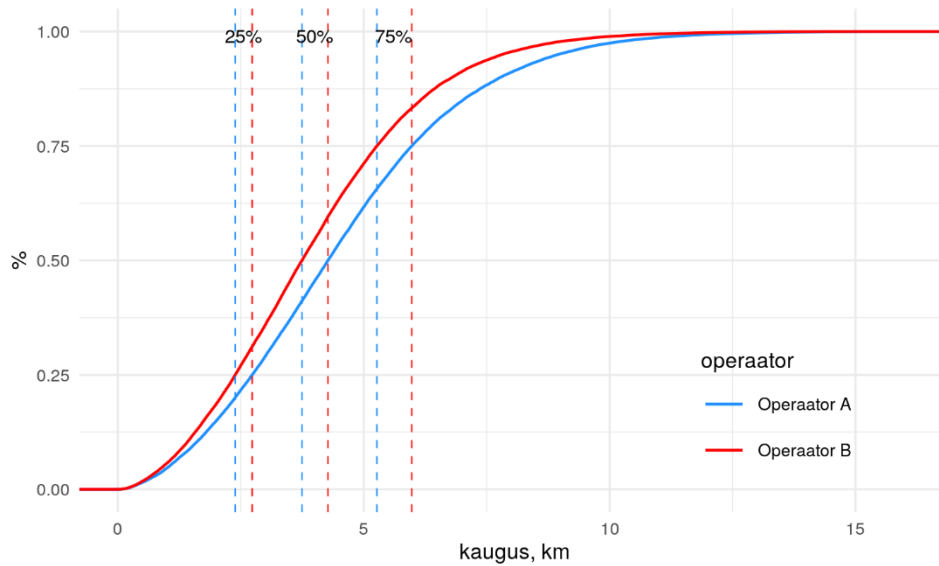
Joonis 12. Kaugus mobiilmastidest kahe uuringusse kaasatud mobiiloperaatori andmete põhjal.

Operaator A puhul on Eesti maismaa pinnal keskmine kaugus mobiilmastist 4,5 kilomeetrit, maksimaalne kaugus ulatub 15,7 kilomeetrini. Operaator B puhul on Eesti maismaa pinnal keskmine kaugus mobiilmastist 3,97 kilomeetrit, maksimaalne kaugus ulatub 14,98 kilomeetrini. Üldiselt võib aga väita, et kahe uuringusse kaasatud mobiilmastide paiknemismustrid on olemuselt väga sarnased ning sellest tingitud erinevusi ja moonutusi analüüsi tulemustes esineda ei tohiks. Mobiilside mastide paiknemine on otseselt seotud asustustihedusega ja maste on rohkem suurema asustustihedusega piirkondades, mida kinnitab ka järgmine joonis (Joonis 13).



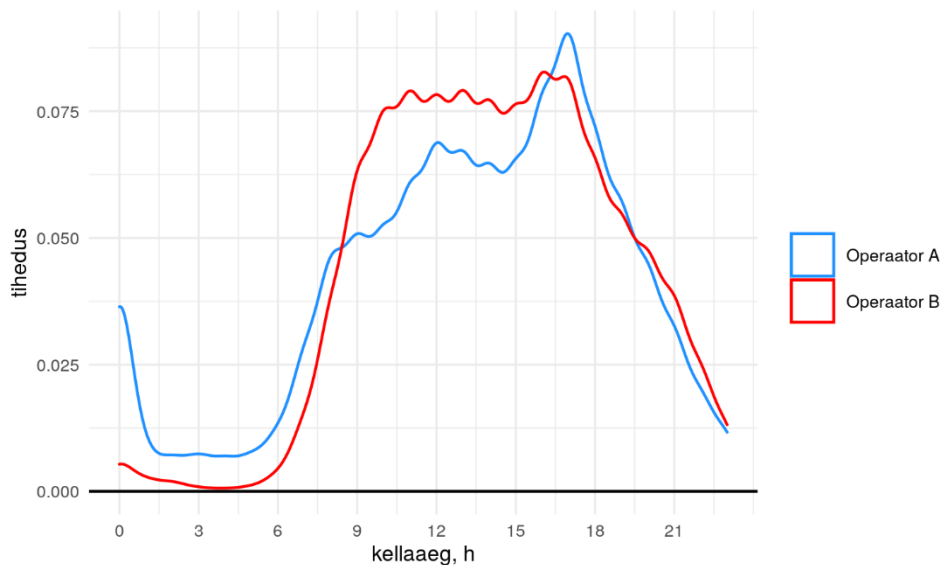
Joonis 13. Keskmine kaugus mobiilmastist asustusüksuse klassi lõikes.

Asustusüksuste hierarhia lõikes rahvastiku paiknemist vaadates näeme, et külades elav rahvastiku keskmine kaugus lähimast mastist on kahe operaatori puhul vahemikus 4,6 ja 5,1 kilomeetrit. Alevikes 1,8 ja 2,4 km ning linnalistes asulates veelgi lähemal, ulatudes Tallinnas (linnaosad) 500 meetri lähedale. 25% rahvastikust elab lähimast mobiilmastist vähem kui 2,6 km kaugusel. Kaugemal kui 10 kilomeetrit jääb elukoht vaid mõnel protsendil Eestis elavatest inimestest.



Joonis 14. Rahvastiku kumulatiivne paiknemine kaugusena lähimast mobiilimastist.

Erindite ja teiste kahtlust äratavate kirjade tuvastamiseks on järgmisele joonisel kantud mõlema mobiiloperaatori kõnetoimingute ööpäevane jaotus.



Joonis 15. Kõnetoimingute jaotus ööpäeva lõikes kahe mobiiloperaatori võrdluses.

Kuigi üldjoontes on mõlema mobiiloperaatori respondentide kõnetoimingute ööpäevane rütm sarnane, päeval ajal on aktiivsus kõrge ja öösel madal, on nähe ka erinevusi. Operaator A puhul on näha suurema kõnetoimingute anomaaliat kesköö kandis. Lähemal vaatlusel selgub, et anomaalia on seotud täpselt südaööl helistamisega. Selle tingib ilmselt mingi tehniline kõnetoiming. Selle moonutuse vältimiseks eemaldatakse kõik 00:00:00 tehtud kõnetoimingud (operatsiooni tulemusena eemaldati operaator A andmestikust 7,8% kõnetoimingutest). Operaator B puhul on 12,3% kirjetest kasutaja identifikaatorita, mistõttu ka need kirjed eemaldati edasisest analüüsist.

Ankurpunktide määramine

Ankurpunktide määramise senise metoodika analüüs ja täiendamise võimalused

Mobiilpositsioneerimise andmete põhjal tähenduslike kohtade ehk ankurpunktide tuvastamine on tähtsal kohal paljudes uuringutes. Käesolevas töös kasutatav metoodika on arendatud Tartu Ülikooli ja OÜ Positium koostöös välja juba üle kümne aasta tagasi (Ahas, Silm, *et al.*, 2010). Passiivse mobiilpositsioneerimise andmed (*Call Detail Record; CDR*) on oma algkujul kõnetoimingute logid, mille alusel mobiiloperaator oma klientidele igakuiseid arveid esitab. Kõige lihtsamal kujul sisaldavad need andmed kasutaja identifikaatorit, kõnetoimingu algusaega ning mobiilantenni kärjekoodi, mis vastavat kõnetoimingu teenindab. Senise praktika kohaselt on Eesti mobiiloperaatorid enne teadlastele andmete edastamist need pseudonümiseerinud. Seejärel edastatakse andmed turvatud andmeside kanalite kaudu anonüümikujul teadlastele. Andmed muutuvad anonüümseks kui need jõuavad teadlaste (ja/või teiste analüütikute) käsutusse. See tähendab, et mobiilsideoperaatoril on võti, millega kindlustatakse, et igal kasutajal on läbi aja kestev unikaalne identifikaator, kuid andmed kustutatakse pärast hoidmiskohustuse lõppu; andmete analüüsijatel seevastu on andmed, aga puudub võti, millega oleks võimalik respondente isikustada.

Elukohaandmed

Käesoleva analüüsi esmasteks lähteandmeteks on kolm erinevat andmestiku. Esimene andmestik sisaldas isikute Rahvastikuregistri (RR) väljavõtet, kus oli märgitud isiku praegune ja endised elukohad ning samuti praegused ja endised lisa-aadressid.

Teine andmestik hõlmas samade isikute passiivse mobiilpositsioneerimise andmeid perioodil 01.02.2017 kuni 01.01.2018 kahelt erinevalt mobiilsideoperaatorilt. Mobiilpositsioneerimise andmestiku põhjal arvutati kõikide osalejate ankurpunktid. Iga ankurpunkt esindas ühe kuu jooksul tehtud kõnetoimingute arvu kindla sidemasti teeniduspiirkonnas. Samuti olid tuvastatud kodu, töö ja teised tähenduslikud kohad.

Kolmas andmestik koosnes samade isikute eneseraporteeritud aadressist, mis saadi uuringusse värbamise ajal läbi viidud küsitluse ja andmete kasutamise nõusoleku sõlmimise ajal. Andmestik oli oluline võrdlemaks RR ja mobiilpositsioneerimise andmeid, eriti kui nimetatud kahe andmestiku vahel esines lahknevusi elukoha määramisel.

Andmete terviklikus

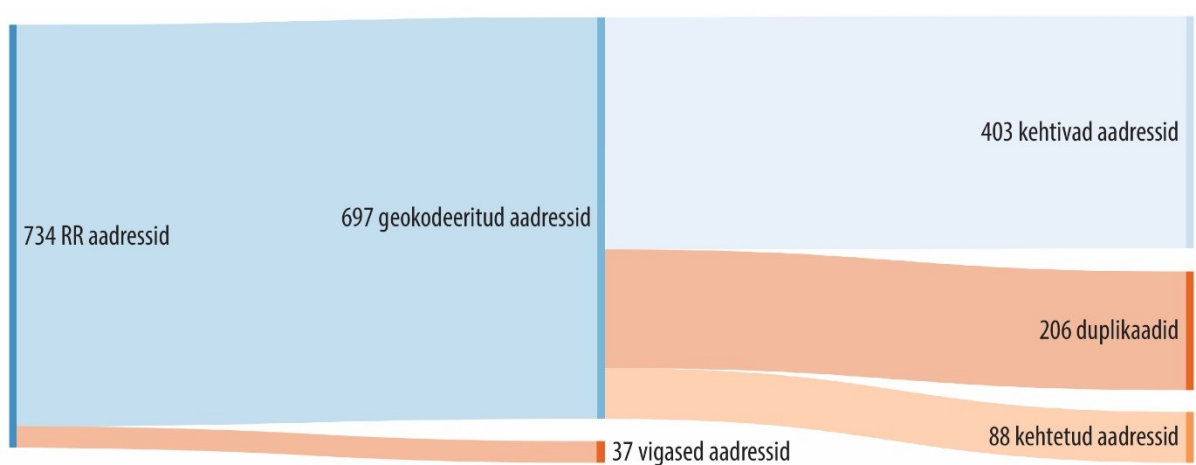
Esindatud isikute arv andmestikes varieerus: RR väljavõttes esines 267 isiku andmed, eneseraporteeritud aadresse oli kogutud 261 isiku ja kodu ankurpunkte 243 isiku kohta. RR väljavõttes olevate elukohtade, kodu ankurpunktide ja eneseraporteeritud elukoha omavaheliseks võrdlemiseks oli vajalik, et kõigis kolmes andmestikus oleks iga isik esindatud. Seetõttu jäid valimisse alles 228 (85%) isiku andmed.

Töövoo kirjeldus

Esimese etapina uuriti lähteandmete kvaliteeti ja eemaldati ebavajalikud või vigased kirjed. Rahvastikuregistri (RR) väljavõttes oli kokku 734 aadressi 267 isiku kohta. Igal isikul oli uuringuperioodi jooksul kehtiv elukoha aadress, üks või mitu kehtivat lisa-aadressi ning endised elukoha ja lisa-aadressid. Aadresside asukoht ruumis määrati koordinaatidega Maa-ameti geokodeerimise teenuse abil. 37 aadressi eemaldati erinevate andmevigade tõttu: kuuel isikul asus lisa-aadress Soomes, üheksal isikul oli lisa-aadressi asemel telefoni number ning nelja isiku lisa-aadressidel ei õnnestunud asukohta määrata (puudus maja number või ei leidnud sellenimelist talu või asus aadressil tühi krunt).

Suur hulk rahvastikuregistri aadressidest olid duplikaadid (30%), mis kustutati analüüsi lihtsustamiseks. Sageli olid isiku elukoht ja lisa-aadress identsed või oli mitu lisa-aadressi kannet, mis

viitasid samale aadressile. Kuna mobiilpositsioneerimise andmed koguti ajavahemiku 01.02.2017 – 01.01.2018 kohta, jäeti alles vaid sellel perioodil kehtivad elukohad ja lisa-aadressid. Tulemusena säilis väljavõttes 403 aadressi 263 isiku kohta (Joonis 16).



Joonis 16. RR väljavõtte aadresside analüüsi etapid ja tulemused.

Kodu ankurpunktide ja elukoha aadressi kattumise uurimine

Kodu ankurpunktide ja elukoha aadressi kattumist uuriti ruumiliselt geograafilise asukoha põhjal. Selleks kasutati kahte erinevat ruumijaotust Eesti territooriumile. Esiteks mobiilsidevõrgu tessellatsiooniga kärgedeks ja teiseks Eesti haldusjaotust omavalitsuse ja asustusüksuse tasanditel.

Passiivse mobiilpositsioneerimise andmestikust võeti välja iga isiku kodu ankurpunkt(id). Nendele omistati koordinaadid sidemasti kärje ID alusel, mis lõi eeldused kodu ankurpunkti(de) võrdlemiseks elukoha aadressiga. Seejärel tuvastati isiku peamine kodu ankurpunkt mastiga ühenduses oldud päevade koguarvu alusel.

Võrdlus Rahvastikuregistriga mobiilsidevõrgu kärgede alusel

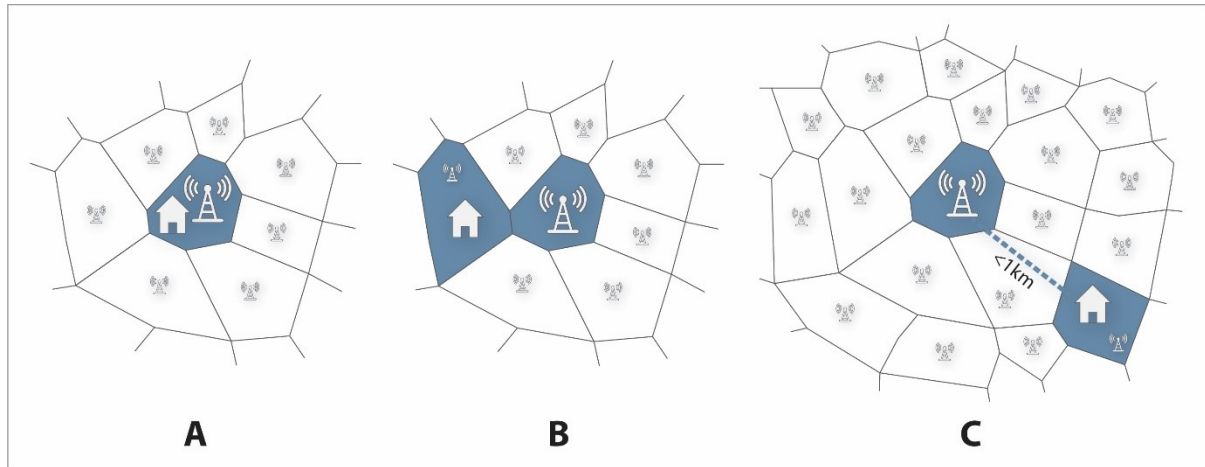
Elukoha andmete ja kodu ankurpunkti(de) kattumise uurimiseks tuvastati esmalt, millises kärjes elukoha aadress asub. Sedasi omistati lisaks kodu ankurpunktile ka igale aadressile kärge-ID tunnus, mis võimaldab isiku kodu ankurpunkti(de) ja elukoha asukoha kattumist uurida. Kui kodu ankurpunkt ja elukoht asusid ühes ja samas kärjes (vt joonis 3, variant A), kinnitas mobiilpositsioneerimine isiku elukoha õigsust.

Asukohtade kattumise uurimist raskendas asjaolu, et mobiiltelefon ei pruugi alati ühenduses olla lähima sidemastiga ning võib asukohta muutmata luua ühenduse teise läheduses oleva sidemastiga. Seda nimetatakse masti viskamiseks (inglise keeles *oscillation*). Seetõttu ei saa pidada elukohta kodu ankurpunktile vastavaks vaid juhul, kui mõlemad asuvad ühes ja samas kärjes. Probleemi lahendamiseks koostati eeskiri, mille alusel määrati, kas kodu ankurpunkt ja elukoht kattuvad. Eeskiri on järgmine:

- kodu ankurpunkt ja elukoht asuvad ühes ja samas kärjes (vt Joonis 17, variant A);
- elukoht asub kodu ankurpunktile piirnevas kärjes (vt Joonis 17, variant B);
- elukoht ei asu kodu ankurpunktile piirnevas kärjes, kuid on eukleidiliselts alla kilomeetri kaugusel kodu ankurpunktist (vt Joonis 17, variant C).

Variant C esines linnades, kus sidemastid asuvad tunduvalt tihedamalt ning seega on variant A ja B esinemine ebatõenäolisem. Juhul, kui peamisele kodu ankurpunktile asus läheduses rohkem kui üks

elukoha või lisa-aadress, võeti arvesse lähim. Seda põhjustas elukoha puhul üksikutel juhtudel elukoha vahetamine uurimisperioodi jooksul.



Joonis 17. Erinevad tingimused, millest lähtuvalt tuvastati kattumist kodu ankurpunktiga ja elukoha vahel.

Kuna lähteandmed sisaldasid kolme tüüpi erinevaid elukoha andmeid (RR järgne elukoht, RR järgsed lisa-aadressid ja eneseraporteeritud aadress), võrreldi elukoha andmete kattumist kodu ankurpunktidega etappides.

Esimese etapina uuriti isiku RR järgse elukoha kattumist peamise kodu ankurpunktiga. Nimetatud andmete ruumilise kattumise puhul ei uuritud antud isikute ülejäänud elukoha andmete kattumist kodu ankurpunktidega, kuna mobiilpositsioneerimise andmed kinnitasid isiku elukoha õigsust.

Teise etapina uuriti nende isikute lisa-aadresside kattumist peamise kodu ankurpunktiga, kelle peamine koduankurpunkt ei vastanud RR järgsele elukohale esimeses etapis.

Kolmanda etapina uuriti RR elukoha andmete kattumist sekundaarsete kodu ankurpunktidega nende isikute puhul, kelle peamine koduankurpunkt ei vastanud RR järgsele elukohale esimeses etapis ja kellel esines rohkem kui üks kodu ankurpunkt. Kui selles etapis ei esinenud samuti mingit kattuvust RR elukoha andmete ja kodu ankurpunktide vahel, järeldati, et mobiilpositsioneerimise andmed ei kinnita isiku RR elukoha andmeid.

Neljanda etapina uuriti nende isikute eneseraporteeritud elukoha andmete kattumist kodu ankurpunktidega, kelle peamine koduankurpunkt ei vastanud RR järgsele elukohale esimeses etapis.

Võrdlus Rahvastikuregistriga haldusjaotuse alusel

Elukoha ja mobiilpositsioneerimise andmete ruumilist kattumist uuriti lisaks mobiilsidevõrgu kärgedele ka haldusjaotuse alusel. Esmalt uuriti, kas isiku elukoht või lisa-aadressid asuvad peamise kodu ankurpunktiga samas asustusüksuses või omavalitsuses või alla kilomeetri kaugusel. Nende isikute puhul, kelle peamine kodu ankurpunkt ei asunud samas haldusüksuses elukohaga, uuriti, kas mõni sekundaarne ankurpunkt asus. Isikute puhul, kelle peamine kodu ankurpunkt ei asunud elukohaga samas haldusüksuses, korraldati kirjeldatud protsessi eneseraporteeritud elukohaga.

Tulemused

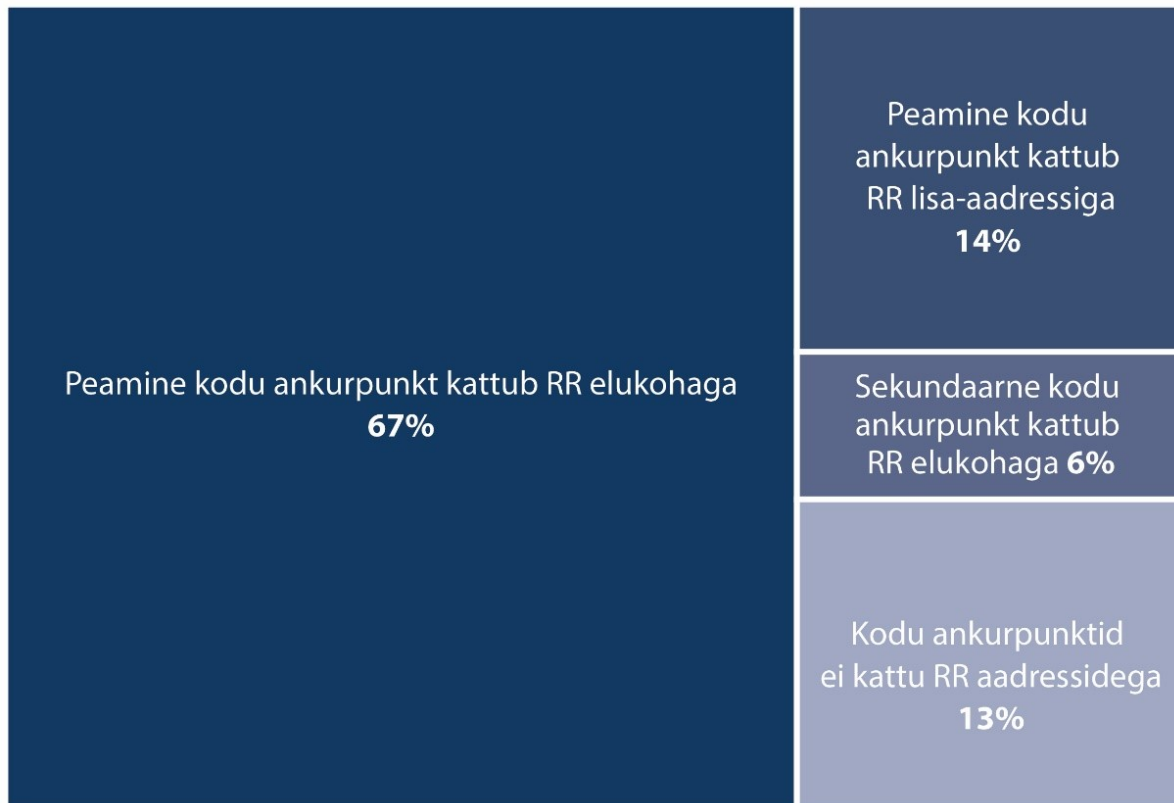
Kodu ankurpunktide kattumine RR addressidega

Mobiilsidevõrgu kärgede alusel

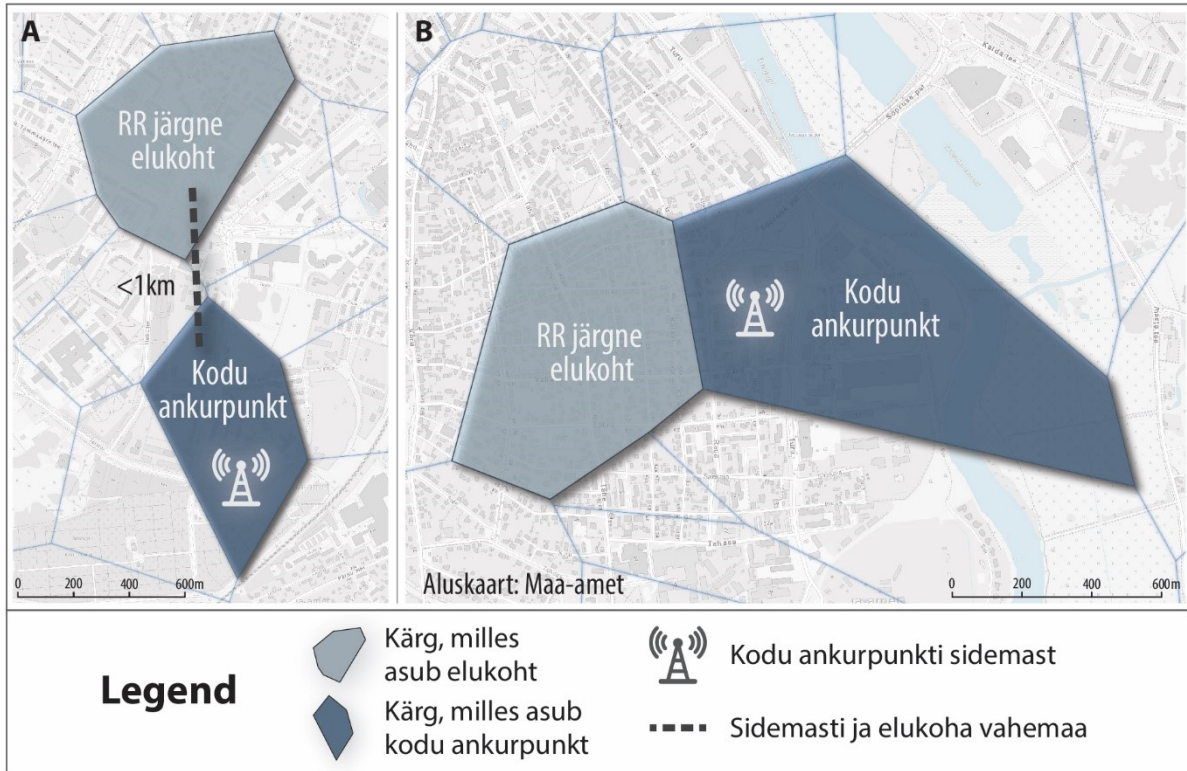
Esimeses etapis uuriti isiku RR järgse elukoha ruumilist kattumist peamise kodu ankurpunktiga. Tulemused on esitatud Tabel 1 ja Joonis 18. 153 (67%) isikul kattusid asukohad eelmises peatükis kirjeldatud meetodikat rakendades (Joonis 19). Nende isikute elukoha andmeid ei olnud vaja edasi uurida, kuna mobiilpositsioneerimise andmed kinnitasid isiku RR järgset elukohta. Ülejäänud 75 (33%) isiku elukoha andmeid uuriti edasi teises etapis.

Tabel 1. Kodu ankurpunktide ja RR aadresside kattuvus mobiilsidevõrgu kärgede alusel.

| | Isikute arv | Osakaal |
|--|-------------|---------|
| Peamine kodu ankurpunkt kattub RR järgse elukohaga | 153 isikut | 67% |
| Peamine kodu ankurpunkt kattub RR järgse lisa-aadressiga | 32 isikut | 14% |
| Sekundaarne kodu ankurpunkt kattub RR järgse elukohaga | 13 isikut | 6% |
| Kodu ankurpunktid ei kattu RR aadressidega | 30 isikut | 13% |

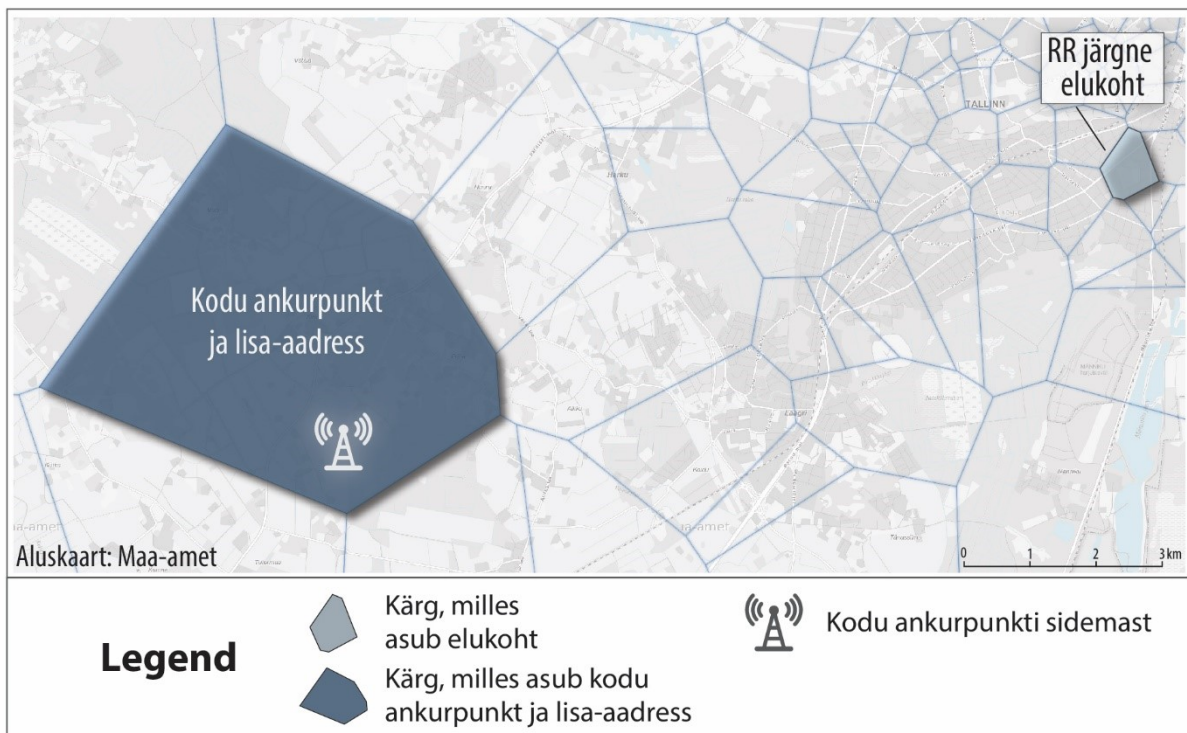


Joonis 18. RR aadresside kattumine kodu ankurpunktidega.



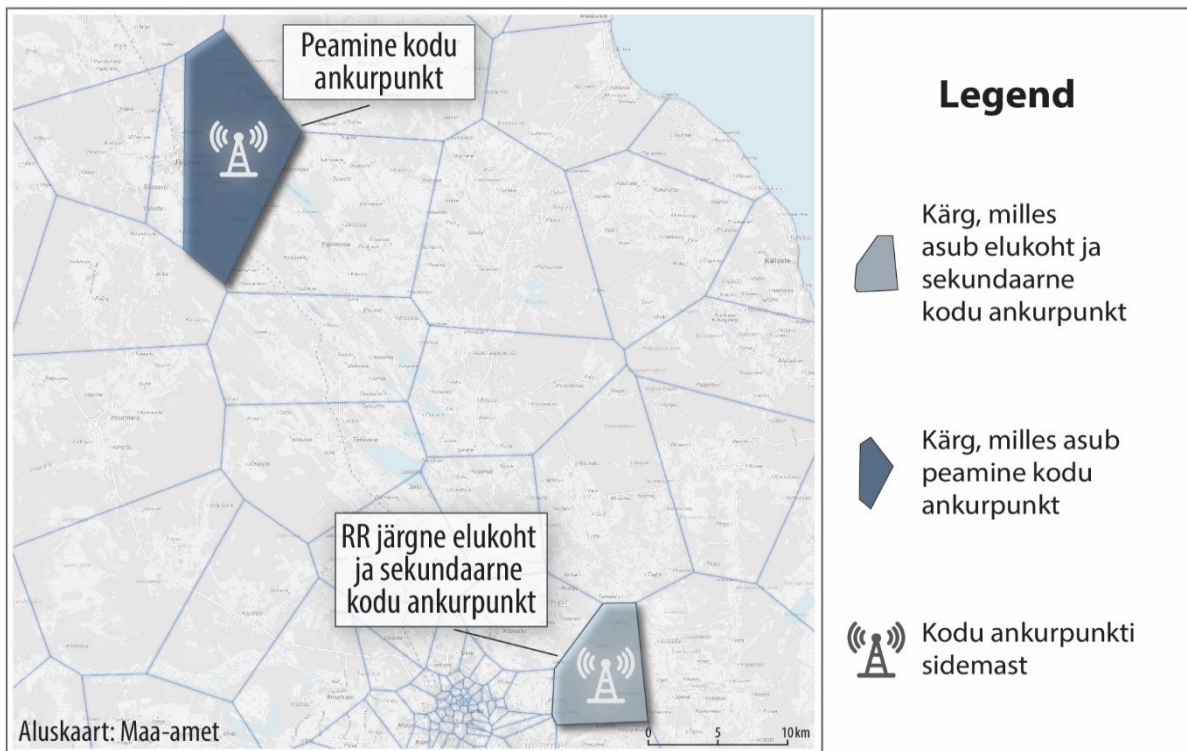
Joonis 19. Visuaalne näide kodu ankurpunkti kattumisest RR järgse elukohaga.

Teises etapis uuriti nende isikute lisa-aadresside kattumist peamise kodu ankurpunktiga, kelle RR järgne elukoht ei kattunud peamise kodu ankurpunktiga (Joonis 20). 32 isikul (14%) kattus lisa-aadress peamise kodu ankurpunktiga.



Joonis 20. Visuaalne näide kodu ankurpunkti kattumisest RR järgse lisa-aadressiga.

Kolmandas etapis uuriti isikuid, kellel oli rohkem, kui üks kodu ankurpunkt. Kui nende peamine kodu ankurpunkt ei vastanud elukohale, uuriti, kas sekundaarne kodu ankurpunkt vastas (Joonis 21). 13 (6%) isikul vastas sekundaarne ankurpunkt elukohale.



Joonis 21. Visuaalne näide sekundaarse kodu ankurpunkti kattumisest RR järgse elukohaga.

Ükski kodu ankurpunkt ei vastanud ühelegi RR aadressile 30 (13%) isikul ehk ei esinenud mingit kattumist mobiilpositsioneerimise andmete ja RR andmete vahel.

Haldusjaotuse alusel

Asustusüksuste ja omavalitsuste alusel uuriti, kas isiku kodu ankurpunktid ja RR aadressid asuvad samas omavalitsuses või asustusüksuses. Tulemused on esitatud Tabel 2.

Tabel 2. Kodu ankurpunktide ja RR aadresside kattuvus haldusjaotuse alusel.

| | Omavalitsuse tasandil | | Asustusüksuse tasandil | |
|--|-----------------------|---------|------------------------|---------|
| | Isikute arv | Osakaal | Isikute arv | Osakaal |
| Peamine kodu ankurpunkt kattub RR järgse elukohaga | 174 isikut | 76% | 143 isikut | 63% |
| Peamine kodu ankurpunkt kattub RR järgse lisa-aadressiga | 22 isikut | 10% | 32 isikut | 14% |
| Sekundaarne kodu ankurpunkt kattub RR järgse elukohaga | 11 isikut | 5% | 8 isikut | 4% |
| Kodu ankurpunktid ei kattu RR aadressidega | 24 isikut | 11% | 45 isikut | 20% |

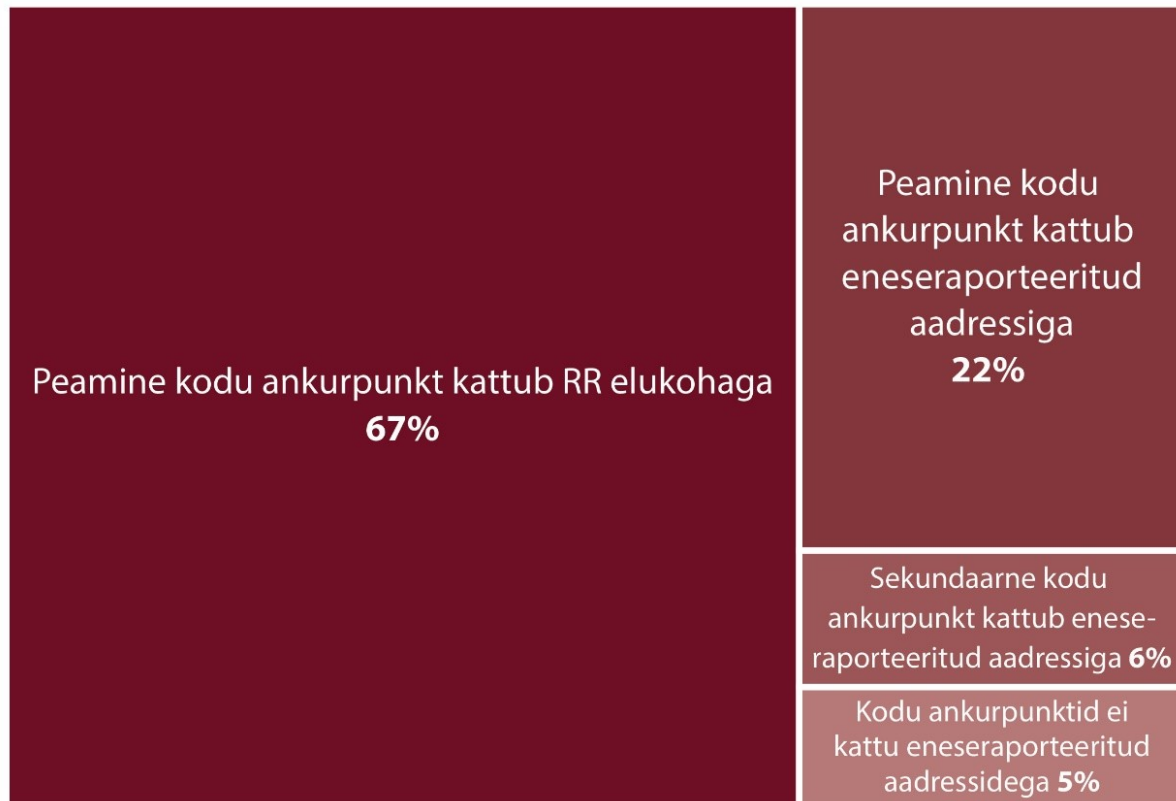
Kodu ankurpunktide kattumine eneseraporteeritud aadressiga

Mobiilsidevõrgu kärgede alusel

Eneseraporteeritud elukohta kattumist peamise kodu ankurpunktiga uuriti nende isikute puhul, kelle RR järgne elukoht ei kattunud peamise kodu ankurpunktiga esimeses etapis ehk 81 isiku puhul. Tulemused on esitatud Tabel 3 ja Joonis 22. 51 (22%) isikul kattus peamine koduankurpunkt eneseraporteeritud aadressiga. 13 (6%) isikul kattus sekundaarne kodu ankurpunkt eneseraporteeritud aadressiga. Seega ei kattunud 11 isikul ükski kodu ankurpunkt eneseraporteeritud aadressiga. Ühegi RR aadressi ega eneseraporteeritud aadressiga ei kattunud ükski kodu ankurpunkt 9 isiku puhul.

Tabel 3. Kodu ankurpunktide kattuvus eneseraporteeritud aadressiga.

| | Isikute arv | Osakaal |
|--|-------------|---------|
| Peamine kodu ankurpunkt kattub RR järgse elukohaga | 153 isikut | 67% |
| Peamine kodu ankurpunkt kattub eneseraporteeritud aadressiga | 51 isikut | 22% |
| Sekundaarne kodu ankurpunkt kattub eneseraporteeritud aadressiga | 13 isikut | 6% |
| Kodu ankurpunktid ei kattu eneseraporteeritud aadressiga | 11 isikut | 5% |



Joonis 22. Eneseraporteeritud aadresside kattumine kodu ankurpunktidega.

Haldusjaotuse alusel

Asustusüksuste ja omavalitsuste alusel uuriti, kas isiku kodu ankurpunktid ja eneseraporteeritud aadress asuvad samas omavalitsuses või asustusüksuses. Tulemused on esitatud Tabel 4.

Tabel 4. Kodu ankurpunktide kattuvus eneseraporteeritud aadressiga haldusjaotuse alusel.

| | Omavalitsuse tasandil | | Asustusüksuse tasandil | |
|--|-----------------------|---------|------------------------|---------|
| | Isikute arv | Osakaal | Isikute arv | Osakaal |
| Peamine kodu ankurpunkt kattub RR järgse elukohaga | 174 isikut | 76% | 143 isikut | 63% |
| Peamine kodu ankurpunkt kattub eneseraporteeritud aadressiga | 33 isikut | 14% | 46 isikut | 20% |
| Sekundaarne kodu ankurpunkt kattub eneseraporteeritud aadressiga | 9 isikut | 4% | 10 isikut | 4% |
| Kodu ankurpunktid ei kattu eneseraporteeritud aadressiga | 12 isikut | 5% | 29 isikut | 13% |

Rahvastikuregistri (RR) elukoha ja mobiilpositsioneerimise andmete kattuvus

Uurimistulemustest selgus (Tabel 5), et sidemasti kärgede põhjal vastas 67% isikutest mobiilpositsioneerimise kaudu tuvastatud peamine kodu ankurpunkt RR järgsele elukohale. Seega saab järeldada, et ligi kahel kolmandikul inimestest vastab RR järgne elukoht nende tegelikku elukohale. Kolmandikul inimestest ei kattu RR elukoha ja mobiilpositsioneerimise andmed eeskujulikult, kuid uurides lisaks nende RR lisa-aadresse ja sekundaarseid kodu ankurpunkte, on võimalik 87% puhul kattumisi andmestike vahel tuvastada.

Asustusüksuste põhjal analüüsid ilmnis sarnane tulemus: 63% vastas peamine kodu ankurpunkt RR järgsele elukohale ja 80% kattusid andmestikud mingil määral (vähemalt üks kodu ankurpunkt asus RR järgse aadressiga samas asustusüksuses). Omavalitsuse tasandil kattusid andmestikud ootuspäraselt paremini, kuna tegu on suurema pindalaga ruumiüksustega: 76% vastas peamine kodu ankurpunkt RR järgsele elukohale ja 89% kattusid andmestikud mingil määral.

Tabel 5. Peamise kodu ankurpunkti kattumine RR aadressidega.

| | Kärgede põhjal | | Omavalitsuse põhjal | | Asustusüksuse põhjal | |
|--|----------------|---------|---------------------|---------|----------------------|---------|
| | Isikute arv | Osakaal | Isikute arv | Osakaal | Isikute arv | Osakaal |
| Peamine kodu ankurpunkt kattub RR järgse elukohaga | 153 isikut | 67% | 174 isikut | 76% | 143 isikut | 63% |
| Peamine kodu ankurpunkt kattub RR järgse lisa-aadressiga | 32 isikut | 14% | 22 isikut | 10% | 32 isikut | 14% |
| Peamine kodu ankurpunkt ei kattu RR aadressidega | 43 isikut | 19% | 32 isikut | 14% | 53 isikut | 23% |

Ruumijaotusele vastavalt 10–14% ei kattunud peamine kodu ankurpunkt RR elukohaga vaid isiku lisa-aadressiga. Selle põhjal saab eeldada, et isik elab enamuse ajast lisa-aadressina märgitud elukohas ja RR järgses elukohas veedab vähem aega. Selline muster on iseloomulik näiteks tudengitele, kes veedavad suurema osa ajast ülikoolilinnas ja vähem aega oma kodukohas. Samuti on muster iseloomulik inimestele, kellel on mitu kodu, näiteks juhul kui veedetakse enamuse ajast suvekodus.

4–6% ei vastanud peamine kodu ankurpunkt ühelegi RR adressile, kuid sekundaarne kodu ankurpunkt vastas RR elukoha adressile (Tabel 6). Võime eeldada, et need isikud ei ole oma peamist elukohta Rahvastikuregistrisse kandnud.

Tabel 6. Sekundaarsete kodu ankurpunktide kattumine RR adressidega.

| | Kärgede põhjal | | Omavalitsuse põhjal | | Asustusüksuse põhjal | |
|--|----------------|---------|---------------------|---------|----------------------|---------|
| | Isikute arv | Osakaal | Isikute arv | Osakaal | Isikute arv | Osakaal |
| Sekundaarne kodu ankurpunkt kattub RR järgse elukohaga | 13 isikut | 6% | 11 isikut | 5% | 8 isikut | 4% |
| Ükski kodu ankurpunkt ei kattu RR adressidega | 30 isikut | 13% | 24 isikut | 11% | 45 isikut | 20% |

Olenevalt ruumijaotusest ei tuvastatud kattumist RR ja mobiilpositsioneerimise andmete vahel 11–20% puhul. Saame järeldada, et nende isikute RR andmed on aegunud ja ei vasta tegelikkusele.

Tulemustest ilmneb tõsiasi, et 33% isikutest ei kattu elukoha ja mobiilpositsioneerimise andmed eeskujulikult ehk teisisõnu ei kattu isiku peamine kodu ankurpunkt RR järgse elukohaga. Seejuures võib rolli omada passiivse mobiilpositsioneerimise ja kodu ankurpunktide tuvastamise mõningane ebatäpsus, kuid suurema osa tingib valede elukoha andmete olemasolu Rahvastikuregistris.

Eneseraporteeritud elukoha ja mobiilpositsioneerimise andmete vastavus

RR elukoha ja mobiilpositsioneerimise andmete erinevuste uurimiseks oli kasutada isikute eneseraporteeritud elukoha andmeid. See lõi võimaluse uurida, kas isiku eneseraporteeritud elukoht vastas kodu ankurpunktile juhul kui RR elukoha andmed ei vastanud, tõstes seeläbi mobiilpositsioneerimise usaldusväärsust. Siinjuures tasub meeles pidada, et RR andmed on oma olemuselt samuti eneseraporteeritud. Andmestikes olevad erinevused tulenevad sellest, et andmed on kogutud eri aegadel.

Seega uuriti, kas isiku peamine ankurpunkt kattub eneseraporteeritud aadressiga juhul kui RR järgse elukoha aadressiga kattumist ei esinenud. Kuna mobiilpositsioneerimise andmed ja eneseraporteeritud elukoha andmed on kogutud samal aja perioodil, võib eeldada, et elukoha andmed kattuvad paremini RR andmetega võrreldes. Tulemustest selgus (vt tabel 7), et 14-22% isikutest kattus peamine kodu ankurpunkt eneseraporteeritud aadressiga, 4-6% sekundaarne kodu ankurpunkt ja 5-13% ei kattunud ükski ankurpunkt. Seega saab järeldada, et mitmed isikud on eneseraporteeritud elukoha samuti valesti esitanud.

Tabel 7. Kodu ankurpunktide kattumine eneseraporteeritud aadressidega.

| | Kärgede põhjal | | Omavalitsuse põhjal | | Asustusüksuse põhjal | |
|--|----------------|---------|---------------------|---------|----------------------|---------|
| | Isikute arv | Osakaal | Isikute arv | Osakaal | Isikute arv | Osakaal |
| Peamine kodu ankurpunkt kattub eneseraporteeritud aadressiga | 51 isikut | 22% | 33 isikut | 14% | 46 isikut | 20% |
| Sekundaarne kodu ankurpunkt kattub eneseraporteeritud aadressiga | 13 isikut | 6% | 9 isikut | 4% | 10 isikut | 4% |
| Ükski kodu ankurpunkt ei kattu eneseraporteeritud aadressiga | 11 isikut | 5% | 12 isikut | 5% | 29 isikut | 13% |

Metoodika rakendamine

Arvestades mobiilsidevõrgu keerulist infrastruktuuri ja mastidele arvutatud teoreetilisi levialasid, mis võivad reaalsusega võrreldes märgatavalt erineda, olid tulemused üllatavalt täpsed. Tulemusi visuaalselt kontrollides oli selgelt märgata, et metoodika toimis ettenähtud viisil ja tulemused olid ootuspärased. Metoodikat oleks seega võimalik rakendada ulatuslikumalt Rahvastikuregistri elukoha andmete võrdlemisel mobiilpositsioneerimise andmetega. Metoodikat saaks täiendada rakendades kolme tasemelist “valgusfoori” süsteemi, kus usaldusväärne kattuvus elukoha ja mobiilpositsioneerimise andmete vahel oleks “roheline”, ülevaatamist ja põhjalikumalt uurimist vajav tulemus “kollane” ja andmete selged erinevused “punane”. Selline süsteem lihtsustaks tunduvalt RR elukoha andmete vastavuse kontrollimist tegelikkusele.

Uurimusest on ilmne, et mobiilpositsioneerimise andmed on suureks abiks isiku elukoha uurimisel, eriti kui registrites on tegu vastuoluliste andmetega. Näiteks, kui isiku elukoht RR andmetel asub ühes asukohas, kuid eneseraporteeritud aadress asub teises asukohas. Kodu ankurpunktide abil on sel juhul võimalik uurida, kumb on inimese peamine elukoht, ning kas inimene ka nimetatud elukohtades üldse viibib.

Passiivse mobiilpositsioneerimise andmestik on ulatuslik ning pidev, mis võimaldab isiku liikumisi uurida pika ajaperioodi jooksul usaldusväärse tulemuse saavutamiseks. Samuti võimaldab mobiilpositsioneerimine uurida lühemaajalisi ja sesoonseid liikumismustreid, mis Rahvastikuregistris esile ei tule. Näiteks on võimalik mobiilpositsioneerimise abil tuvastada tudengite sesoonseid elukoha vahetumisi kodulinna ja ülikoolilinna vahel.

Mobiilpositsioneerimise rakendamisega kaasneb ilmselt mitmeid probleeme. Esimene probleemide ring on seotud andmete kättesaadavusega. Pärast ELi andmekaitse üldmääruse rakendamist on enamasti mobiilioperaatoreid Eestis olnud väga tõrjuvad, et väljastada uuringuteks passiivse mobiilpositsioneerimise nõu toorandmeid. Kui esimestel aastatel oli peaarvumendiks, et seda ei tohi teha sideseadusest lähtuvalt, siis täna on need regulatsioonid kehtestanud mobiilioperaatorid ise. Kasvanud on ka operaatorite vastava valdkonna ärihuvid ning arendatakse oma tooteid (nt Telia Crowd Insights). Enamasti pakutakse andmeid, kus mastide asukohta on üldistatud ja kasutaja identifikaator uueneb iga 24–48 tunni pärast. Selline andmestik teeb aga longituudsed ankurpunktide kontseptsioonil põhinevad uuringud ja analüüsid võimatuks. Musta kasti tüüpi karbitooted ei saa ka olla aluseks riiklikule statistikale, samuti tekib nii vastava analüüsimeetodika tootmise operaatoripoolne monopol ning erinevate mobiilioperaatorite mobiilsusandmete võrdlus kujuneb väga keeruliseks kui mitte võimatuks. Ilmselt on naiivne loota, et mobiilioperaatorid ise tahavad sotsiaalset vastutust tundes avatumaks muutuda ning lihtsaimaks lahenduseks riigi poolt vaadatuna on seadusandluses muudatuste tegemiseks, mis kuulutab mobiilpositsioneerimise andmed (*CDR*, *DDR*, *signaling*) suure avaliku huviga andmeteks kohustades operaatoritel võimaldada andmetele ligipääsu. Oluline on, et protsess toimuks kooskõlastatult erinevate huvitatud osapooltega, koostatakse selge raamistik nii andmekaitse kui ka -analüüsi standardite ning nõuete näol.

Mobiilpositsioneerimise rakendatavus rahvastiku paiknemise ja liikuvuse statistika tootmiseks

Kui eespool tutvustatud mobiilpositsioneerimise andmestiku kasutamise roll elukohaandmete täpsustamiseks saab ilmselt olla eelkõige registri ja tegeliku elukoha lahknevusi kontrolliv ja mitme elukohaga inimesi tuvastav, siis rahvastiku paiknemise ja mobiilsuse statistika tootmiseks on tegemiseks hindamatu andmeallikaga. Kui konkreetse isiku puhul on oluline, et tema elukoht oleks fikseeritud eluruumi täpsusega, siis rahvastiku või selle väiksemate sotsiaalsete rühmade paiknemise- ja mobiilsusmustrite kirjeldamisel võime modelleerimisprotsessis inimesi ka „tükeldada“ (0,75 elab

Kesklinnas, 0,25 elab Südalinnas) ning asumipõhiselt saame tegelikkust adekvaatselt ja objektiivselt kirjeldavad numbrid. Mis sobivad suurepäraseks andmesisendiks erinevate teenuste planeerimisel ja pakkumisel.

Lisaks on oluline ka defineerida, millises ajalisel aknas täpset elukohta soovitakse. Kui me tahame elukohti teada aasta täpsusega, siis tekivad ühed nõudmised, kui aga soovime teada elukoha infot kuu täpsusega, siis on nõuded ning tegelikkuse ja registrite lahknevus märksa suurem. Elukohavahetus pole enamasti diskreetse ajahetkega seotud sündmus vaid pikema aja peale jaotunud protsess nii et mingitel perioodidel ei oska ilmselt inimene isegi öelda, kus ta elukoht asub. Kui sellele lisada nädalase ja kuise tsükliga tööranne ning sesoonne teine kodu, siis tulebki tõdeda, et peale selle, et mobiilpositsioneerimise täpsusel on piirid ees, on seda ka väga kõrge täpsuse (nii ajas kui ruumis) saavutamine rahvastikuregistri puhul.

Eesti Töötukassa tööturuteenuste ja -toetuste infosüsteemi andmete kasutamise võimalused inimeste elukohaandmete täpsustamiseks

EMPIS ehk Eesti Töötukassa tööturuteenuste ja -toetuste infosüsteemi pidamise eesmärk on tagada Eesti Töötukassale ja Sotsiaalministeeriumile neile „Tööturuteenuste ja -toetuste seaduse“, „Välismaalaste seaduse“, „Töötuskindlustuse seaduse“, „Sotsiaalmaksuseaduse“ ning Euroopa Liidu õigusaktide, välislepingute ning riiklike programmide ja projektidega pandud ülesannete täitmiseks vajalike andmete olemasolu ning tagada Töötukassa organisatsioonile efektiivsed menetlusprotsessid ([link](#)). Vastavalt registri põhimäärusele ([link](#)) esitavad sinna andmeid:

- 1) töötu;
- 2) tööotsija;
- 3) tööturuteenuse osutamist taotlenud ja tööturuteenust saanud isik, kes ei ole töötuna ega tööotsijana arvele võetud (edaspidi muu tööturuteenust saanud isik);
- 4) tööandja;
- 5) tööturuteenuse osutaja;
- 6) töötukassa;
- 7) tuumakatastroofi, tuumakatsetuse või aatomielektijaama avarii tagajärgede likvideerimisest osa võtnud mittetöötav isik, kelle eest töötukassa maksab sotsiaalmaksu (edaspidi sotsiaalmaksu maksmist taotlenud isik).

Muude andmete hulgas talletatakse registris subjekti elukoha aadressiandmed ja kontaktandmed (sidevahendi liik ja number ning e-posti aadress). Kuigi määrus seda ei sätesta, on väljastatud andmetes näha ka iga inimese postiaadress. Lisaks võib EMPIS-es olla ühe isiku kohta mitu kirjet.

Algselt 35 kirjet, 11 respondendi kohta. Valimist 5%. Ainuüksi see fakt näitab, et tegemist on potentsiaalselt väheolulise andmeallikaga elukohaandmete täpsustamiseks.

EMPIS- registris olevast 11 uuringu osaleja puhul on registris 1 elukohta ja 1 postiaadress. Nelja osaleja puhul on registris 2 elukohta ja 2 postiaadressi. EMPIS-e 35 aadressist on 22 ADS-OID klassifikatsiooni järgi eluruumides ja 13 elukondlikes hoonetes. 8 osaleja puhul on andmed olemas nii Rahvastikuregistrist, EMPIS-est kui ka mobiilpositsioneerimisest.

EMPIS-e elukoha ja postiaadressid olid identsed kõigil vaatlusalustel juhtudel.

Registrite võrdlus viidi läbi erinevatel tasemetel. ADS_OID, asustusüksus, omavalitsus, maakond. Kuna kattuvate andmeallikatega osalejaid jäi alles vaid 8, siis vaadatakse järgnevalt kõiki osalejaid eraldi (isikud on kodeeritud tähtedena: A, B, C, D, E, F, G, H).

Isikul A on EMPIS-baasis 2 aadressi (elukoht + postiaadress). Mõlemad neist kattuvad rahvastikuregistri elukoha täpse aadressiga. Respondent A kohta puuduvad mobiilpositsioneerimise andmed.

Isikul B on EMPIS-es 5 aadressi (2 elukohta + 3 postiaadressi). 1 elukoht ja 2 postiaadressi kehtivad perioodil 26.07.2011-19.06.2017. Teised (elukoht + postiaadress) hakkavad kehtima alates 19.06.2017. Neist esimese perioodi elukoht ja üks postiaadress (asuvad samal aadressil) ei kattu rahvastikuregistriga, üks postiaadressi kattub rahvastikuregistri ADS-ROID-ga. Hilisema perioodi elukoht ja postiaadress kattuvad rahvastiku registri ADS-ROID'ga ja ka varasema perioodi ühe postiaadressi ADS-ROID'ga. Seega on EMPIS-aadressil täpne kattuvus rahvastikuregistri elukoha aadressiga olemas. Vastavalt mobiilpositsioneerimise andmetele asuvad selle osaleja kodud Pärnu linnas, Rae vallas ja Viimsi vallas, neist kõige olulisem on koduankurpunkt aga Pärnu linnas, mis ei kattu ei rahvastikuregistri ega ka EMPIS-e andmetega.

Isikul C on EMPIS baasis kaks erinevat elukoha ja postiaadressi, samuti on tal ka kaks rahvastikuregistri aadressi. See tähendab, et aadressid on uuritava perioodil muutunud mõlemas registris. Seejuures ei kattu uuritava perioodi mõlema registri esimesed elukoha-aadressid ka maakonna tasemel. Samas kattub asustusüksuse tasemel uuritava perioodi esimene EMPIS-e elukoha-aadress rahvastikuregistri teise aadressiga. Täielik ADS-ROID tasemel on mõlemas registris kattuvus EMPIS-e esimese postiaadressi, EMPIS-e teise elukoha-aadressiga ja rahvastikuregistri teise elukoha-aadressiga. Tegemist tundub olevat kõige keerulisema juhtumiga, kus uuringus osalenud isik on teadlikult vahetanud elu- ja postiaadresse kahe erineva aadressiobjekti vahel. Kas osaleja C puhul ei kattu mobiilpositsioneerimise alusel leitud elukoht (Juuru alevik) ja teised ankurpunktid ühegi RR ja EMPIS-e aadressiga.

Isikul D langevad rahvastikuregistri elukohaga täielikult kokku nii EMPIS-e elukoha- kui ka postiaadress. Samas on mobiilpositsioneerimise järgne alukoht hoopis Pärnu linnas.

Isikul E on EMPIS-e registris uuritava perioodil kaks erinevat aadressi, rahvastikuregistris aga üks elukoha aadress. Kõik kombinatsioonid kattuvad vähemalt asustusüksuse tasemel (Paide linn). Täielikult kattuvad rahvastikuregistri aadressiga EMPIS-e esimene postiaadress ja teise perioodi elukoha ja postiaadress. Uuritava perioodi mobiilpositsioneerimise põhised elukoha aadressid (2 tk) on aga tuvastatud hoopis Anija vallas. Lisaks on tuvastatud ankurpunkte veel Pärnu linnas, Võru ja Saue vallas, mis samuti ei kattu rahvastikuregistri ja EMPIS-e baasiga.

Isikul F kattuvad täielikult rahvastikuregistriga nii EMPIS-e elukoha kui ka postiaadress, st kõik on samal aadressil. Selle osaleja puhul kattub rahvastikuregistri ja EMPIS-e elukoha aadress mobiilpositsioneerimise abil leitud elukohaga. Osaleja väga selgele ühele kodule viitab asjaolu, et kõik koduankrud on määratud läbivalt Rapla linna piiridesse.

Isikul G ei kattu ka maakonna tasemel rahvastikuregistriga ükski EMPIS-e aadress. Samas kattuvad mobiilpositsioneerimise ja rahvastikuregistri järgne püsiva elukoha aadress (Pirita linnaosa).

Isikul H kattuvad asustusüksuse tasemel kõik EMPIS-e aadressid rahvastikuregistriga. Täielikult ei kattu rahvastikuregistri elukoha ja EMPIS-e teine postiaadress. Mobiilpositsioneerimise elukoht on määratud Narva-Jõesuu linna ja rahvastikuregistri elukoht Narva linna. Samas on antud osaleja tegevusruum piirdunud uuritava perioodi vältel väga kompaktselt Narva ja Narva-Jõesuu linnas. Vastavalt

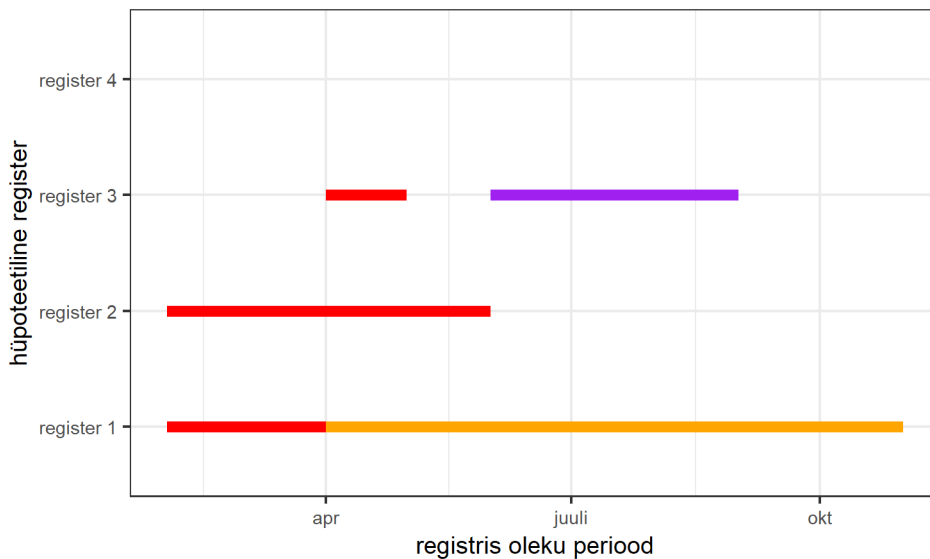
algoritmile on tuvastatud koduankrud mõlemasse linna. Ilmselt võib väita, et kuigi osaleja veetis märkimisväärse osa ajast mõlemas omavalitsus, on tema rahvastikuregistri ja EMPIS-e elukohta aadressid korrektsed.

EMPIS andmestiku kasutamise potentsiaal elukohta andmete täpsustamisel on madal. Selle peamiseks põhjuseks on ühiskonna vähene kaetus (käesoleva uuringu valimist vaid 5%). Lisaks puudub põhjuslik alus võtta vastus otsus, et kui erinevate registrite elukohta aadressid kuhjuvad samasse kohta, siis kasvab ka tõenäosus, et just seal asub uuritava inimese tegelik püsielukoht. EMPIS-e, rahvastikuregistri ja mobiilpositsioneerimise alusel määratud elukohtade lahknevust iseloomustab järgnev joonis (Joonis 23).

| EMPIS | | | | | | rahvastikuregister | | | | | | mobiilpositsioneerimine | | | |
|---------|------------|------------|---------|-------------|-------------------|--------------------|----------|-----------------|----------------|---------|-------------|-------------------------|---------|-------------|--------------|
| osaleja | algus | lõpp | maakond | omavalitsus | aadressi tüüp | tüüp | kehtivus | kehtivuse algus | kehtivuse lõpp | maakond | omavalitsus | asustusüksus | maakond | omavalitsus | asustusüksus |
| C | 26.01.2016 | 10.08.2017 | Tartu | Tartu linn | Tartu linn ELUK_1 | ELUKOHT | ENDINE | 03.05.2016 | 27.11.2017 | Võru | Võru vald | Puiga küla | Rapla | Rapla vald | Juuru alevik |
| C | 26.01.2016 | 10.08.2017 | Tartu | Tartu linn | Tartu linn ELUK_1 | ELUKOHT | ENDINE | 28.11.2017 | 14.04.2020 | Tartu | Tartu linn | Tartu linn | Rapla | Rapla vald | Juuru alevik |
| C | 26.01.2016 | 10.08.2017 | Tartu | Tartu linn | Tartu linn POST_1 | ELUKOHT | ENDINE | 03.05.2016 | 27.11.2017 | Võru | Võru vald | Puiga küla | Rapla | Rapla vald | Juuru alevik |
| C | 26.01.2016 | 10.08.2017 | Tartu | Tartu linn | Tartu linn POST_1 | ELUKOHT | ENDINE | 28.11.2017 | 14.04.2020 | Tartu | Tartu linn | Tartu linn | Rapla | Rapla vald | Juuru alevik |
| C | 26.01.2016 | 10.08.2017 | Tartu | Tartu linn | Tartu linn POST_2 | ELUKOHT | ENDINE | 03.05.2016 | 27.11.2017 | Võru | Võru vald | Puiga küla | Rapla | Rapla vald | Juuru alevik |
| C | 26.01.2016 | 10.08.2017 | Tartu | Tartu linn | Tartu linn POST_2 | ELUKOHT | ENDINE | 28.11.2017 | 14.04.2020 | Tartu | Tartu linn | Tartu linn | Rapla | Rapla vald | Juuru alevik |
| C | 10.08.2017 | NA | Tartu | Tartu linn | Tartu linn ELUK_2 | ELUKOHT | ENDINE | 03.05.2016 | 27.11.2017 | Võru | Võru vald | Puiga küla | Rapla | Rapla vald | Juuru alevik |
| C | 10.08.2017 | NA | Tartu | Tartu linn | Tartu linn ELUK_2 | ELUKOHT | ENDINE | 28.11.2017 | 14.04.2020 | Tartu | Tartu linn | Tartu linn | Rapla | Rapla vald | Juuru alevik |
| C | 10.08.2017 | NA | Tartu | Tartu linn | Tartu linn POST_3 | ELUKOHT | ENDINE | 03.05.2016 | 27.11.2017 | Võru | Võru vald | Puiga küla | Rapla | Rapla vald | Juuru alevik |
| C | 10.08.2017 | NA | Tartu | Tartu linn | Tartu linn POST_3 | ELUKOHT | ENDINE | 28.11.2017 | 14.04.2020 | Tartu | Tartu linn | Tartu linn | Rapla | Rapla vald | Juuru alevik |

Joonis 23. EMPIS-e, rahvastikuregistri ja mobiilpositsioneerimise abil tuvastatud elukohtade lahknevus (punase joonega on ühendatud kattuvad aadressid).

Lisaks elukohtade lahknevusele valmistab probleeme ka erinevatesse registritesse kantud aadresside kehtivuse aeg, mis võib olla väga erinev. Joonis 24 on esitatud konstrueeritud näide võimalusest, kus uuritav isik on küll erinevates registrites aadressiga esindatud, kuid erinevatel perioodidel ning erinevates kohtades.



Joonis 24. Elukohale viitava või kontaktaadressi esinemise fragmenteerituse hüpoteetiline näide (värv tähistab aadressi).

Kuigi osadel juhtudel perioodid ja kohad võivad kattuda tervikuna või osaliselt, on siiski väga raske objektiivselt põhjendada kus ja millal kõnealune hüpoteetiline isik elada võis.

Kinnistusraamatu andmete kasutatavus inimeste elukohtaandmete täpsustamiseks

Uuringus osalejatega seotud Kinnistusraamatu kandeid oli algseid 208 isiku kohta. Pärast ühe operaatori (kust andmeid ei saadud) klientide eemaldamist jäi alles 122 Kinnistusraamatu kandega uuringus osalejat, kellel kokku tuvastati 315 kinnistusraamatu kannet. Sageli on kinnisvara omanikel

kinnisturaamatus mitu kannet, mis omakorda võivad koosneda lahustükkidest, millel aadressisüsteemis on omakorda mitu vastet. Osalejate kinnistusraamatu kannetest 186 (59%) olid hooneosad ja (129) 41% katastriüksused. Katastriüksuseid ja nende iseseisvaid ADS-OID-koodiga lahustükke, millel Eesti topograafia andmekogu andmetel puudus hoonestus oli kokku 21 (15 erinevat uuringus osalejat). Sellised kirjed eemaldati, kui aadressid, kus ei saa elada. Samuti tasub hoonete kihist eemaldada hooned, mille tüüp on vare või vundament (käesolevas uuringus selliseid juhtumeid polnud).

122 alles jäänud osalejast 68 (56,2%) oli hooneosa või hoonestusega kandeid kinnistusraamatus 1, 34 osalejal (28,1%) oli 2 ning 12 osalejal (9,9%) 3 kannet. Viiel osalejal (4,1%) oli 5 kannet ning kahe osaleja puhul 8 kannet.

Rahvastikuregistri püsielukoha aadressi ja Kinnistusraamatu kannete võrdlemisel selgus, et 70 osaleja (57,4%) puhul need kattusid ADS-OID tasemel. 1 Kinnistusraamatu kandega uuringus osalejast (68) kattusid aadressid ADS-OID tasemel 33 isiku (48,5%) puhul, 51 isikul (75,0%) oli kattuvus vähemalt asustusüksuse tasemel, 53 (78,0%) omavalitsuse ja 57 (83,8%) maakonna tasemel.

Kahe kinnistusraamatu kandega osalejate (34) puhul kattus rahvastikuregistri elukoha-aadressiga vähemalt 1 Kinnistusraamatu kanne 24 juhul. Kahe kahe kandega osaleja puhul ei kattunud rahvastikuregistri püsielukoht kummagi Kinnistusraamatu kandega ka maakonna tasemel ning ühel juhul oli kattuvus maakonna tasemel.

Osalejatest (12), kelle puhul leiti Kinnistusraamatust 3 hoonestusega vastet, oli 5 juhul üks täpne kanne (ADS-OID) Rahvastikuregistriga vastavuses 5 juhul, 6 osalejal oli vastavus vähemalt asustusüksuse tasemel ning 1 osaleja puhul ei leitud vastet ka maakonna tasemel.

Rangema filtrina võib määrata tingimuseks, et arvestatakse ainult neid kandeid, millel paiknevad ka eluhooned. Samas tuleb arvestada, et osades mitmeluruumidena registreeritud hooned on samuti kasutusel inimeste eluruumidena.

Niisiis võib Rahvastikuregistri püsielukoha aadresside ja Kinnistusraamatu omandikandeid võrreldes väita, et kattuvus on päris hea. Samas on paljudel juhtudel inimestel rohkem kui üks Kinnistusraamatu kanne ning suur hulk on ka neid, kellel selline omand üldse puudub. Ka puudub põhjuslik seos, mille alusel võiks väita, et kui Rahvastikuregistri ja Kinnistusraamatu elukoha aadressid kattuvad, siis inimene elabki just selles kohas. Seda kinnitab ka võrdlus mobiilpositsioneerimise andmetega. Nii näiteks kattub ühe Kinnistusraamatu kandega osalejatest 45,6% elukoha asustusüksus mobiilpositsioneerimise asukohaga. Omavalitsuse tasemel on kattuvus 41 osalejal (60,3%).

Võttes põhjendamata eelduseks, et mida rohkem erinevatest andmeallikatest pärit elukoha-aadresse satub samasse kohta, siis viitab see ka tegelikule elukohale, näeme, et vaid 57% juhtudest kattub selliselt leitud elukoht mobiilpositsioneerimise elukoha algoritmi tulemusega. Laiaulatuslikum inimese enda öeldud elukoha-aadresside kogumine ei garanteeri, et see aitab täpsustada tegelikku püsielukohta. Seda illustreerib ilmekalt ka järgnev näide ühe uuringus osaleja juhtumist (Joonis 25).



Joonis 25. Ühe uuringu osaleja kinnistusraamatu kannete (6) paiknemine rahvastiku registri järgse, mobiilpositsioneerimise põhise ja tegeliku elukoha suhtes.

Mobiilpositsioneerimise andmete registritega võrdluse kokkuvõte

Eelnevalt esitatud analüüsi põhjal võib väita, et passiivse mobiilpositsioneerimise andmed võiks olla väga olulise tähtsusega Rahvastikuregistri elukohaandmete täpsustamiseks. Tulemustest selgub, et vaid 67% uuringu osalejate puhul kattus rahvastikuregistri elukoht mobiilpositsioneerimise abil määratud elukohaga. Samas tuvastati 14% osaleja puhul mobiilipõhise elukoha vastavus rahvastikuregistri lisa-aadressiga. 6% osalejatest kattus mobiilpositsioneerimise andmetel leitud sekundaarne elukoht rahvastikuregistri elukohaga. Ühtegi kattuvust mobiilandmetega ei tuvastatud 13% inimeste puhul. See tähendab, et 87% osaleja puhul õnnestus tuvastada mobiilipõhiste püsi- ja sekundaarsete elukohtade vastavus rahvastikuregistri püsielukoha või lisa-aadressiga. Seejuures tuleks analüüsi läbiviijate arvates eelistada mobiilpositsioneerimise abil tuvastatud elukohaga kokku langevat rahvastikuregistri aadressi, kuna see peegeldab piirkonnas viibimise fakte.

Lisaks peab ka riik teadvustama ning määratlema, millistel juhtudel on oluline teada inimeste elukohta isikustatud kujul või piisab teadmisest, et vastavas kohas elatakse ning kui palju on elanikke ning millistel juhtudel on seda teavet vaja eluruumi täpsusega ning millal piisab üldisemat täpsusest (nt asustusüksus, linnaosa, asum vmt).

Nagu juba eespool tõdeti, ei ole passiivse mobiilpositsioneerimise abil võimalik määrata inimese elukohta täpsemalt kui mobiilivõrgu antenni leviala ulatus. Seejuures on üksikantennide levialad väga erineva suurusega (Joonis 5) ning võivad hõredalt asustatud maapiirkondades ulatuda mõnekümne

kilomeetrini. Samas elab kolm neljandikku elanikkonnast mobiilantennidele lähemal kui 6 km ja kaugemal kui 10 km mobiilantennist elab vaid mõni protsendipunkt inimestest. See tähendab, et mobiilpositsioneerimise andmetel on väga kõrge potentsiaal kontrollimaks, kas isiku taotleva elukohapõhise hüve tingimused on täidetud või mitte. Samas pole võimalik vastavat otsust automaatselt langetada. Kui tekib kahtlus Rahvastikuregistri ja tegeliku elukoha lahknevuses, on võimalik küsida inimese elukoha kohta täiendavaid andmeid, viia läbi kohapealt vaatlust vmt.

Arvestada tuleb ka asjaoluga, et kuigi aktiivsete SIM-kaartide suhtarvu poolest on Eestis 100 elaniku kohta 135 aktiivset SIM-i, siis on olemas teatud sotsiaalsed grupid, kellel mobiiltelefon kas puudub täiesti või kasutavad nad seda nii harva, et juhuslik kasutuskoht ei peegelda tegelikku elukohta. Ilmselt on selliste isikute osakaal siiski väga väike.

Kasutatud kirjandus

- Aasa, A. *et al.* (2021) 'Spatial interpolation of mobile positioning data for population statistics', *Journal of Location Based Services*. doi: 10.1080/17489725.2021.1917710.
- Ahas, R., Aasa, A., *et al.* (2010) 'Daily rhythms of suburban commuters' movements in the Tallinn metropolitan area: Case study with mobile positioning data', *Transportation Research Part C: Emerging Technologies*, 18(1), pp. 45–54. doi: 10.1016/j.trc.2009.04.011.
- Ahas, R., Silm, S., *et al.* (2010) 'Using mobile positioning data to model locations meaningful to users of mobile phones', *Journal of Urban Technology*, 17(1), pp. 3–27. doi: 10.1080/10630731003597306.
- Ahas, R. *et al.* (2015) 'Everyday space–time geographies: using mobile phone-based sensor data to monitor urban activity in Harbin, Paris, and Tallinn', *International Journal of Geographical Information Science*, 29(11), pp. 2017–2039. doi: 10.1080/13658816.2015.1063151.
- Ahas, R. and Mark, Ü. (2005) 'Location based services - New challenges for planning and public administration?', *Futures*, 37(6), pp. 547–561. doi: 10.1016/j.futures.2004.10.012.
- Järv, O., Ahas, R. and Witlox, F. (2014) 'Understanding monthly variability in human activity spaces: A twelve-month study using mobile phone call detail records', *Transportation Research Part C: Emerging Technologies*, 38, pp. 122–135. doi: 10.1016/j.trc.2013.11.003.
- Kamenjuk, P., Aasa, A. and Sellin, J. (2017) 'Mapping changes of residence with passive mobile positioning data: the case of Estonia', *International Journal of Geographical Information Science*, 31(7). doi: 10.1080/13658816.2017.1295308.
- Miller, H. J. (2005) 'A measurement theory for time geography', *Geographical Analysis*, 37(1), pp. 17–45. doi: 10.1111/j.1538-4632.2005.00575.x.
- Mooses, V. *et al.* (2020) 'An ethno-linguistic dimension in transnational activity space measured with mobile phone data', *Humanities and Social Sciences Communications*, 7(1), pp. 1–13. doi: 10.1057/s41599-020-00627-3.
- Newsome, T. H., Walcott, W. A. and Smith, P. D. (1998) *Urban activity spaces: Illustrations and application of a conceptual model for integrating the time and space dimensions*.
- Novak, J. *et al.* (2013) 'Application of mobile phone location data in mapping of commuting patterns and functional regionalization: A pilot study of Estonia', *Journal of Maps*, 9(1), pp. 10–15. doi: 10.1080/17445647.2012.762331.
- Saluveer, E. *et al.* (2020a) 'Methodological framework for producing national tourism statistics from mobile positioning data', *Annals of Tourism Research*, 81, p. 102895. doi: 10.1016/j.annals.2020.102895.
- Saluveer, E. *et al.* (2020b) 'Methodological framework for producing national tourism statistics from mobile positioning data', *Annals of Tourism Research*, 81. doi: 10.1016/j.annals.2020.102895.
- Schönfelder, S. and Axhausen, K. (2010) 'Urban rhythms and travel behaviour: Spatial and temporal phenomena of daily travel', *Urban Rhythms and Travel Behaviour: Spatial and Temporal Phenomena of Daily Travel*, pp. 1–230.
- Silm, S. and Ahas, R. (2010) 'The seasonal variability of population in Estonian municipalities', 42. doi: 10.1068/a43139.