

SVEUČILIŠTE U RIJECI
TEHNIČKI FAKULTET
Diplomski studij računarstva

Diplomski rad

**Metodologija za usporedbu
kontekstualiziranih
polazišno-odredišnih matrica**

Rijeka, prosinac 2019.

Vjera Turk
0069064924

SVEUČILIŠTE U RIJECI
TEHNIČKI FAKULTET
Diplomski studij računarstva

Diplomski rad

**Metodologija za usporedbu
kontekstualiziranih
polazišno-odredišnih matrica**

Mentor: prof.dr.sc. Renato Filjar

Rijeka, prosinac 2019.

Vjera Turk
0069064924

Umjesto ove stranice umetnuti zadatak za završni ili diplomski rad

Naslov zadatka: Metodologija za usporedbu kontekstualiziranih polazišno-odredišnih matrica

Thesis title: Methodology for contextualised origin – destination matrices comparison

Polje znanstvenog područja: Računarstvo

Grana znanstvenog područja: Informacijski sustavi

Sadržaj zadatka: Polazišno-odredišna matrica (POM) omogućuje sustavnu statističku procjenu migracija stanovništva u zadanom prostorno-vremenskom okviru. Za razliku od tradicionalnog pristupa brojanja putovanja i putnika, za procjenu POM-e danas se sve više koristi statistička analiza podataka iz suvremenih informacijskih i komunikacijskih sustava (zapisi o aktivnostima u javnoj pokretnoj mreži, združena očitavanja prijamnika za satelitsku navigaciju i sl.), čime je omogućeno poboljšanje kvalitete procjene preslikavanjem POM-e na kontekst. **Pojavljuje se potreba za objektivnom procjenom kvalitete POM-e u odnosu na referentnu (kontrolnu).** U ovom radu potrebno je definirati odnosne parametre kvalitete POM-e te razviti metodologiju usporedbe dviju POM-a dobivenih različitim postupcima procjene i s podacima iz različitih izvora. Usporedbu obaviti korištenjem numeričkog i grafičkog oblika POM-e. Metodologiju je potrebno izvesti u programskom okruženju za statističko računarstvo R te demonstrirati njenu primjenu na slučaju usporedbe dviju POM-a. Komentirati dobivene rezultate sa stajališta apsolutne i relativne točnosti matrica.

Izjava o samostalnoj izradi rada

Izjavljujem da sam samostalno izradila ovaj rad.

Rijeka, prosinac 2019.

Ime Prezime

Sadržaj

Popis slika	viii
Popis tablica	x
1 Uvod	1
2 Polazišno-Odredišna Matrica	3
2.1 Definicija Polazišno odredišne matrice	3
2.2 Tradicionalni pristupi procjeni Polazišno-Odredišnih Matrica	6
2.2.1 Izvori ulaznih podataka	6
2.2.2 <i>Traffic Analysis Zone</i>	8
2.2.3 Model prometa	9
2.2.4 Vremenski nezavisno i vremenski zavisno modeliranje	11
2.3 Suvremeni pristupi procjeni Polazišno-Odredišnih Matrica	12
2.3.1 Pokretna osjetila	12
2.3.2 Polazišno-Odredišna Matrica iz zapisa o telekomunikacijskim aktivnostima u javnoj pokretnoj mreži	14
2.3.3 Diskretizacija Vremena	17

3	Pregled postupaka i istraživanja o vrednovanju Polazišno-Odredišnih Matrica	19
3.0.1	Linearna regresija - mjera r^2	20
3.0.2	Vrednovanje suvremenih Polazišno-Odredišnih Matrica	22
3.0.3	Grafički oblik i sličnost strukture Polazišno-Odredišnih Matrica	22
4	< Naslov >	26
4.1	Uvodna razmatranja i pregled istraživanja	26
4.2	Kontekstualizacija Polazišno-Odredišnih Matrica	29
4.2.1	Primjeri korištenja konteksta u analizi kretanja	29
4.3	Teza	32
4.3.1	Alternativni pristup vrednovanju Polazišno-Odredišnih Matrica	32
4.4	Odnosni parametri kvalitete	32
4.4.1	Prostorno obuhvaćanje	32
4.4.2	Rezolucija	33
4.4.3	Zrnatost	35
4.4.4	Putovanje kao parametar	36
4.4.5	Gustoća informacija	39
4.4.6	Ukupna širina toka	40
5	Metodologija	42
5.1	Uvjeti usporedbe	42
5.2	Usporedba prema odnosnim parametrima kvalitete	44
5.2.1	Postupak odlučivanja	45
6	Rezultati	47

Sadržaj

7	Diskusija	48
8	Zaključak	49
	Pojmovnik	57
	Sažetak	58
A	Postojeće metrike za vrednovanje Polazišno-Odredišnih Matrica	59
	A.0.1 Metrike za procjenu sličnosti Polazišno-Odredišnih Matrica s referentnom	59
B	Prilagodba prostorne podjele	61
	B.1 Agregacija baznih stanica	61
	B.2 Interpolacija	62
	B.3 Posljedice agregacije prostornih ćelija	63
C	OpenCellID	65
D	OpenStreetMap	66

Popis slika

2.1	Polazišno-Odredišna Matrica. Prema: [1]	5
2.2	Usmjereni graf kretanja i pripadajuća Polazišno-Odredišna Matrica. (Zone su označene slovima abecede). Prema: [2]	5
2.3	Prikaz prosječnog volumena na vezama (prometnicama) sa stanicama za prebrojavanje radnim danom 2005. Orlando, Florida. [14]	8
2.4	Mreža za analizu prometa sastoji se od 221,14 cesta - plave linije i 3,199 TAZ zona - granice označene iscrtanom crnom linijom [17] . .	9
2.5	Voronoi dijagram 1090 ćelija baznih stanica - kineski grad Shenzhen [27]	17
3.1	Primjer linearne regresije između vrijednosti dvije POM. Vrijednosti POM-e A su na x i vrijednosti POM-e B su na y osi. Pravac linearne regresije označen je crvenom, reziduali sivim linijama. $a = 0.0071273$, $b = 0.631128$, $r^2 = 0.4026$	21
3.2	Narušenu strukturalnu sličnost lakše je uočiti nego narušen MSE. SSIM (i MSSIM) u odnosu na referentnu POM-u može biti korišten kao dodatna informacija (<i>goodness of fit</i> mjera) prilikom vrednovanja POM-e [35]	25
4.1	Širina jutarnjeg vrhunca prometne potražnje - vremenski zavisno mo- deliranje [20]	28

4.2	Distribucija udaljenosti putovanja u kilometrima, Senegal 2016. [19]. U studiji su izbačena putovanja kraća od $3km$ (prostorna rezolucija > $3km$)	38
4.3	Dvije POM-e - produkti različitih definicija putovanja. Dobivene su iz potpuno istog izvora CDR zapisa. [19]	41
5.1	Broj putovanja javnim prijevozom zabilježenih <i>OV-chipkaart</i> pamet- nim karticama u Nizozemskoj izražen u u milijunima po tjednima u godini. Zelena linija označava prosječnu vrijednost radnog tjedna [32]	43
B.1	Prostorna podjela na anketne jedinice u gradu Santiagu. Gustoća tornjeva baznih stanica po anketnim jedinicama. Gustoća varira od 1 do 250 tornjeva po anketnoj jedinici. [46]	62
B.2	Primjer mozaika <i>krhotina</i> za interpolaciju između Voronoi ćelija i kvadratne mreže. Nijansa odražava postotak površine zgrada u <i>kr-</i> <i>hotini</i> u odnosu na cijelu Voronoi ćeliju kojoj pripada. Rezultat se dobiva na osnovu tih <i>težinskih faktora</i> polazišne i odredišne <i>krhotine</i> te pripadajućih ("roditeljskih") Vornoi ćelija. Sumom svih rezultata <i>krhotina</i> jedne kvadratne ćelije izračuna se konačni tok. [15]	63
B.3	Distribucija CDR putovanja na području Santiaga, Čile. Polazišno- Odredišna Matrica (POM) je normalizirana prema redovima (po- lazištima) (*) [46]	64

Popis tablica

2.1	Primjeri zapisa o telekomunikacijskim aktivnostima (pozivi) [23] . .	15
2.2	Primjeri zapisa o telekomunikacijskim aktivnostima [24]. Iz zapisa je uklonjen datum, a identifikacijska oznaka bazne stanice (pozivatelja) zamijenjena je njenim položajem. Podatci o baznim stanicama dostupni su u bazama poput OpenCellID (vidi Dodatak C)	15
5.1	Tablica usporedbe	46

Poglavlje 1

Uvod

Polazišno-Odredišna Matrica je alat koji omogućuje opis i sustavnu statističku procjenu migracija stanovništva na nekom području u zadanom prostorno-vremenskom okviru. POM-a služi za opis grupne mobilnosti i mjerenje socio-ekonomske aktivnosti u nekoj regiji, a najčešće se koristi u prometnoj znanosti za analizu i strateško planiranje prometnog opterećenja i prometne infrastrukture.

Osim tradicionalnog pristupa prebrojavanja putovanja i putnika na raskrižjima, anketiranjem ili pomoću različitih strateški postavljenih osjetila, za procjenu POM-a koristi se i statistička analiza podataka iz suvremenih informacijskih i komunikacijskih sustava (zapisi o aktivnostima u javnoj pokretnoj mreži, združena očitavanja prijamnika za satelitsku navigaciju i sl.). Takav moderan pristup omogućuje poboljšanje kvalitete procjene preslikavanjem POM-e na kontekst. Ako nosi informaciju o svrsi kretanja, POM-a specificira socio-ekonomske aktivnosti koje su pokretači putovanja odnosno motivatori individualne i grupne mobilnosti.

Nedostaci kao što su mali uzorci stanovništva u anketama, neodređenost polazišta i odredišta putovanja, korištenje podataka koji nisu ažurni, čine tradicionalne metode za procjenu POM-a sve manje kompetentnima u rapidno promjenjivim urbanim sredinama.

Pojavom i razvojem velikog broja usluga zasnovanih na lokaciji, prikupljanje podataka o položaju u svrhu pružanja usluge omogućuje analizu kretanja stanovništva. Suvremeni izvori podataka o položaju i načini procjene POM-a koji uključuju kon-

Poglavlje 1. Uvod

tekstualizaciju su nov doprinos u području analize kretanja stanovništva.

Postojeće metode vrednovanja POM-a definiraju vrijednost nove matrice razmatranjem njene sličnosti s postojećom matricom za isto područje. Postoji potreba da se definira i kvantizira kvaliteta POM-e kroz objektivne parametre.

... dan je okvir za donošenje odluka o kvaliteti POM-a koje se uspoređuju. (...)

Poglavlje 2

Polazišno-Odredišna Matrica

Polazišno-Odredišna Matrica (*eng. Origin-destination Matrix (ODM) ili Trip Table*) je uobičajeni način mjere pokretljivosti stanovništva na nekom području u prostorno-vremenskom okviru. POM-a služi za opis grupne mobilnosti i mjerenje socio-ekonomske aktivnosti u nekoj regiji, a najčešće se koristi u prometnoj znanosti za analizu i strateško planiranje prometnog opterećenja i prometne infrastrukture. POM-e mogu opisivati prometnu potražnju samo na jednom raskrižju, na području grada, na području jedne regije, cijele države pa i šire. Veličinom područja na kojem se nastoje opisati kretanja povećava se i zahtjevnost prikupljanja podataka za procjenu prometne potražnje i izradu POM stoga postoji više postupaka prikupljanja podataka i modela za procjenu prometa.

2.1 Definicija Polazišno odredišne matrice

Definicija 2.1.1. Polazišno-odredišna matrica je matrica čiji svaki element predstavlja broj putovanja između odgovarajućeg para zona unutar promatranog područja u promatranom vremenskom razdoblju τ , u svakom smjeru zasebno. Redovi u POM-i $i = 1, 2, 3 \dots n$ predstavljaju polazišta (izvore) prometnog toka, a stupci $j = 1, 2, 3 \dots n$ njihova odredišta (ušća). Element f_{ij} predstavlja broj putovanja iz pripadajuće polazišne zone i u pripadajuću ciljnu zonu j . Broj putovanja naziva se i širina prometnog toka (*eng. flow*). (slika 2.1)

Poglavlje 2. Polazišno-Odredišna Matrica

POM-a se može smatrati matričnim zapis težinskog, usmjerenog grafa, gdje su težinski faktori usmjerenih veza širine prometnih tokova između čvorova mreže koji predstavljaju zone na koje je promatrano područje podijeljeno (slika 2.2). Zone u POM-i su poredane od zapada prema istoku, te od od sjevera prema jugu.

POM-a može biti procijenjena za sadašnje ili planirano prometno opterećenje. Može biti prikazana ukupno ili po pojedinim načinima prijevoza (modalitetima prometnog sustava - osobna vozila, vozila javnog prijevoza, teretna vozila, pješaci itd.) i/ili različitim svrhama putovanja (putovanja na posao, putovanja kući...). [3]

Tranzitna POM-a (eng. t-ODM) je vrsta POM-e u kojoj su zabilježene poznate tranzicije između zona. Svaka tranzicija predstavlja onaj dio putovanja koji je poznat, a početak i završetak cjelokupnog putovanja kojem pripada su nepoznati [4].

Poglavlje 2. Polazišno-Odredišna Matrica

Polazišta (Origins)	Odredišta (Destinations)							$\sum_{j=1}^n f_j = O_j$
	1	2	3	...	j	...	n	
1	f_{11}	f_{12}	f_{13}	...	f_{1j}	...	f_{1n}	O_1
2	f_{21}	f_{22}	f_{23}	...	f_{2j}	...	f_{2n}	O_2
3	f_{31}	f_{32}	f_{33}	...	f_{3j}	...	f_{3n}	O_3
...
i	f_{i1}	f_{i2}	f_{i3}	...	f_{ij}	...	f_{in}	O_i
...
n	f_{n1}	f_{n2}	f_{n3}	...	f_{nj}	...	f_{nn}	O_n
$\sum_{i=1}^n f_i = D_i$	D_1	D_2	D_3	...	D_j	...	D_n	$\sum_{i=1}^n \sum_{j=1}^n f_{ij} = F_{ij}$

Slika 2.1 Polazišno-Odredišna Matrica. Prema: [1]



Slika 2.2 Usmjereni graf kretanja i pripadajuća Polazišno-Odredišna Matrica. (Zone su označene slovima abecede). Prema: [2]

2.2 Tradicionalni pristupi procjeni Polazišno-Odredišnih Matrica

2.2.1 Izvori ulaznih podataka

Ankete

Tradicionalni pristupi procjeni POM-a uključuju metode provođenja anketiranja. Vrste anketa su: ankete u kućanstvima, presretanje vozila, anketiranje prijevoznih kompanija i tvrtki, tranzit na granici, turisti u hotelima i parking [5]. Ankete su zahtjevne iz perspektive utrošenog vremena, uloženog truda i radne snage stoga se ne provode često. Anketiranje putovanja u kućanstvima provodi se na područjima većine razvijenih svjetskih metropola tek **svakih 5 do 10 godina** [6]. U većini na provođenje anketa odlazi najveći postotak investicija u planiranje prometa, u prosjeku oko 7.4 milijuna dolara godišnje [5]. Podaci koje ankete stanovništva bilježe osim broja stanovnika su broj vozila, broj zaposlenih, broj putovanja određenim načinom prijevoza (automobil, javni prijevoz...) po kućanstvu [7]. Ostale ankete općenito pokrivaju malen dio populacije te su POM-e dobivene iz anketa opterećene neizbježnom pristranošću [8]. Studija provedena na području Belgije pokazala je da se zadovoljavajuća razina kvalitete POM-a procijenjenih isključivo iz anketa u kućanstvima postiže tek ako uzorak obuhvaća 50% populacije [5].

Prebrojavanje vozila (engl. *Traffic Counts*)

Metode koje uključuju prebrojavanje vozila na ključnim čvorištima prometne infrastrukture manje su zahtjevne za izvođenje u odnosu na ankete i značajno smanjuju vrijeme i troškove prikupljanja podataka. Postupci uključuju: analizu nadzornih snimaka prometa, automatski sustav za prepoznavanje i očitovanje registarskih pločica engl. Automatic Number Plate Recognition (ANPR), osjetila prometnog toka (radarska, magnetna, video-senzori, zvučna..) i ručno prebrojavanje. **Primjenom ovih postupaka nije moguće pouzdano odrediti stvarna polazišta i odredišta putovanja** [9], već se ona određuju matematički modeliranjem uz uvažavanje pretpostavki (odlomak 2.2.3). Raspodjela stanica ili osoba koje rade obzervaciju

prometa treba biti plansko, a broj dostupnih uređaja/ljudi ograničavajući je faktor za veličinu područja koje je moguće nadzirati [10]. Točnost procijenjenih matrica ovisi o: korištenom modelu procjene, grešci ulaznih podataka i infrastrukturi detektora koju je skupo izmijeniti ili proširiti zbog troškova održavanja i instalacije [10] [11]. Podaci koje generira oprema koja radi na temelju vizualne percepcije mogu postati neiskoristivi ako su prikupljeni za vrijeme trajanja loših vremenskih uvjeta, a kao nadomjestak se koristi ručno prebrojavanje [12].

Raspodjela stanica za prebrojavanje

Točnost matrice povećava se s brojem stanica- položaja u infrastrukturi s kojih će se vršiti prebrojavanje. Stanice na različitim položajima u prometnoj mreži imaju različit stupanj utjecaja na procjenu POM-e [10]. Važno je odrediti optimalan broj stanica prebrojavanja i njihovih položaja u prometnoj mreži. Optimalni uvjeti dobit će se poštujući slijedeće četiri pretpostavke:

1. Barem jedna stanica prebrojavanja po paru zona
2. Točke prebrojavanja trebaju biti na onim prometnim vezama (cestama) između para zona gdje protječe najveći mogući prometni tok između te dvije zone
3. Odabrana prometna veza treba presijecati što je više moguće tokova
4. Rezultati prebrojavanja prometa na svim vezama trebaju biti linearno nezavisni

Prošireni modeli za raspodjelu stanica u današnjim programskim podrškama imaju mogućnost ograničavanja investicija u infrastrukturu za promatranje. **Prva pretpostavka neće biti ispunjena ako ako nisu osigurana sredstva [10]. U praksi, model prometne potražnje može imati 2,000 Traffic Analysis Zone (TAZ) što znači da pripadajuća POM-a ima 4 milijuna ćelija, dok je na prometnoj mreži postavljeno svega 1,000 stanica za prebrojavanje [13].** Na slici 2.3 prikazana je karta grada Orlando s istaknutim važnim prometnicama (sivom) i s prosječnim prometnim tokovima zabilježenim uređajima za prebrojavanje vozila. [14]

Poglavlje 2. Polazišno-Odredišna Matrica



Slika 2.3 Prikaz prosječnog volumena na vezama (prometnicama) sa stanicama za prebrojavanje radnim danom 2005. Orlando, Florida. [14]

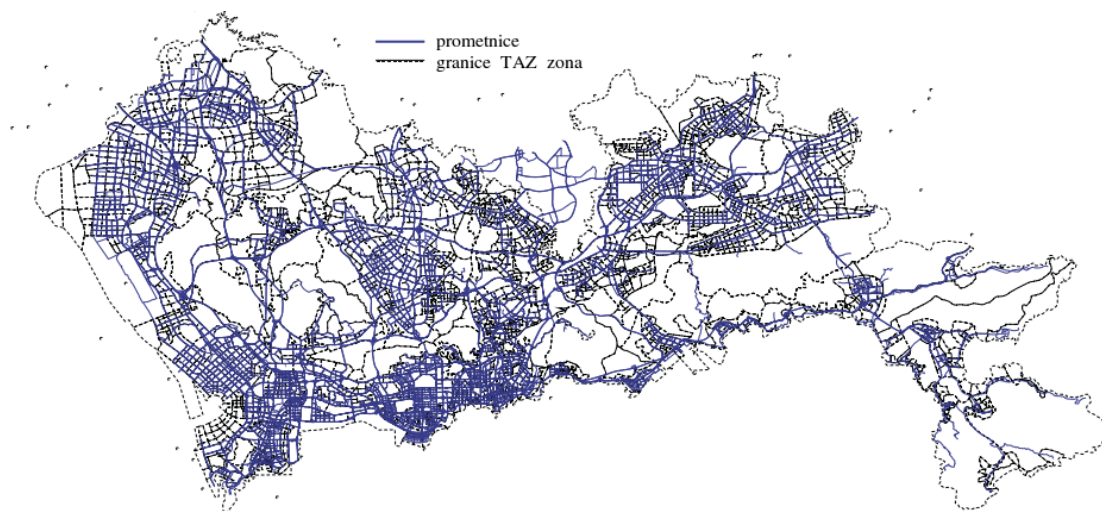
2.2.2 Traffic Analysis Zone

Traffic Analysis Zone (TAZ) je strani naziv za jedinicu prostorne podjele (zonu) kod analize prometa. Površina tih jedinica varira i nije uniformna. Podjela se može temeljiti na populaciji ili broju zaposlenih na nekom prostoru, primjerice svaka zona pokriva prostor na kojem obitava otprilike tri tisuće ljudi, ili zone generiraju podjednak broj izlaznih putovanja (jednaka širina izlaznih tokova) [15].

U hijerarhijskoj podjeli kod ankete stanovništva općenito postoje manje i veće jedinice podjele, a nema podjele koja bi precizno odgovarala TAZ. Anketni blokovi određeni su kao relativno homogene jedinice kad je riječ o karakteristikama populacije, ekonomskom statusu i uvjetima života. Do 2000. godine zastupljen je bio jedan model predviđanja prometa, a veličina TAZ bila je približno jednaka anketnom bloku od 600 do 3,000 ljudi. Pojavom kompleksnijih modela

predviđanja prometa počinju se koristiti sve manje prostorne jedinice jer se povećava i razina detalja koju noviji modeli mogu ostvariti. Jedinica veća od TAZ uvedena 2010. naziva se Traffic Analysis District (TAD) i obuhvaća oko 20,000 stanovnika, a koristi se za analize na razini države.[16]

Kineski grad Shenzhen ima preko 10 milijuna stanovnika i proteže se na 1,991 km^2 . Istraživanje objavljeno 2017. koristi podjelu na 3,199 TAZ (Vidi sliku 2.4). Iz danih informacija zaključuje se da je prosjek površine TAZ u gradu Shenzhen 0.62 km^2 i u prosjeku jedna zona obuhvaća više od 3,125 stanovnika.



Slika 2.4 Mreža za analizu prometa sastoji se od 221,14 cesta - plave linije i 3,199 TAZ zona - granice označene iscrtanom crnom linijom [17]

2.2.3 Model prometa

Modeliranje prometa zahtjeva veliku količinu podataka - informacija o prometnoj mreži i prometnoj potražnji. Točnost modelirane (planirane) prometne situacije ovisi o kvaliteti dostupnih informacija te kako su podaci kombinirani, koji težinski faktori su primijenjeni za različite izvore. Prometna potražnja ključna je komponenta te stoga gotovo svaki prometni model zahtjeva POM-u koja specificira prometnu potražnju između zona u prometnoj mreži. Točna POM-a osnova je za odluke kod

Poglavlje 2. Polazišno-Odredišna Matrica

strateških planiranja prometne infrastrukture i u Inteligentnim Transportnim Sustavima (engl. *Intelligent Transportation Systems*) (ITS) koji su zauzvrat ključni za izbor ruta u različitim sustavima navigacijskog navođenja.

Neke od pretpostavki kojima se vode u modeliranju prometne potražnje su da je ukupna dnevna potražnja podjednaka u oba smjera kod svakog para zona na području grada [4], da svi korisnici imaju saznanja o situaciji na cesti [15] te da je mobilnost općenito manja tijekom zime nego tijekom ljeta [9] i da je broj putovanja po kućanstvu proporcionalan dohotku [7]. Klasičan bazni model za procjenu prometne potražnje je Gravitacijski model (eng. *Gravity model*) koji kod proračuna "privlačnosti" 2 zone uzima u obzir **gustoću populacije, udaljenost i troškove putovanja**. *Broj putovanja između dvije zone srazmjeran je veličini aktivnosti odnosno nastajanju i privlačenju putovanja u tim zonama, a obrnuto proporcionalan prostornoj udaljenosti između zona ili nekoj drugoj vrsti otpora (vrijeme putovanja, pješaćenja, čekanja, parkiranje, vozarina,.. ili kombinacija svih tih faktora)* [18]. Izvedene verzije integriraju i neke druge (socio-ekonomske) faktore [19].

Ankete se provode na ograničenom podskupu populacije, a ograničen je i broj stanica za prebrojavanje. Procjena POM iz prebrojanih vozila je matematički pododređena (više je nepoznanica nego poznatih informacija). To znači da bi velik broj matrica odgovarao tako prebrojanim tokovima vozilima. Oba izvora kombiniraju se u procjeni POM-a. Temeljna ideja kod modeliranja je [13]:

1. Pronaći postojeću POM-u za isto područje i raspodjelom njenih tokova na prometnu infrastrukturu simulirati rezultate prebrojavanja na odabranim točkama u prometnoj mreži.

Ako postoji više POM-a za isto područje potrebno je prikupiti informacije i odabrati onu koja najviše odgovaraju za konkretni slučaj. Većina algoritama koristi takozvanu engl. *seed* POM-u koja je *najbolja procjena* dosadašnje potražnje, a koja procjenjuje prometnu potražnju u prošlosti- npr. iz anketiranja po kućanstvima, nedavnu prometnu potražnju neprecizno ili je potražnja razvijena iz *principa tipičnog ponašanja vozača*. [13]

2. Početnu POM-u modificirati iterativnim procesom evaluacije omjera tokova

dodijeljenih prometnim vezama iz POM-e i prebrojenih tokova na vezama gdje se promet prebrojavao. [13]

U tom iterativnom procesu prometni tokovi iz modificirane POM-e ponovno se dodjeljuju prometnoj infrastrukturi (sve dok se ne postigne zadani broj iteracija ili zadovoljavajući stupanj sličnosti početne i nove POM-e - detaljnije u poglavlju 3).

2.2.4 Vremenski nezavisno i vremenski zavisno modeliranje

Modeliranje prometa dijeli se na dvije grane, vremenski nezavisnu (statičku, analitičku) i vremenski zavisnu (simulacijsku). U **statičkom modeliranju cilj je analiza prometa za strateška, investicijska planiranja izgradnje nove infrastrukture. Modeli procjenjuju prosječnu dnevnu prometnu potražnju (period od 16 ili 24 sata). Radi se o vremenski nezavisnoj analizi.** [13]

Vremenski zavisno modeliranje

Vremenski zavisno modeliranje koristi se npr. kod upravljanja prometnom signalizacijom i definiranja rasporeda vožnje javnog prijevoza. Riječ je o vremenski zavisnoj analizi prometa. Period u kojem se procjenjuje prometna potražnja manji je od jednog dana, primjerice period τ od 15 minuta [10][7].

Neki *Peak-hour* vremenski modeli koji opisuju potražnju za vrijeme dnevnog vrhunca prometne potražnje, procjenjuje potražnju za svaku minutu ($\tau = 1$ min). [20]. Modeliranje u stvarnom vremenu svrstava se u vremenski zavisno modeliranje.

Većina razvijenih modela za vremenski zavisno modeliranje iz postupka prebrojavanja (engl. *link count*) primjenjuju se na malim, zatvorenim prometnim mrežama kao što su raskrižja, petlje i manje autoceste. [10]

2.3 Suvremeni pristupi procjeni Polazišno-Odredišnih Matrica

Informacijske i telekomunikacijske usluge generiraju masivne skupove podataka vezanih za gibanje u prostoru (mobilnost). Podatci koje generiraju pametni telefoni i pametne kartice omogućuju podatkovnim analitičarima razumjeti ponašanje pojedina u različitim domenama, uključujući i njihov obrazac kretanja. Danas ti podatci postaju suvremeni izvori podataka o dnevnim mobilnosti stanovništva. Podaci iz tih suvremenih izvora nisu ciljano prikupljani za analizu mobilnosti stanovništva već za pružanje usluge zasnovane na lokaciju te se koristi naziv pasivno prikupljani polazišno-odredišni podaci (engl. passively collected origin-destination data). **Suvremeni izvori podataka o položaju kao što su zapisi o telekomunikacijskoj aktivnosti nisu vezani uz prometnu infrastrukturu niti za određen način prijevoza (kretanja).**

Pružatelji usluga zasnovanih na lokaciji i operatori javne pokretne mreže moraju voditi računa o privatnosti pojedinaca kada je riječ o podacima koji se prikupljaju, uključujući podatke o položaju korisnika. Ukoliko ustupe podatke za razvoj POM-a, ti podaci moraju biti anonimizirani te ne smije postojati mogućnost povratnog inženjerstva kojim bi se povratio izvorni identitet. Važno je osigurati anonimnost i dozvolu pojedinca da se podaci koriste u jasno definirane svrhe. Idealno je da operator ustupi već anonimizirane podatke stranci koja iz njih procjenjuje POM-e.

2.3.1 Pokretna osjetila

Pokretno osjetilo je uređaj koji je postavljen na vozilo koje se potom kreće prometnom infrastrukturom, postajući tako dijelom prometnog toka, te koji je opaža vrijednosti definiranih parametara gibanja vozila na koje je postavljeno te ih pohranjuje i prenosi radi naknadnog procesiranja. Osnovni podatci koje osjetilo prikuplja su položaj, brzina i smjer kretanja vozila. [3]

Zahtjev koji nužno mora biti ispunjen je poznavanje položaja vozila. Uređaj određuje položaj pomoću Global Navigation Satellite System (GNSS) prijemnika ili se vozilo identificira na određenim točkama u prometnoj mreži, na kojima može

ujedno i prenijeti podatke prikupljene od posljednje točke. Anglosaksonski naziv za ovu metodu prikupljanja podataka o stanju u prometu je Floating Car Data (FCD). FCD metoda zastupljena je u ITS. **U prikupljanju podataka sudjeluju uređajima opremljena taksi vozila, službena vozila, vozila za distribuciju robe, i vozila javnog prijevoza.** Kod ITS koji zahtijevaju podatke u realnom vremenu mora biti uspostavljen komunikacijski sustav za učestali prijenos podataka o položaju, a najčešće se odvija preko Global System for Mobile (Communications) (GSM), Wi-fi i Bluetooth komunikacijskih sustava. [3]

Ukoliko pokretno osjetilo nije specijalizirani uređaj već pametni telefon koristi se naziv Floating Phone Data (Floating Cellular Data) (FPD). FPD se odnosi na prikupljanje podataka pomoću mobilnih (pametnih) telefona neovisno o načinu određivanja položaja - korištenjem GNSS ili pomoću javne pokretne mreže. Položaj pokretne stanice unutar javne pokretne mreže može se odrediti triangulacijom signala koje pokretna stanica hvata iz više baznih stanica ili približno odrediti iz jednačavanjem položaja pokretne stanice s položajem (područjem pokrivanja) bazne stanice s kojom je pokretna stanica povezana u trenutku očitavanja položaja.

Preciznost određivanja položaja

GNSS osnovni postupak određivanja položaja sadržani neke pogreške određivanja položaja. Preciznost u rangu od 10-20 m je moguća uz dostupnost signala od najmanje 4 satelita, no u urbanom okruženju zgrade i druge visoke strukture često blokiraju vidljivost između satelita i prijamnika. Rezultat tog je degradacija točnosti određenog položaja. Napredno određivanje položaja podrazumijeva usporedbu i usklađivanje inicijalno određene putanje kretanja s kontekstom (npr. poznatom infrastrukturom) (engl. *map matching*). Njime se postiže preciznost u rangu 0-15 m. [21][22]

2.3.2 Polazišno-Odredišna Matrica iz zapisa o telekomunikacijskim aktivnostima u javnoj pokretnoj mreži

U postupku praćenja aktivnosti za potrebe pružanja i naplatu usluga te za održavanje operativnosti same mreže zapisuju se podaci o aktivnostima u javnoj pokretnoj mreži. Pokretni uređaj (npr. pametni telefon) u javnoj pokretnoj mreži je kontinuirano povezan s mrežom putem odgovarajuće bazne stanice, neovisno o telekomunikacijskoj aktivnosti uređaja. U postupku uspostave poziva, signalizacija uključuje i identifikacijsku oznaku bazne stanice s koje je poziv uspostavljen. Navedena oznaka bilježi se u zapisu o telekomunikacijskoj aktivnosti.

Zapisi o telekomunikacijskoj aktivnosti i detaljima telekomunikacijskih transakcija vode se u odgovarajućim komponentama javne pokretne mreže pod imenima: Call Data Records ili Charging Data Records (CDR). CDR obuhvaća transakcijske podatke o pozivima, razmjeni kratkih poruka (engl. Short Messaging System, SMS) i prijenosu podataka pokretnim internetom. Transakcijski podatci uključuju, između ostalog: vremenske trenutke početka i završetka telekomunikacijske transakcije (korištenja komunikacijske usluge), vrstu komunikacije, identifikacijske oznake pošiljatelja i primatelja, te identifikacijske oznake baznih stanica na koje su u trenutku započinjanja i završetka transakcije bili spojeni pošiljatelj i primatelj. (Tabela 2.2)

Zapisuju se i podatci o prekapčanju pokretnog uređaja na drugu baznu stanicu (engl. *handover data*) tijekom trajanja telekomunikacijske aktivnosti radi postizanja i održavanja potrebne kvalitete telekomunikacijske usluge, *Location Area Updates* - podatci prikupljani svakih 20 min do 2 sata o tome na koju skupinu baznih stanica je pokretna stanica spojena (svaka LA skupina ima od 150 do 500 baznih stanica, [9]), te podaci kada je pokretna stanica uključena u mrežu i isključena iz mreže i opterećenje pojedinog tornja [23].

Novi postupci koriste upravo anonimizirane, često višemjesečne CDR, kao izvor podataka za procjenu POM-a [25][19][8]. Metodologija korištenja odnosnih elemenata CDR-a za procjenu POM predložena je još 1994. godine [19].

Definicija 2.3.1. Skup minimalno n događaja (telekomunikacijskih aktivnosti) u nizu, istog korisnika, koji su zabilježeni na istom položaju unutar definiranog mini-

Poglavlje 2. Polazišno-Odredišna Matrica

Tablica 2.1 Primjeri zapisa o telekomunikacijskim aktivnostima (pozivi) [23]

UserID	Event Time	Cell Tower	Caller	Callee	In/Out	Duration (s)
38DA6	2015-05-01 18:26:50	1921	38DA6	163B7	Out	52
78EC3	2015-05-01 14:16:09	2189	53880	78EC3	In	600

Tablica 2.2 Primjeri zapisa o telekomunikacijskim aktivnostima [24]. Iz zapisa je uklonjen datum, a identifikacijska oznaka bazne stanice (pozivatelja) zamijenjena je njenim položajem. Podatci o baznim stanicama dostupni su u bazama poput OpenCellID (vidi Dodatak C)

SIM Card ID	Time	Latitude	Longitude
0055555556	13:10:09	114.0397917	22.5740278
0055555556	16:08:30	113.8056944	22.7613889
0055555556	20:50:17	113.8465972	22.7725
0055555557	23:39:23	114.043854	22.530757
0055555558	22:13:27	114.043854	22.530757

malnog vremenskog perioda čine jedno Zaustavljanje (engl. *Stop*) .

Zahtjevom da se radi o više ($n > 1$), a ne jednom ($n = 1$) zabilježenom događaju pokušava se osigurati da se ne radi o tranzitnom već o stvarnom zaustavljanju i suzbija se bilježenje lažnog kretanje kod privremenog spajanja na susjednu baznu stanicu. Ako je $n > 1$ zaustavljanje ima 2 vremenske oznake, trenutak kada je zabilježen početak i trenutak kada je zabilježen kraj zaustavljanja.

Definicija 2.3.2. Putovanje je skup dva zaustavljanja koja zadovoljavaju unaprijed definirana minimalna i maksimalna ograničenja međusobne vremenske i prostorne udaljenosti. Kraj polazišnog zaustavljanja postaje početak putovanja, a početak odredišnog zaustavljanja kraj putovanja. Niz kontinuiranog zaustavljanja na različitim položajima otkriva putanju kretanja.

Prostorna podjela - Voronoi teselacija

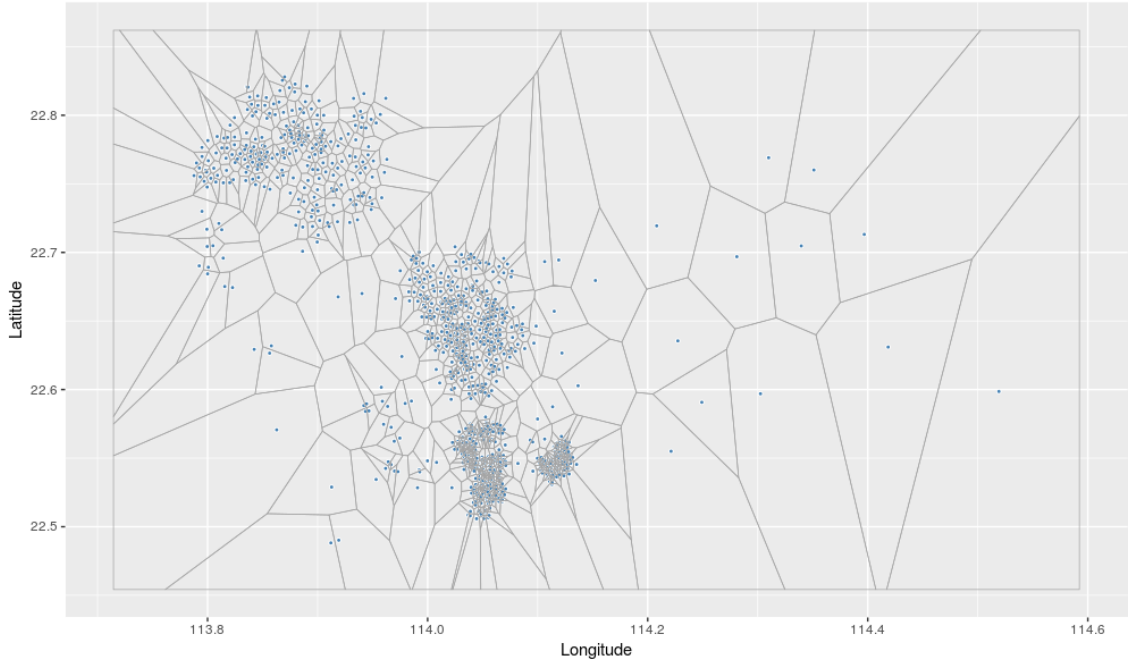
Najfinija prostorna rezolucija koju je moguće postići kada je riječ o POM-i dobivenoj iz CDR postiže se podjelom prema Voronoiovoj teselaciji.

Definicija 2.3.3. Voronoi teselacija definirana je skupom točaka P u ravnini, gdje za svaku točku p_i Voronoi ćeliju K_i čine sve točke ravnine bliže toj točki negoli ikoj drugoj točki iz skupa. Svaka takva ćelija dobiva se presjecištem pola prostora između susjednih točaka skupa, pa stoga čini konveksni poligon. Dijelovi Voronoiova dijagrama su sve točke u ravnini koje su ekvidistantne prema dva najbliža položaja (granice poligona). Voronoiovi vrhovi su točke ekvidistantne tri (ili više) položaja. [26]

Točke skupa su položaji baznih stanice, a ćelije oko njih aproksimacija površine koju ta bazna stanica pokriva (vidi sliku 2.5). Površine prekrivanja u stvarnosti se preklapaju jer područje prekrivanja nije sasvim fiksno i ovisi o trenutnoj aktivnosti svake bazne stanice [9].

Gustoća Voronoi ćelija dobro je korelirana s gustoćom stanovništva te stoga bolju prostornu rezoluciju ima na područjima veće gustoće populacije [19]. Ako poistovjetimo položaj korisnika s položajem tornja bazne stanice, mjerna nesigurnost određivanja položaja iznosi od približno stotinu metara u gusto naseljenom, do nekoliko kilometara u ruralnom području [28]. Problem predstavlja preklapanje područja pokrivanja u stvarnosti [19]. Zbog uravnoteženja opterećenja u mreži (eng. *load balancing*) uređaj mijenja baznu stanicu na koju je spojen iako svo vrijeme zadržava isti položaj. Učinak koji uravnoteženje opterećenja ima na definiranje putovanja može se ublažiti određivanjem minimalne udaljenost putovanja, zanemarivanjem putovanja u susjedne ćelije [19] ili agregacijom područja pokrivanja baznih stanica čija je udaljenost manja od 75 m (s dodjelom događaja najaktivnijem tornju). [15]

Analiza snage signala i trenutnih opterećenja okolnih baznih stanica mogu dati jasniji uvid u opterećenje mreže u vremenu početka (i trajanja) telekomunikacijske aktivnosti. [23]



Slika 2.5 Voronoi dijagram 1090 ćelija baznih stanica - kineski grad Shenzhen [27]

2.3.3 Diskretizacija Vremena

Suvremeni izvori podataka o kretanju bilježe vremenski žig uz podatak o položaju što ih čini pogodnima za vremenski zavisnu analizu. Zbog ograničene granularnosti CDR diskretizacija vremena je praktički neizbježna [23]. U literaturi je korištena različita vremenska rezolucija (razlučivost) - podjela na vremenske periode τ od 15 min, 30 min [29], 1 sat, 2 sata, 3 sata [30][23]. Putovanja ili zaustavljanja bivaju dodijeljena periodima u kojima se odvijaju. Još jedna od zastupljenih podjela je na specifične periode - *radni sati*, *noćni sati*, *komutacijski sati* [29][6][19] u kojoj se promatra gdje pojedinci pretežito borave u tipičnom radnom vremenu te tokom noći kroz cijelo razdoblje prikupljanja, \mathcal{T} .

Rekonstrukcija putanja

Kretanje pojedinca tokom dana najčešće je rascjepkano, odnosno u nizu postoje "rupe" kada je informacija o kretanju nepoznata zbog nepostojanja aktivnosti u tom vremenu.

Definicija 2.3.4. Potpunost putanje definirana je izrazom

$$potpunost = \frac{n_L}{\tau \times \mathcal{T}} \quad (2.1)$$

gdje je:

n_L broj zabilježenih položaja (zaustavljanja)

τ vremenski period

\mathcal{T} duljina ukupnog vremenskog razdoblja za koje su dostupni zapisi.

Idealna potpunost je postignuta ako postoji barem jedan zabilježeni položaj (zaustavljanje) po periodu. [23]

Veća potpunost putanja postiže se ako se osim poziva kao komunikacijskih aktivnosti u obzir uzmu SMS poruke i prijenos mobilnih podataka [23]. **Odbacivanje korisnika s niskim stupnjem potpunosti zastupljeno je u istraživanjima razvoja POM iz CDR što dovodi do iskorštavanja tek vrlo malog dijela korisnika - tek 1.67% i 0.45% u nekim istraživanjima [23].**

Rekonstrukcija putanja na osnovu aktivnosti u cijelom vremenskom razdoblju promatranja \mathcal{T} (npr. 7, 15, 30, 90 dana) nameće se kao novi postupak koji povećava broj i točnost rekonstruiranih putanja i smanjuje mogućnost povratnog inženjeringa identiteta. Rekonstrukcijom putanja dolazi do povećanja broja istih putanja (više pojedinaca) koje se odvijaju isto vrijeme, čime se štite pojedinci koji prije postupka jedini imaju veliku potpunost putanja. [23]

Poglavlje 3

Pregled postupaka i istraživanja o vrednovanju Polazišno-Odredišnih Matrica

Postupak vrednovanja je objektivna procjena kvalitete, vrijednosti objekta koji se vrednuje. Pregledom literature o vrednovanju POM utvrđeno je da se jednoznačno koriste pojmovi točnost, pouzdanost i kvaliteta POM-a. Vrednovanje POM-a vrši se postupcima koji koriste mjere razine sličnosti, odnosno razlike (greške) s referentnom matricom - a temelj je razlika na razni svakog elementa POM.

Tradicionalni postupak vrednovanja POM-e svodi se na određivanje njene sličnosti s *ground truth* matricom [31]. Ground truth POM-a je POM-a koja je uzeta kao ona koja najtočnije opisuje stvarno stanje. Ground truth POM-a je dobivena tradicionalnim postupcima (anketiranje i/ili prebrojavanje vozila). Statističke mjere kvantiziraju razliku procijenjenih i pravih vrijednosti (vidi dodatak A - Postojeće metrike za vrednovanje Polazišno-Odredišnih Matrica). Iz statističkih mjera sličnosti izdvojen je parametar r^2 koji je opisan u narednom potpoglavlju.

Upravo ta *ground truth* POM-a je ujedno i početna POM-a (iz odlomka 2.2.3) u postupku procjene nove POM-e iz rezultata prebrojavanja vozila [13]. POM će biti vrednovana usporedbom s POM-om koja se koristi u postupku procjene te iste POM-e. Postupak vrednovanja nove POM-e

stoga je ovisan o kvaliteti početne POM-e, koja je procijenjena u odnosu njenu početnu matricu i tako rekurzivno.

3.0.1 Linearna regresija - mjera r^2

Tehnika vrednovanja linearnom regresijom procjenjuje vrijednosti ćelija jedne POM-e (zavisna varijabla - y) na temelju vrijednosti druge - referentne *ground truth* POM-e (nezavisna varijabla - x). Vrijednosti u POM promatraju se kao skup vrijednosti $i = 1, 2, 3, \dots, n^2$ ne uzimajući u obzir njihovu matričnu strukturu. Pretpostavka kod linearne zavisnosti je da postoje koeficijenti a i b takvi da se za svaku vrijednost nezavisne varijable x_i zavisna varijabla y_i može napisati kao:

$$y_i = bx_i + a + \epsilon_i \quad (3.1)$$

gdje je :

pogreška ϵ slučajna varijabla s normalnom raspodjelom i očekivanom vrijednošću 0
 b koeficijent smjera pravca linearne regresije
 a slobodni član (predstavlja križište linije regresije s y osi).

[32] (vidi sliku 3.1)

Korelacijski koeficijent određenja ili obuhvaćanja varijance r^2 je mjera koja pokazuje koliko je slaganje između vrijednosti izračunatih modelom procjene i izmjerenih vrijednosti (*ground truth*). Općenito, to je mjera dobrote pristajanja (engl. *goodness of fit*) koja govori koliko uspješno model opisuje zavisnost (korelaciju) između zavisne i nezavisne varijable. r^2 predstavlja omjer modelom opisanih varijacija u odnosu na ukupne varijacije podataka:

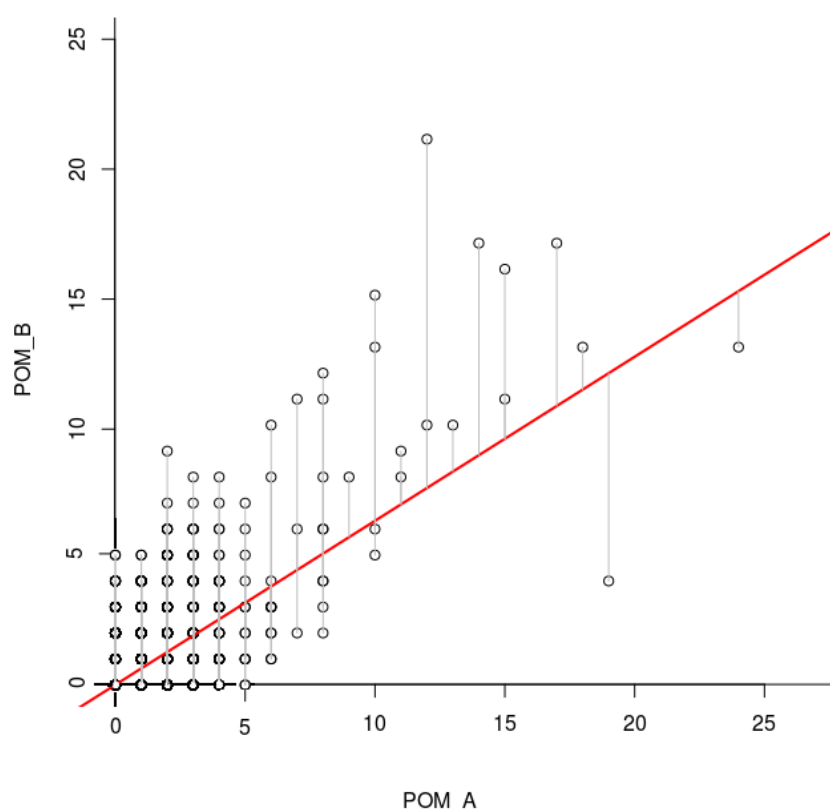
$$r^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{\sum (y_i - \bar{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (3.2)$$

Raspon mjere r^2 je od 0 do 1, 0 ukazuje na nepostojanje objašnjenja varijance između izračunatih i izmjerenih podataka, a 1 ukazuje da je varijanca u potpunosti objašnjena modelom.

Ako su matrice potpuno iste, vrijednost r^2 iznosi 1, vrijednosti a 0 i vrijednost b 1. Što su vrijednosti bliže navedenima, matrice su sličnije [32]. Kod usporedbe 2

POM-e u odnosu na referentnu *ground truth*, ona koja je sličnija referentnoj bolje opisuje kretanje.

Zadovoljavajuća vrijednost r^2 u usporedbi s *ground truth* navodi se 0.99 [32]. Ta vrijednost ne smatra se praktičnom za usporedbu POM-a dobivenih različitim pristupima procijeni što dolazi do izražaja i kada se radi o POM-i iz suvremenih izvora [32].



Slika 3.1 Primjer linearne regresije između vrijednosti dvije POM. Vrijednosti POM-e A su na x i vrijednosti POM-e B su na y osi. Pravac linearne regresije označen je crvenom, reziduali sivim linijama.
 $a = 0.0071273$, $b = 0.631128$, $r^2 = 0.4026$

3.0.2 Vrednovanje suvremenih Polazišno-Odredišnih Matrica

Unaprjeđenje metodologije procjene POM-e koja će pružiti povećanu kvalitetu i transparentnost uz ispunjavanje suvremenih zahtjeva na tehnologije (cijena, zaštita privatnosti i podataka itd.) privlači pažnju i van područja prometne znanosti. Suvremeni pristupi koriste alternativne izvore podataka o kretanju za promatranje obrazaca putovanja i pristupe procjeni POM-a, no prema [33] nailaze na problem da prometna struka zahtijeva da se njihova vrijednost dokazuje usporedbom s tradicionalnim pristupima, dobro poznatim u praksi. Za tradicionalnu usporedbu suvremene POM čija prostorna podjela nije TAZ s POM čija je prostorna podjela na TAZ traži da se prostorna podjela prilagodi nekoj novoj ili prostornoj podjeli jedne od matrica te da se nakon toga matrice usporede *elementwise*. Prilagodba se u praksi čini agregacijom i/ili interpolacijom tokova. (Detaljnije u Dodatku B - Prilagodba prostorne podjele.)

3.0.3 Grafički oblik i sličnost strukture Polazišno-Odredišnih Matrica

Vrijednosti u matrici mogu se normalizirati na opseg vrijednosti koji je moguće prikazati u obliku slike tako da svaka vrijednost predstavlja vrijednost piksela. Različite palete korištene su za prikaz raspona, od raspona 0-100 koji predlažu [30], prikazom nijansama jedne boje 0-255 engl. *grayscale* [34], paletom nijansa 2 boje, ili kojom drugom proizvoljnom paletom s više boja .

Nekolicina autora ističe važnost strukturalne sličnosti s referentnom matricom kao važnu mjeru kvalitete POM-e jer visoka razina strukturalne sličnosti može biti prisutna i kod POM-a s manjom razinom sličnosti prema statističkim mjerama. Također, strukturalna sličnost je (vizualno) vidljiva u grafičkom obliku POM-e. **Dobro odgovara ljudskoj vizualnoj percepciji sličnosti slike.**

Više-razinski indeks strukturalne sličnosti

Više-razinski indeks strukturalne sličnosti (engl. Mean Structural Similarity index) (MSSIM) dolazi iz područja računalne obrade slike i koristi se kao mjera usporedbe

digitalnih slika (*eng. measure of comparison*). Ideja o korištenju MSSIM u analizi urbanog kretanja stanovništva za mjerenje sličnosti POM-a prvi puta se spominje i demonstrira na simuliranim POM-ma dobivenim iz referentne POM-e dodavanjem šuma. [35]

Informacija o strukturi slike definira se kao atributi slike koji predstavljaju strukturu objekata na sceni, i neovisni su o prosječnom osvjetljenju i kontrastu. Jer osvjetljenje i kontrast mogu znatno varirati na sceni, moraju se u obzir uzeti samo njihove lokalne vrijednosti.

Indeks strukturalne sličnosti (engl. Structural Similarity Index) (SSIM) se bazira na degradaciji strukturalnih informacija na jednoj slici u usporedbi s drugom (referentnom) slikom. SSIM se računa za svaki kvadratni blok veličine N elemenata na način da se jezgra (da bi obuhvatila novi blok) pomiče ćeliju po ćeliju dok ne prođe preko cijele slike. MSSIM je srednja vrijednost svih SSIM.

Osnovni višerazinski indeks strukturalne sličnosti

Neka su X i Y matrice koje uspoređujemo a $x = \{x_n | x = 1, 2, \dots, N\}$ i $y = \{y_n | y = 1, 2, \dots, N\}$ parovi vrijednosti kvadratnih prozora veličine jezgre na istim pozicijama u X i Y ; $SSIM$ je određen prosječnim vrijednostima μ_x i μ_y sa standardnim devijacijama σ_x i σ_y i kovarijancom σ_{xy}

$$l(x, y) = (2\mu_x\mu_y + C1)/(\mu_x^2 + \mu_y^2 + C1) \quad (3.3)$$

$$c(x, y) = (2\sigma_x\sigma_y + C2)/(\sigma_x^2 + \sigma_y^2 + C2) \quad (3.4)$$

$$s(x, y) = (\sigma_{xy} + C3)/(\sigma_x\sigma_y + C3) \quad (3.5)$$

Izraz $l(x, y)$ opisuje razliku u osvjetljenju, $c(x, y)$ razliku u kontrastu, a $s(x, y)$ razliku u strukturi između x i y . $C1$, $C2$ i $C3$ su konstante uvedene da se izbjegne "nestabilnost" kada su nazivnici bliski 0. Opća forma $SSIM$ definira se kao

$$SSIM(x, y) = [l(x, y)^\alpha][c(x, y)^\beta][s(x, y)^\gamma] \quad (3.6)$$

gdje su α , β i γ parametri relativne važnosti svake komponente. Za $SSIM$ vrijedi slijedeće:

$$SSIM(x, y) \leq 1 \quad (3.7a)$$

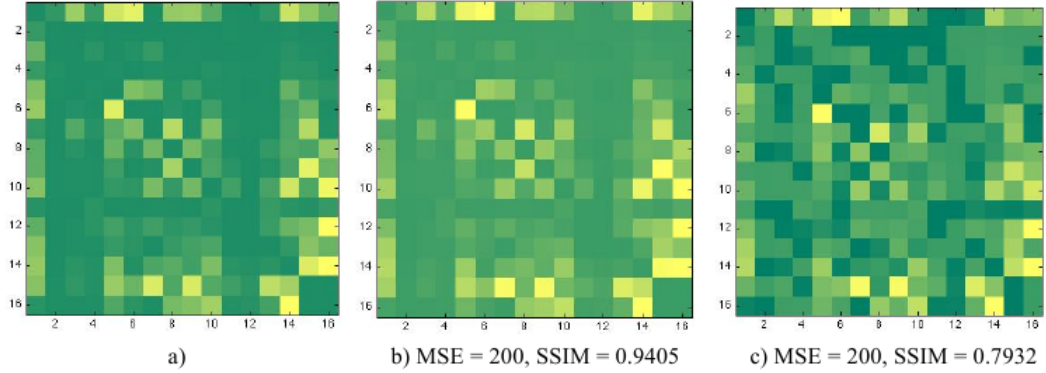
$$SSIM(x, y) = SSIM(y, x) \quad (3.7b)$$

$$SSIM(x, y) = 1 \iff x = y \quad (3.7c)$$

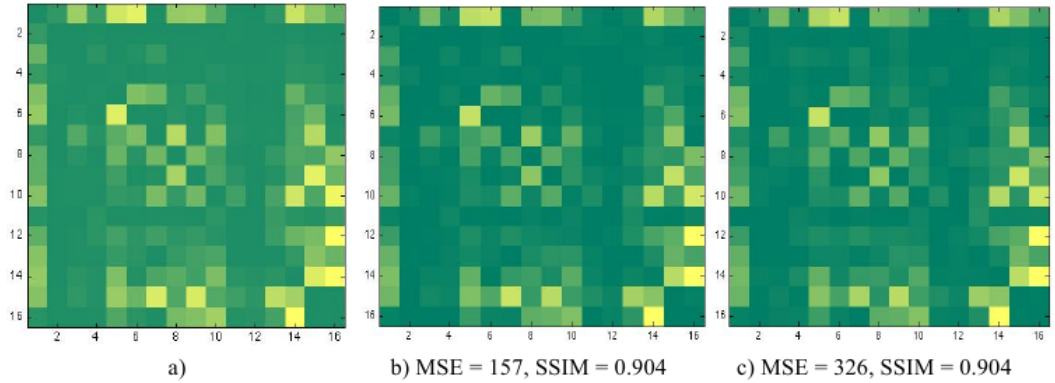
$$MSSIM(X, Y) = \frac{1}{M} \sum_{m=1}^M SSIM(x_m, y_m) \quad (3.8)$$

Poboljšani višerazinski indeks strukturalne sličnosti

Nekoliko godina nakon prvog spominjanja $MSSIM$ kao metrike usporedbe POM-a [34] se dotiče 3 problema postavljajući pitanja: Koliko treba biti veliki blok? Kako usporediti "guste" i "rijetke" POM-e? Koja je prihvatljiva vrijednost $MSSIM$? Autori definiraju poboljšani model koji nazivaju $4D-MSSIM$ u izračun kojeg dodaju stvarne euklidske udaljenosti prostornih zona. Tehnike bazirane na strukturalnoj sličnosti i njihov doprinos još su uvijek otvorene za raspravu. [33][36][34]



- (a) U odnosu na referentnu *ground truth* POM-u a), vrijednost metrike točnosti MSE jednak je za obje POM-e b) i c), dok se indeks strukturalne sličnosti znatno razlikuje, gdje POM-a b) ima veću strukturalnu sličnost s a) nego što ima c). Podsjetimo se da za SSIM vrijedi (3.7). Lako je na matrici c) *golim okom* uočiti narušenost strukturalne sličnosti.



- (b) U odnosu na referentnu POM-u a), vrijednost metrike strukturalne sličnosti iznosi 0.904 za obje POM-e b) i c), dok se MSE razlikuje (veći MSE znači veću razliku). Obje POM-e dobivene su dodavanjem Gaussianovog šuma na referentnu matricu.

Slika 3.2 Narušenu strukturalnu sličnost lakše je uočiti nego narušen MSE. SSIM (i MSSIM) u odnosu na referentnu POM-u može biti korišten kao dodatna informacija (*goodness of fit* mjera) prilikom vrednovanja POM-e [35]

Poglavlje 4

< Naslov >

4.1 Uvodna razmatranja i pregled istraživanja

Širenje gradova i rast stanovništva rezultirali su rastućim brojem sve ozbiljnijih prometnih zastoja u gradovima diljem svijeta. Strategije upravljanja prometom i gradom traže uvažavanje dinamike razvoja stanovništva i njegove migracije u kontekstu suvremenih društvenih i gospodarskih kretanja. Djelotvorno upravljanje i kontrola prometa doprinose povećanju sigurnosti, kvalitete prometne i kvalitete transportno-logističke usluge, potiče ekonomski rast i smanjuje zagađenje okoliša. Suvremeni društveno-ekonomski odnosi promijenili su prirodu kretanja ljudi i dobara zbog čega tradicionalni pristup postaje zastario i neprimjeren.

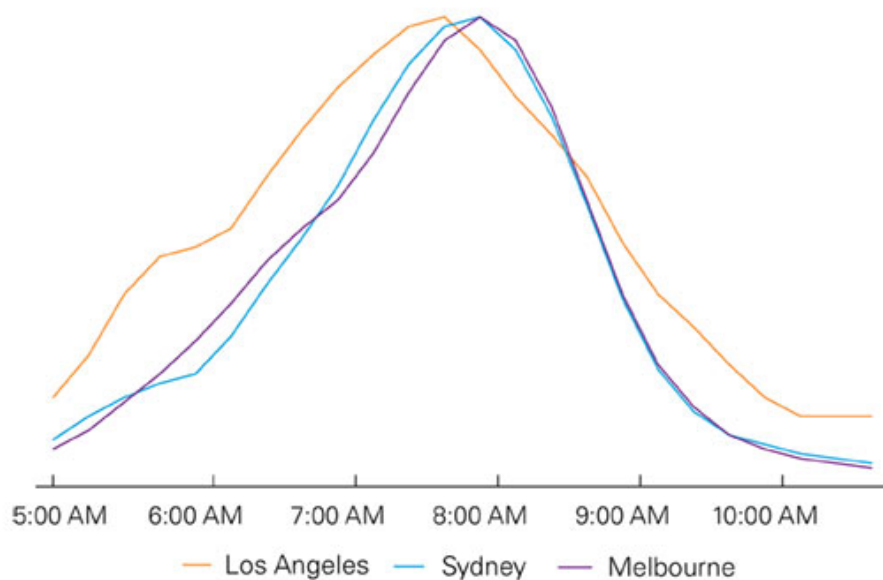
Analiza prometne potražnje tijekom dnevnih vrhunaca i pokazatelja kako će se taj vrhunac mijenjati u bliskoj i dalekoj budućnosti zahtjeva odabir dobre vremenske rezolucije. Tradicionalni modeli procjene pretpostavljaju da se širina vremenskog vrhunca u budućnosti neće mijenjati iako se ona već sad razlikuje od grada do grada (Slike 4.1) [20].

Pretpostavka da je referentna matrica *a priori* točna ne omogućuje da se sagleda je li nova POM-a točnija i/ili kvalitetnija od nje. Kako prikazati kvalitetu POM-e bez oslanjanja na *ground turth* koja porastom značajnijih i bržih promjena povećava utjecaj pristranosti kretanju u prošlosti? Postoji potreba za objektivnom procjenom kvalitete POM-a definiranjem odnosnih parametara kva-

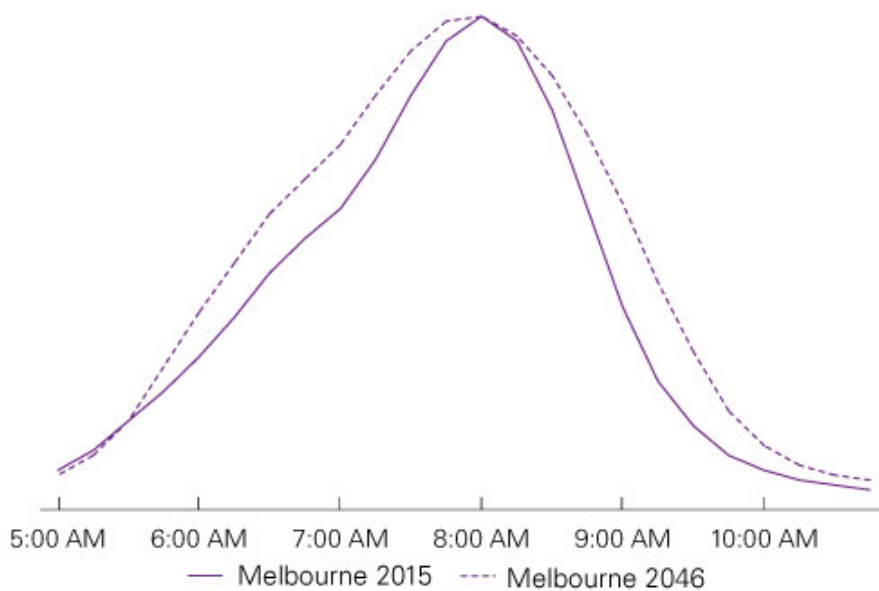
litete.

CDR se ističe kao izvor gotovo jednako zastupljen u zemljama u razvoju i razvijenim zemljama, s penetracijom oko 40%-50% stanovništva za pojedine operatore javne pokretne mreže [9]. **Praksa odbacivanja ogromnog dijela uzorka zbog niske potpunosti putanja može se zamijeniti rekonstrukcijom putanja uvažavajući njihov kontekst [23].**

Analiza kretanja stanovništva koja uzima u obzir i kontekst proširuje područje primjene zaključaka analize kretanja stanovništva van prometa [37]. Mogućnost (naknadnog) preciziranja odredišta s razine zone na razinu zgrade i određivanje svrhe (motiva) kretanja nije obuhvaćeno procjenom kvalitete POM.



(a) Širina jutarnjeg vrhunca prometne potražnje u 3 grada (oznake vremena su u lokalnom vremenu za svaki grad)



(b) Predviđena širina vrhunca prometne potražnje

Slika 4.1 Širina jutarnjeg vrhunca prometne potražnje - vremenski zavisno modeliranje [20]

4.2 Kontekstualizacija Polazišno-Odredišnih Matrica

Definicija 4.2.1. Putovanje kao informacijski objekt je entitet kojeg čine: polazišni položaj iz kojeg je putovanje krenulo, odredišni **položaj** u kojem se putovanje završava, **vremenske oznake** kada je počelo i kada je završilo, a kada je poznat, u navedeno se ubraja i **kontekst** putovanja.

Definicija 4.2.2. Kontekst putovanja je skup podataka koji definiraju polazište, odredište, **način prijevoza**, **svrhu** i **trajektoriju** putovanja.

Kontekst može biti izvučena iz samih podataka o kretanju ili iz dodatnih vanjskih izvora podataka. Putovanja svake kontekstne kategorije čine po jednu kontekstualiziranu POM-u.

Kontekst specificira socio-ekonomske aktivnosti koje su pokretači putovanja odnosno motivatori individualne i grupne mobilnosti. Preciznost određivanja konteksta definira izbor metode procjene POM-e.

Ako svrha putovanja nije izvorno navedena kao opisna veličina (prediktor) u ulaznim podacima za razvoj POM-e, moguće je o kontekstu putovanja zaključiti iz baza neizravnih kontekstualiziranih podataka trećih strana (npr. zapisi o korištenju telekomunikacijske mreže dopunjeni prostornim podacima Geoprostornog Informacijskog Sustava (GIS)) uz korištenje odgovarajućih postupaka prostorne statistike i analitike.

4.2.1 Primjeri korištenja konteksta u analizi kretanja

Studija A - 3 tipa privlačenja

Alhazzani i ostatak autora 2016. kvantizirali su vezu između ukupno 23 namjenske kategorije Točaka Interesa (engl. Points of Interest, POIs) iz službene gradske baze (npr. tvornice, bolnice, javne škole, religijski objekti, hoteli, knjižnice, sportski objekti) i onoga što nazivaju 3 tipa privlačenja. Dolaze do zaključka da **postoje 3 osnovna tipa privlačenja od kojih svaki ima karakterističan tok (gustoću ukupnog toka, udaljenost i disperziju polazišta tokova usmjerenih prema**

objektu tog tipa privlačenja). Svaku od kategorija karakterizira jedan od ta 3 tipa privlačenja. U radu koriste podatke iz službene baze s 12,000 POIs i CDR zapisima iz razdoblja od mjesec dana [8]. Zone su okarakterizirane objektima koji se nalaze u njima.

Studija B - Model Raspodjele Toka

Inovativni pristup kontekstualizaciji toka korištenjem OpenStreetMap baze predstavljen je u radu sa sudjelovanjem autora ovog diplomskog rada [24]. Prostorni objekti na području interesa kategorizirani su na osnovu pripadajućih opisnih podatak prema tipu socio-ekonomske aktivnosti u koju su uključeni (vidi dodatak D - OpenStreetMap). Definirano je ukupno 6 kategorija: Dom (Home), Posao (Work), Zdravlje (Health), Edukacija (Education), Zabava (Leisure) i Ostalo (Other). **Razvijen je vjerojatnosni model koji, na osnovu broja objekata pojedine kategorije u odredišnoj zoni i promatranog vremenskog okvira, ukupni ulazni tok u odredišnu ćeliju dijeli na 6 tokova usmjerenih prema objektima tih kategorija.** Jedna POM dijeli se na 6 kontekstualiziranih POM-a, po jednu za svaku od kategorija. POM-e dobivene iz javno dostupnih, anonimiziranih telekomunikacijskih zapisa na području kineskog grada Shenzhena korištene su u demonstraciji modela [30][27]. Definiranjem konteksta putovanja na ovaj način sačuvana je mogućnost postizanja dobre rezolucije u vremenu, a istovremeno je dan kontekst koji uključuje tipične komutacijske *Home* i *Work* kategorije, te umjesto općenite kategorije *Other* precizira 4 kategorije.

Definiranjem konteksta ulaznih tokova definiraju se nove informacije o kretanju unutar ćelije, odnosno o konačnoj destinaciji nakon ulaska u ćeliju. Saznaje se o internom kretanju, čak i za metodu procjene koja bez kontekstualizacije nema preciznost određivanja položaja veću od prostorne zone (CDR).

Studija C - Planiranje održivih gradova

Modeli za procjenu potrošnje energije na razini zgrada, a posljedično i na razini gradova temelje na prostornom kapacitetu zgrada, nedovoljno uvažavajući činjenicu da

ispunjenost tog kapaciteta varira u vremenu. **Analizom kretanja stanovništva moguće je procijeniti ispunjenost kapaciteta zgrada i odrediti prosječnu potrošnju energije.** Korištenje CDR zapisa prikladno je za ovakvu analizu potrošnje energije na razini grada zbog velikog uzoraka stanovništva i neovisnosti o prometnoj infrastrukturi. Zgrade iz baze zgrada na području od interesa (grad Boston) podijeljene su u 3 tipa - stambene, industrijske i komercijalne. Model uzima u obzir ukupnu površinu zgrada i "radno vrijeme" zgrada koje nisu stambene. Zgrade u zonama su u procesu grupirane u blokove unutar zone, a vjerojatnosnim modelom ulazna putovanja su dodijeljena blokovima, a potom pojedinoj zgradi u bloku. Ovisno o pretpostavljenom odnosu između opterećenja i kapaciteta zgrada, u odnosu na standardne modele procjena potrošnje energije razlikuje se od +1% do - 15% za stambene te -4% do -21% za komercijalne zgrade [37] (standardni modeli precijenjuju potrošnju energije).

Studija D - Predviđanje odredišta u stvarnom vremenu

Studija se bavi predviđanjem odredišta za vrijeme trajanja započetog putovanja u pseudo-stvarnom vremenu na osnovu prepoznavanja svrhe putovanja. Svrha putovanja određuje se metodama strojnog učenja iz podataka o korištenju zemljišta (POIs) i poznatih uzoraka putovanja (model uči iz 5, 15 ili 30 GPS putanja prethodnih putovanja istog pojedinca). Predstavljen je unaprijeđeni model predviđanja odredišta koji koristi svrhu putovanja kao kontekst u određivanju krajnjeg odredišta. Rezultati pokazuju značajno poboljšanje u točnosti i brzini predviđanja u odnosu na predviđanje bez uzimanja svrhe u obzir. [38]

Studija E - Rekonstrukcija rijetkih trajektorija

[23]

4.3 Teza

Postoji skup parametara kojima se objektivno mogu usporediti dvije na različiti način ili na isti način dobivene POM-e.

Rad nastoji tezu potkrijepiti opisom metodologije za vrednovanje matrica prema njihovoj kvaliteti uvažavajući definirane odnosne parametre kvalitete.

4.3.1 Alternativni pristup vrednovanju Polazišno-Odredišnih Matrica

Točnost, rezolucija i obujam podataka su ključni parametri geografske informacije i znanje o čemu oni zavise je nužna baza za veće pothvate procjene vrijednosti geografskih informacija kao društvenog dobra [39].

Vrednovanje POM-e je proces koji ima za cilj procjenu obilježja kvalitete POM-e i znači identifikaciju obilježja kvalitete POM-e, njenog sadržaja, njene smislenosti, a izražava se slijedećim indikatorima: prostorno obuhvaćanje, rezolucija (prostorna , vremenska i tematska), zrnatost (prostorna , vremenska i tematska), putovanje kao parametar, gustoća informacija i širina toka.

4.4 Odnosni parametri kvalitete

4.4.1 Prostorno obuhvaćanje

Geoinformacijska znanost (geoinformatika) definira prostorno obuhvaćanje kao:

Definicija 4.4.1. Prostorno obuhvaćanje je geografsko područje gdje su prikupljeni podatci, mjesto koje je predmet opažanja ili lokacija u fokusu aktivnosti. [40]

Definicija u ovom obliku odgovara definiciji prostornog obuhvaćanja kao parametra kvalitete POM. Prostorno obuhvaćanje kao kategorička mjera za kvalitetu POM podrazumijeva kategorije:

1. isključivo prometna infrastruktura (ceste, željeznica, javni prijevoz)

2. područje neovisno o prometnoj infrastrukturi (CDR, GNSS)

Prostorno obuhvaćanje uvjetuje koji načini kretanja su obuhvaćeni POM-om.

Većina zemljama ima neki oblik zakona o zabrani korištenja mobilnih telefona tijekom uspravljanja vozilom [41] istraženo odražava li se to i kako na zastupljenost takvih putovanja u POM-i dobivenih iz CDR.

4.4.2 Rezolucija

Geoinformacijska znanost rezoluciju općenito definira kao količinu detalja u setu podataka [39]. Degbelo u radu "Pet općih svojstava rezolucije" [39] kao jedno od svojstava navodi da je **rezolucija (geografskih podataka) pozitivno korelirana s točnošću i količinom podataka**. Što je veća razina detalja u prikazu, veća je sličnost tog prikaza 'istini' (savršenom prikazu). Što se više detalja pohranjuje, veća je količina podataka.

Isti rad napominje da *prisutnost rezolucije otvara pitanje "Koja je optimalna rezolucija i postoji li optimum?"*. Trud tek treba biti uložen u razvoj potpuno razrađene teorije optimalne **prostorne, vremenske i tematske rezolucije** koja će uzeti u obzir specifičnosti prikupljanja podataka (osjetila, ljudska osjetila, anketa, jesu li podaci proizašli iz postojećih opažanja ili ne) i zadatak od interesa [39].

Vremenski nezavisna POM ima prostornu, a vremenski zavisna POM prostornu i vremensku rezoluciju, kontekstualizirana ima tematsku rezoluciju. Alternativni pristup vrednovanju POM definira prostornu i vremensku rezoluciju kao:

Definicija 4.4.2. Prostorna rezolucija je najmanja prijedena udaljenost koju je moguće identificirati.

Udaljenost je kontinuirana veličina, a radi pojednostavljenja metodologije definirana je kao kategorička s kategorijama:

1. $< 100m$
2. $100m - 500m$
3. $500m - 1km$

Poglavlje 4. < Naslov >

4. $1km - 2km$

5. $2km - 3km$

6. $> 3km$

Definicija 4.4.3. Vremenska rezolucija je najmanji vremenski odmak koji je moguće identificirati.

Kod suvremenih pristupa vremenska rezolucija ovisit će o periodu uzorkovanja položaja.

Tematska rezolucija

Geoinformacijska znanosti definira tematsku rezoluciju kao količinu kategoričkih detalja na *tematskoj karti*. Izražava se brojem tipova pokrova (engl. land cover) ili brojem tipova namjene zemljišta (engl. land use) [42]. Ovakva definicija dobro se uklapa u tematsku rezoluciju kao parametar kvalitete kontekstualiziranih POM koje su procijenjene na temelju namjene zemljišta ili objekata. Razvrstavanjem putovanja u zasebne POM prema kategorijama dobije se broj POM jednak tematskoj rezoluciji.

Definicija 4.4.4. Tematska rezolucija svrhe je broj kategorija svrhe putovanja definirane iz namjene krajnjeg odredišta putovanja.

Primjer kategorija i njihovih tematskih rezolucija:

- dom, posao (2)
- stambeni, industrijski, komercijalni (3) [37]
- dom, posao, edukacija, razonoda, zdravlje, ostalo (6) [24]
- tvornice, ambasade, sveučilišti objekti, bolnice, javne škole, ostali obrazovni objekti, stanice javnog prijevoza, knjižnice, klinike, religijski objekti, hoteli, sportski objekti, banke ... (23) [8]

Definicija 4.4.5. Tematska rezolucija načina kretanja je broj kategorija načina kretanja po kojim je moguće razvrstati putovanja.

Primjer kategorija i njihovih tematskih rezolucija:

- nadzemna željeznica (gradski javni prijevoz) (1) - izvor zapisi transakcija sustava Pametnih Kartica (engl. *Smart Card*) za prijevoz [32]
- taxi (1) - izvor zapisi GPS putanja taksi vozila
- osobna vozila, kamioni, javni prijevoz (3)
- auto, vlak, avion (3) - izvor CDR [43]
- pješaci, osobna vozila, vozila javnog prijevoza, kamioni (4)

4.4.3 Zrnatost

Zrnatost je stupanj do kojeg je sustav rastavljiv u manje dijelove, bilo sam sustav ili njegov opis ili opažanje - stupanj do kojeg je veći entitet djeljiv. [44]

Za razliku od rezolucije prema [39], zrnatost i točnost su neovisne. Odgovori na pitanje "Gdje se nalazite?" "U Hrvatskoj", "U Rijeci", "U svojoj kući" i "Za svojim stolom" svaki je točan u smislu da govori 100%-tnu istinu'.

Prostorna zrnatost se kao odnosni parametar kvalitete izražava kroz broj dijelova na koje je područje podijeljeno.

Definicija 4.4.6. Prostorna zrnatost POM je broj dijelova (zona) na području od interesa. Iščitava se iz dimenzije POM.

Prostorna Zrnatost ovisit će o prostornoj podjeli područja, gustoći baznih stanica, stupnju agregacije zona i prostornoj rezoluciji.

Definicija 4.4.7. Vremenska zrnatost je broj vremenskih okvira u jednom danu. Iščitava se iz broja POM procijenjenih za vremensko razdoblje $\mathcal{T} = 1$ dan.

Istaknute su one kategorije koje su zastupljene u literaturi. Vremenski okvir i pripadajuća zrnatost:

1. < 15 min (> 86)
2. **15 min (86)** [10][7]
3. 20 min (72)

Poglavlje 4. < Naslov >

4. 30 min (48) [29]
5. **1 h (24)** [45][19][43]
6. 2 h (12)
7. **3 h (8)** [30][23]
8. 4 h (6)
9. 6 h (4)
10. **8 h (3)**
11. 12 h (2)
12. 24 h (1) [43]

Najveća Vremenska Zrnatost koju je moguće postići procjenom POM ovisit će o vremenskoj rezoluciji u kojoj postoje podaci za procjenu.

Definicija 4.4.8. Tematska zrnatost je najmanja prostorno razlučiva jedinica koja ima tematsku rezoluciju.

1. zona
2. objekt [37] [24]

Objekt u zoni je najmanja prostorno razlučiva jedinica kod procjene matrica koja uključuje kontekstualizaciju toka na temelju objekata u odredišnoj zoni (studija B 4.2.1).

4.4.4 Putovanje kao parametar

Putovanja koja se protežu kroz više perioda u pravilu se dodjeljuju vremenskom okviru u kojem započinju ili u kojem završavaju. Autori [10] [30] [19] sva putovanja dodjeljuju intervalu u kojem je putovanje započeto, dok neki procjenjuje obje vrste POM-a [43]. Naredni odjeljci ne daju definiciju putovanja kao parametra već samo ističu problematiku koja se javlja kod definicije putovanja općenito.

Ograničenja na putovanja iz istog izvora podataka - Charging Data Records

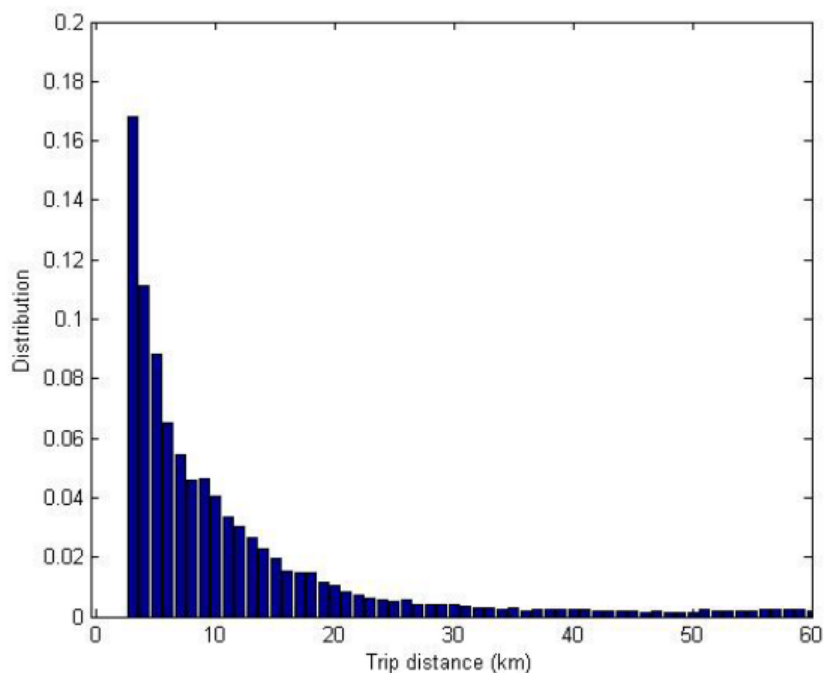
Putovanja kod vremenski zavisnih POM-a kreiranih iz CDR moguće je svrstati u nekoliko kategorija na osnovu prihvatanja ili odbacivanja određenih kretanja kao putovanja. Putovanjima se ne smije obuhvatiti lažno kretanje uzrokovano uravnoteženjem opterećenja i lažno zaustavljanje uzrokovano neuniformnom gustoćom zapisa pojedinaca u vremenu, što različiti autori nastoje postići postavljanjem navedenih ograničenja:

1. putovanja u sve ćelije, bez internih putovanja (bez dijagonale)
2. putovanja u sve ćelije osim u susjedne
3. agregacija susjednih ćelija - POM-e s internim putovanjima

Definicija 4.4.9. Interno putovanje je svako putovanje koje se odvija unutar jedne zone, odnosno čije se polazište i odredište nalaze unutar iste prostorne zone. Suprotno internom je eksterno putovanje.

4. putovanja ograničena minimalnom i/ili maksimalnom udaljenosti [46]
5. putovanja ograničena minimalnom duljinom trajanja zaustavljanja
6. putovanja ograničena minimalnom i/ili maksimalnom duljinom trajanja kretanja
7. putovanja ograničena brojem događaja koji potvrđuju zaustavljanje i maksimalnim vremenskim razmakom između dva uzastopna događaja

Za 2 različite definicije putovanja, za POM-e iz identičnih podataka na području Senegala generirano je u prosjeku 0.7 putovanja dnevno po korisniku u jednom, a 0.06 u drugom. Ograničenja su bila: minimalna udaljenost 3 km, maksimalna 100 km 4 u prvom i vremensko na 1 sat minimalnog trajanja (kategorija 6) i prostorno na udaljenost od barem 5 udaljenosti baznih stanica (kategorija 4). Najveća je bila zastupljenost kratkih putovanja. [19] Distribuciju možemo vidjeti na slici 4.2



Slika 4.2 Distribucija udaljenosti putovanja u kilometrima, Senegal 2016. [19]. U studiji su izbačena putovanja kraća od $3km$ (prostorna rezolucija $> 3km$)

O ograničenjima na putovanja ovisit će njihov konačan broj, odnosno širina toka.

Kontekstualizacija

Komutacijska putovanja (engl. *commuting trips*), odnosno putovanja na posao i putovanja kući značajna su po tome što su regularna - slijede dnevni ritam [29], što ih čini predvidivima [15]. Goulding u svom priručniku za kreiranje POM iz CDR ističe da je važno definirati što se smatra putovanjem - zastane li netko na benzinskoj crpki na putu s posla kući, je li riječ o jednom putovanju ili dva? Prihvaćanjem svakog zaustavljanja kao svojevrsnog *kraja tranzicije* (primjerice ukidanjem ograničenja 5) te prebrojavanjem tranzicija dobiva se tranzitna POM (spomenuta u 2.1). Tranzitna POM uopće neće opisivati kretanja između polazišta i odredišta *u suštini* [15].

Ako nije poznat kontekst putovanja, nastoji se prema duljini zaustavljanja odrediti je li neko zaustavljanje doista kraj putovanja [15]. Poznavanjem konteksta putovanja, objekata u neposrednoj blizini zaustavljanja, omogućuje da ne temeljimo kraj putovanja isključivo na osnovu duljine trajanja zaustavljanja već da se u obzir uzme svrha zaustavljanja.

Prednost suvremenih izvora podataka s višednevnim ili višemjesečnim vremenom promatranja je uočavanje obrazaca i repetativnosti putovanja na dnevnoj ili tjednoj bazi [23]. Poznavanje tipičnog kretanja pojedinca u to vrijeme omogućuje se rekonstrukcija putovanja u onom periodu kada podaci nedostaju. Čen sva putovanja jednog pojedinca u prošlosti i budućnosti od promatranog putovanja naziva upravo *kontekstom* u metodlogiji opisanoj u radu [23].

Mjera koja bi govorila o kvaliteti odabrane definicije putovanja prema razmotrenoj problematici mogla bi biti omjer tranzitnih i ne-tranzitnih putovanja (putovanja sa svrhom) zabilježenih u POM, no kako izostaje konvencionalni dogovor oko definicije putovanja i tranzicije ova mjera nije obuhvaćena ovim radom.

-Kontekstualizacija i interna putovanja- * kontekstualizacija donosi da znamo interna kretanja/ putovanja između objekata

4.4.5 Gustoća informacija

Definicija 4.4.10. Gustoća informacija je omjer ispunjenih ćelija i Nul-ćelija u POM-i.

POM-e dobivene iz anketa u kućanstvima i iz Road Side Interview (RSI) rijetko su popunjene, odnosno većinski su ispunjene Nul-ćelijama (eng. *Null-cells*, *Zero-cells*) [5]. POM-e kreirane iz GPS ili CDR zapisa gusto su ispunjene, najčešće vrlo malim vrijednostima umjesto nula [34]. POM izrađene iz uzoraka nakon procjene se skaliraju da bi obuhvatile cijelu populaciju. Postupci skaliranja nisu obuhvaćeni ovim radom. Skaliranje uključuje množenje POM-e određenim faktorom, a u tom slučaju Nul-ćelija zapravo ne mijenja vrijednost.

Dva su uzroka praznim poljima, nepostojanje putovanja između tih parova zona ili su postupkom procjene putovanja nezabilježena.

4.4.6 Ukupna širina toka

Definicija 4.4.11. Ukupna širina toka je ukupan broj odlazaka/dolazaka po vremenskom okviru za cijelu matricu - suma svih elemenata POM-e. (vidi sliku 2.1 F_{ij}).

Svi dosad spomenuti parametri utjecat će na ukupnu širinu toka. Veličina i reprezentativnost uzorka stanovništva te k-anonimizacija.

Izdvajanje najaktivnijih korisnika (onih s najviše zapisa) zastupljeno je u istraživanjima procjene POM iz CDR [29] [6]. Calabrese [28] uzima samo slučajno odabranih 25% korisnika (njih 1 milijun) da pojednostavi analizu, a [46] samo 100,000, kako bi procijenjena POM bila usporediva s onom dobivenom iz ankete provedene na uzorku iste veličine. Rekonstrukciju putanja kao rješenje koje osigurava zadržavanje velikog broja (manje aktivnih) korisnika predlaže Chen (odlomak 4.2.1).

Slabo obuhvaćanje krajnjih dobnih skupina zbog slabe zastupljenosti mobilnih uređaja kod najstarije i najmlađe dobne skupine ne tumači se kao pristranost, uz napomenu da su to ujedno skupine sa generalno slabom mobilnosti u odnosu na ostale [9]. Goulding smatra da se eventualno postojanje pristranosti može smatrati zanemarivim kod tržišta gdje veliki postotak stanovništva (npr. 92%) posjeduje mobilni uređaj [15].

K-anonimizacija

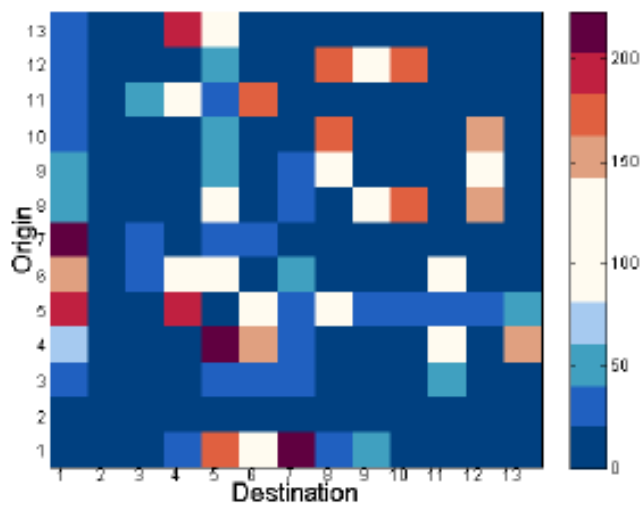
K-anonimnost (engl. *k-anonymity*) određuje najmanju dozvoljenu širinu toka koja se dozvoljava, u suprotnom se tok u potpunosti zanemaruje (postaje Nul-ćelija). Razlog je zaštita od eventualne mogućnosti identificiranja pojedinaca, pogotovo u vrlo rijetko naseljenim područjima [15]. K-anonimizacijom treba osigurati da tok u svakoj ćeliji bude veći od:

$$k \times m \tag{4.1}$$

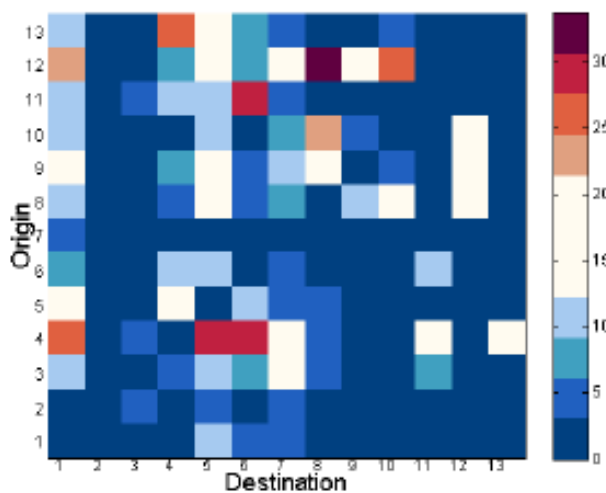
gdje je:

m - maksimalni broj putovanja za svakog pojedinca

k - odabrana vrijednost kojom se osigurava da na temelju širine toka pojedinac ne može biti razazan između $k - 1$ drugih pojedinaca.



(a) POM-a u kojoj je putovanje definirano kao zabilježeni pomak između 2 antene - *kratka putovanja*



(b) POM-a u kojoj je putovanje definirano kao promjena položaja u odnosu na položaj u prethodnom periodu. Položaj u periodu je onaj na kojem korisnik pretežito boravi. Za jedan period se pretpostavlja da je korisnik na prebivalištu, za drugi da je na radnom mjestu - matricu čine putovanja od prebivališta na posao engl. *commuting trips*

Slika 4.3 Dvije POM-e - produkti različitih definicija putovanja. Dobivene su iz potpuno istog izvora CDR zapisa. [19]

Poglavlje 5

Metodologija

opis sklopa/programske podrške. Opis rješavanja zadatka i odgovarajući proračuni

5.1 Uvjeti usporedbe

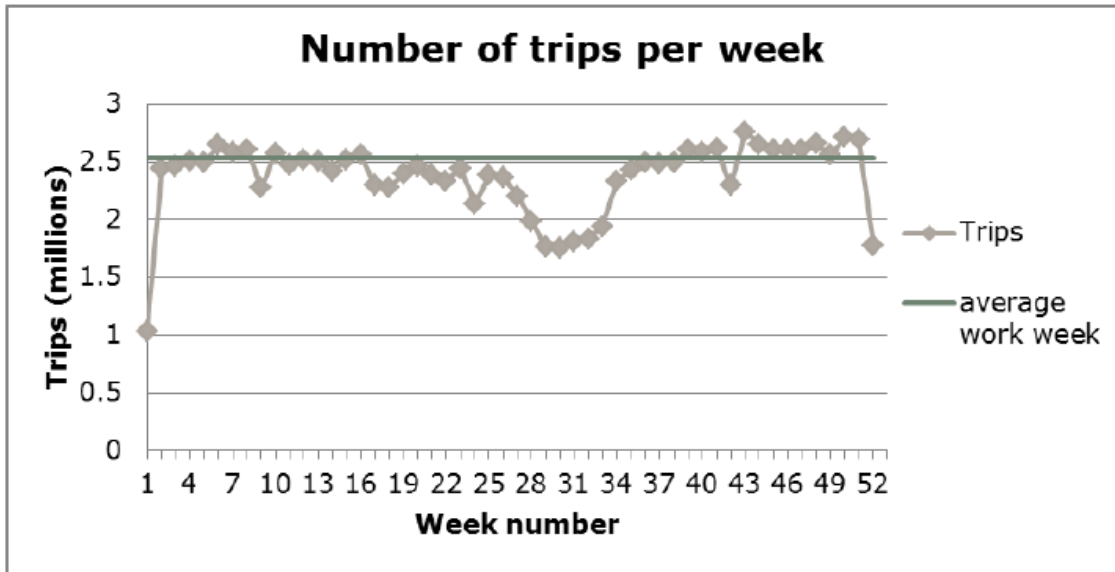
POM-e koje se uspoređuju trebaju ispunjavati uvjete usporedbe. Riječ je o slijedećim zajedničkim obilježjima:

- isto područje/grad/država
- isto doba godine

Zabilježenu varijaciju po tjednima u godini iz istraživanja procjene POM-a iz podataka o javnom prijevozu na području nizozemske možemo vidjeti na slici 5.1. Prednost novih postupaka procjene POM iz CDR u odnosu na ankete stanovništva je da mogu uhvatiti tjedne i sezonske uzorke [28]

- ako se radi o jednom danu - idealno isti dan u tjednu

Rezultati nacionalne ankete u Sjedinjenim Američkim Državama pokazuju da je prosječan broj dnevnih putovanja po stanovniku 4.18 tijekom radnog dana te 3.86 tijekom neradnih dana u tjednu, što je u skladu s zaključcima modernih postupaka procjene.[28][47]



Slika 5.1 Broj putovanja javnim prijevozom zabilježenih *OV-chipkaart* pametnim karticama u Nizozemskoj izražen u milijunima po tjednima u godini. Zelena linija označava prosječnu vrijednost radnog tjedna [32]

Osim očekivane razlike između radnih i neradnih dana, uočeno je da je petak, kao zadnji radni dan u tjednu, drugačiji od ostalih radnih dana [28][48][31]. Uočena je pojačana mobilnost petkom na razini regije (Boston Metropolitan Area) [28] i na razini cijele države (Obale Bjelokosti) s povećanjem od 35% u odnosu na nedjelju [48] na temelju POM-a izrađenih iz CDR. Uočena je smanjena mobilnost vezana za javni prijevoz (izvor za procjenu POM-a *Pametne kartice za javni prijevoz*) na razini regije (Southeast Queensland, Australija). Osim prosječnog broja dnevnih putovanja, petak ima i nešto drugačiji dnevni uzorak po satima (pojačanje prometa prema kraju radnog vremena započinje ranije).

Državni praznici i izvanredni događaji kao što su vremenske nepogode unose odstupanja od ustaljenih uzorka kretanja. Privremene turističke atrakcije i sportski događaji također mogu utjecati na odstupanje od uobičajenog uzorka kretanja. [36]

- ista definicija dana

Dio putovanja započinje u jednom a završava u drugom danu. Autori [29][6] zastupaju definiranje početka/završetka dana u 3:00 idućeg dana, umjesto u 0:00, kako dio putovanja ne bi bio izostavljen zato jer se proteže kroz 2 dana.

- ista kategorija putovanja (tranzitna, kontekstualizirana, *komutacijska*)
- obje POM-e napravljene objektivno na reprezentativnom uzorku

5.2 Usporedba prema odnosnim parametrima kvalitete

Kada se ustanovi da POM-e ispunjavaju uvjete usporedbe, uspoređuje ih se prema parametrima definiranim u 4.4.

Optimalne vrijednosti pojedinih parametara naposljetku će ovisiti o području primjene POM i njihovo određivanje zahtjeva dodatna istraživanja koja bi se nadovezala na metodologiju predstavljenu u ovom radu. Primjerice, rezolucija koja dobro odgovara za primjenu u jednom području može biti nepotrebno visoka za primjenu u drugu svrhu. Niža rezolucija može smanjiti količinu pohranjenih podataka i vrijeme izvođenja postupka procjene.

- Prostorno Obuhvaćanje: **Matrica je bolja ako je prostorno obuhvaćanje veće** (područje neovisno o prometnoj infrastrukturi > područje ovisno o prometnoj infrastrukturi)
- Gustoća Informacija: **Matrica je bolja što je gustoća informacija veća (što je manji broj Nul-ćelija)**
- Ukupna Širina Toka: **Matrica je bolja što je ukupna širina toka veća**
Podrazumijeva se isti vremenski okvir npr. od 6:00 do 9:00, cijeli jedan dan...
- Zrnatost: **Matrica je bolja što je zrnatost veća, pod uvjetom da je dovoljno mala da osigurava k-anonimnost na svakom toku POM-e.**

Poglavlje 5. Metodologija

K-anonimnost osigurava da identitet pojedinca ne može biti razazan od k-1 drugih pojedinac (4.4.6). Povećanjem bilo koje zrnatosti tokovi postaju sve uži. Uvažavanje k-anonimnosti postavlja ograničenje na zrnatost u svrhu osiguranja očuvanja garantirane privatnosti [15]. Tradicionalni postupci procjene POM (podatak za Ujedinjeno Kraljevstvo) dozvoljavaju širinu jednog toka između 12 i 15. Agencija za analizu kretanja iz CDR Teralytics [43] koristi k između 5 i 10.

Prostorna Zrnatost

Vremenska Zrnatost

Tematska Zrnatost

- Rezolucija: Odlomak 4.4.2 govori o pozitivnoj korelaciji rezolucije i točnosti. Odbacivanjem točnost kao jedinog pokazatelj kvalitete pitanje postaje koja rezolucija je optimalna za prikaz dnevnih urbanih migracija u POM. Niska zrnatost umanjuje potrebu za visokom rezolucijom, a visoka rezolucija omogućuje visoku zrnatost. Primjerice ako je vremenski okvir širine 12 sati (2 perioda u danu) nema potrebe da je početak i završetak putovanja određen u minutu, no ako možemo odrediti završetak putovanja u minutu, vremenski okvir teoretski može biti $< 15min$.

Rezolucija nije svojstvo POM koliko je svojstvo izvora podataka i postupka procjene.

Prostorna Rezolucija

Vremenska Rezolucija

Tematska Rezolucija

5.2.1 Postupak odlučivanja

Metodologija definira svaki parametar kao jednako vrijedan u donošenju konačne odluke o kvaliteti matrice.

Mogli biste uz definiciju ovih kriterija, samo sliku ubacite iz ovih nekih simu-

Poglavlje 5. Metodologija

lacija, Varaždina, Koprivnice, kao ilustraciju (bez odlaska u dubinu). Imate sliku koja pokazuje da...(ne bi to spadalo u rezultate?) možete staviti u definiranju kriterija odlučivanja- koprivnica bez susjednih putovanja i koprivnica sa susjednim, evo ZAŠTO je TO bolje (bez ičega drugog, kako su to radili drugi)

Treba raspisati do kraja kako ću odlučiti je li nešto bolje po ovom ili onom kriteriju - poglavlje 5 - i tu smo jasni, opet slika ili prebrojavanje - nema drugog. U poglavlju 5. napisat ćete samo da brojite, i gledate slike. A u poglavlju 6 ćete staviti slike i napisati rezultate zbrajanja- ne više od toga.

Simulirati kontekstualizaciju sa nekim random postocima na koje podijelim matricu?

I nakon toga na primjeru Čilea ili čega god hoćete

Koja ima najviše pluseva, ta će ispasti bolja- sve je jednako vrijedno (svaki parametar?!)

Poglavlje 6

Rezultati

iznošenje rezultata i analiza rezultata do kojih je došlo rješavanjem problematike zadatka, rezultati mjerenja

u poglavlju 6 ćete staviti slike i napisati rezultate zbrajanja- ne više od toga.

Tablica 6.1 Tablica usporedbe

POM	Prostorno Obuhvaćanje	Gustoća Informacija	Prostorna Zrnatost	Vremenska Zrnatost	Tematska Rezolucija Načina kretanja	Ukupna Širina Toka
A	neovisno o prometnoj infrastrukturi	0.008830772	1090×1090	8	0	30878
B	cesta	0.08683223	1090×1090	8	1	427646

Poglavlje 7

Diskusija

U diskusiji recite da ste uveli jedan svoj vlastiti pogled temeljen na vlastitim definicijama zato što je cijelo područje prilično neodređeno. Da ste vi napravili svoj okvir, egzaktna definicije pojmova, i da ste radili po tim definicijama, da ste napravili tezu, da ste po toj tezi definirali elemente za usporedbu i postupak odlučivanja. U kojem ste rekli, ako po ovom parametru dobijemo da je vrijednost veća da je matrica bolja i da ste na temelju toga izveli određene zaključke. Kažete što vam je procijenjeni nedostatak: da niste obuhvatili ili potrebno je još istražiti koliko klizeći prozori mogu utjecati na kvalitetu matrice, dali postoji neka mogućnost poboljšanja u samom postupku kreiranja marica (nemojte zalaziti puno u kreiranje i ići pisati ono što ste radili na kampu).

Pomični vremenski okvir, analiza dijelova matrice prema parametrima (dio koji je bolji, dio koji je slabiji) – > ispitati prisustvo neuniformnosti

Poglavlje 8

Zaključak

Establishing access policies for spatial information is a pressing societal need requiring a better understanding of all values involved. It is complicated by the fact that information about indoor and geographic spaces gets collected and shared by almost everybody. This phenomenon of crowd-sourced or Volunteered Geographic Information (VGI, (Goodchild 2007)) is profoundly altering the values related to spatial information, from economic as well as institutional, ethical, and legal perspectives. <http://www.economist.com/node/1788311>

Bibliografija

- [1] Prema slici. , s Interneta, <https://www.zukunft-mobilitaet.net/103615/analyse/datenbasierte-verkehrsplanung-big-data-mobilfunkdaten-optimierung-trajektorien/> (4.10.2019.).
- [2] Prema slici. , s Interneta, <https://transportgeography.org/wp-content/uploads/odgraph.png> (4.10.2019.).
- [3] N. Jelusic, “Telematicka sucelja (nastavni tekst),” 2016.
- [4] A. Peterson, “The Origin-Destination Matrix Estimation Problem- Analysis and Computations,” Ph.D. dissertation, Linköping Studies in Science and Technology, 2007.
- [5] M. Cools, E. Moons, and G. Wets, “Assessing the Quality of Origin-Destination Matrices Derived from Activity Travel Surveys,” Transportation Research Record: Journal of the Transportation Research Board, 2010.
- [6] J. L. Toole, S. Colak, B. Sturt, L. P. Alexander, A. Evskoff, and M. C. González, “The path most traveled: Travel demand estimaton using big data resources,” Transport Research Part C, 2015.
- [7] I. Espada. (2013.) Overview of the four-step transport demand model. Australian Road Research Board. , s Interneta, <https://www.youtube.com/watch?v=h2rxC-0rZLU> (20.9.2019.).
- [8] M. Alhazzani, F. Alhasoun, Z. Alawwad, and M. C. González, “Urban Attractors: Discovering Patterns in Regions of Attraction in Cities,” Public Library of Science, 2016.

BIBLIOGRAFIJA

- [9] P. Bonnel, E. Hombourger, A. maria Olteanu-Raimond, and Z. Smoreda, “Passive Mobile Phone Data to Construct Origin-Destination Matrix: Potentials and Limitations,” Transportation Research Procedia, Elsevier, no. 11, pp. 381.–398., 2015.
- [10] S. Bera and K. V. K. Rao, “Estimation of origin-destination matrix from traffic counts: the state of the art,” European Transport Trasporti Europei n. 49, 2011.
- [11] A. Kuharic, “Prikupljanje podataka o protoku vozila uporabom mobilnih telefona,” Master’s thesis, Fakultet prometnih znanosti, Sveuciliste u Zagrebu, 2016. , s Interneta, <https://zir.nsk.hr/islandora/object/fpz:683/datastream/PDF/>
- [12] M. S. Iqbal, C. F. Choudhury, P. Wang, and M. C. González, “Development of Origin-Destination Matrices Using Mobile Phone Call Data: A Simulation Based Approach,” Transportation Research Part C: Emerging Technologies 40, Elsevier, 2014.
- [13] C. A. Segovia, “Origin-Destination Trips and Skims Matrices,” 2015. , s Interneta, http://www.fsutmsonline.net/images/uploads/fsutms_training/FSUTMS_Webinar_Series__Origin_and_Destination_Skims_and_Matrix_Information_20150917.pdf
- [14] OOCEA System Traffic Data & Statistical Manual. (2005.) Central Florida Expressway Authority. , s Interneta, <http://files.meetup.com/515268/orlando%20Highway%20traffic.pdf> (1.9.2019.).
- [15] J. Goulding, Best Practices and Methodology for OD Matrix Creation from CDR data, N/LAB, University of Nottingham, 2016.
- [16] H. Zhao and Y. Zhao, “Traffic analysis zones - how do we move forward?” AASHTO CTPP and TRB Joint Conference, Applying Census Data for Transportation, 2017.
- [17] J. Xie, Y. Nie, and X. Liu, “Testing the proportionality condition with taxi trajectory data,” Transportation Research Part B Methodological, May 2017.,

BIBLIOGRAFIJA

- scientific Figure on ResearchGate. , s Interneta, https://www.researchgate.net/figure/GIS-data-of-the-city-of-Shenzhen_fig3_316912924
- [18] prof. dr. sc. Ljupko Simunovic. (2019.) Prometno planiranje. , s Interneta, https://www.weboteka.net/fpz/Osnove%20prometnog%20in%C5%BEenjerstva/OPI_predavanje_6.-Lj_Simunovic.pdf (5.10.2019.).
- [19] D. Gundlegard, C. Rydergren, N. Breyer, and B. Rajna, "Travel demand estimation and network assignment based on cellular network data," COMPUTER COMMUNICATIONS, 2016.
- [20] P. Thakur and R. Grace. (2018., July) Rethinking peak hour with new transport models. KPMG Australia. , s Interneta, <https://home.kpmg/au/en/home/insights/2018/07/avoid-peak-hour-new-transport-models.html> 20.9.2019.
- [21] G. R. Jagadeesh, T.Srikanthan, and X. D. Zhang, "A Map Matching Method for GPS Based Real-Time Vehicle Location," 2004. , s Interneta, <https://www3.ntu.edu.sg/home/asgeorge/pdf/J04.pdf>
- [22] "Materijali kolegija usluge zasnovane na lokaciji," 2018.
- [23] G. Chen, S. Hoteit, A. C. Viana, M. Fiore, and C. Sarraute, "Individual Trajectory Reconstruction from Mobile Network Data," Tech. Rep., January 2019. , s Interneta, <https://hal.inria.fr/hal-01675570v2>
- [24] I. Stupar, P. Martinjak, V. Turk, and R. Filjar, "Socio-Economic Origin-Destination Matrix Derivation Through Contextualization of Material World," 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2018.
- [25] V. Frías-Martínez, E. Frías-Martínez, and C. S. Ruiz, "Estimation of urban commuting patterns using cellphone network data," 2012.
- [26] Geomatematicki rjecnik. Struna- Hrvatsko Strukovno Nazivlje. , s Interneta, <http://struna.ihjj.hr/naziv/voronoiov-dijagram/33840/> (20.1.2019.).
- [27] Podatci dostupni na:. , s Interneta, <https://www-users.cs.umn.edu/~tianhe/BIGDATA/>

BIBLIOGRAFIJA

- [28] F. Calabrese, G. D. Lorenzo, L. Liu, and C. Ratti, “Estimating Origin-Destination flows using opportunistically collected mobile phone location data from one million users in Boston Metropolitan Area,” IEEE Pervasive Computing 10, no. 4, pp. 36–44., April 2011., 2011.
- [29] C. M. Schneider, V. Beik, T. Couron  , Z. Smoreda, and M. C. Gonz  lez, “Unraveling daily human motifs,” J. R. Soc. Interface, no. 10, May 2013.
- [30] M. Fili  , R. Filjar, and K. Vidovi  , “Graphical Representation of Origin-Destination Matrix in R Statistical Environment,” 36. skup o prometnim sustavima s medjunarodnim sudjelovanjem AUTOMATIZACIJA U PROMETU, November 2016.
- [31] A. Travassoli, A. Alsg  r, M. Hickman, and M. Meshbah, “How close the models are to the reality? Comparison of Transit Origin-Destination Estimates with Automatic Fare Collection Data,” Australian Transport Research Forum, 2016.
- [32] W. Kuhlman, “The construct of purpose specific od matrices using public transport smart card data,” Tech. Rep., 2015.
- [33] T. Pollard, N. Taylor, T. van Vuren, and M. MacDonald, “Comparing the Quality of OD Matrices in Time and Between Data Sources,” European Transport Conference, 2013.
- [34] T. van Vuren, “256 shades of gray - comparing OD matrices using image quality assessment techniques,” Scottish Transport Applications Research (STAR), 2015.
- [35] T. Djuki  , “Reliability assessment of dynamic OD estimation methods based on structural similarity index,” 92nd meeting of the Transportation Research Board, 2013.
- [36] J. Varun, H. Sunil, R. Vasisht, S. Saragur, and A. Lele, “On the Analysis and Visualisation of Anonymised Call Detail Records,” 2013.
- [37] E. Barbour, C. C. Davila, S. Gupta, C. Reinhart, J. Kaur, and M. C. Gonz  lez, “Planning for sustainable cities by estimation building occupancy with mobile phones,” Nature Communications, 2019.

BIBLIOGRAFIJA

- [38] C. Krause and L. Zhang, “Short-term travel behaviour predictions with gps, land use and point of interest data,” Elsevier, 2018.
- [39] A. Degbelo and W. Kuhn, “Five general properties of resolution,” GIO 2014 - Workshop on Geographic Information Observatories 2014, 2014.
- [40] Spatial coverage. , s Interneta, <https://documentation.ands.org.au/display/DOC/Spatial+coverage> (4.10.2019.).
- [41] Global status report on road safety 2018. World Health Organisation. , s Interneta, https://www.who.int/violence_injury_prevention/road_safety_status/2018/Table_A9_Mobile_Phone_laws.pdf?ua=1 (2.10.2019.).
- [42] A. K. Smith, “An evaluation of high-resolution land cover and land use classification accuracy by thematic, spatial, and algorithm parameters,” Master’s thesis, University of Waterloo, 2017. , s Interneta, https://uwspace.uwaterloo.ca/bitstream/handle/10012/12506/Smith_Alexander.pdf?sequence=3&isAllowed=y
- [43] (03.06.2019) Teralytics. , s Interneta, <https://thenewstack.io/teralytics-takes-big-data-approach-human-movement/>
- [44] Granularity. , s Interneta, <http://wiki.gis.com/wiki/index.php/Granularity> (4.10.2019.).
- [45] E. Mellegard, “Obtaining origin-destination- matrices from cellular network data,” Master’s thesis, Chalmers University of Technology, Department of Mathematical Sciences, 2011.
- [46] E. Graells-Garrido and D. Saez-Trumper, “A Day of Your Days: Estimating Individual Daily Journeys Using Mobile Data to Understand Urban Flow,” 2016.
- [47] F. Bahoken and A.-M. Olteanu-Raimond, “Designing Origin-Destination Flow Matrices from Individual Mobile Phone Paths: The effect of spatiotemporal filtering on flow measurement,” in ICC 13 - 26th International Cartographic Conference, Aug 2013, DRESDEN, Germany. ICC 13 - 26th International Cartographic Conference 15p, 2013., hal-01011987v2.

BIBLIOGRAFIJA

- [48] S. Scepanovic, I. Mishovski, P. Hui, J. K. Nurminen, and A. Yla-Jaaski, “Mobile Phone Call Data as Regional Socio-Economic Proxy Indicator,” PLos ONE, 2015.
- [49] M. Coscia and R. Hausmann, “Evidence that Calls-based Mobility Networks are Isomorphic,” Working Papers - Cernter for International Development at Harvard University, 2015.
- [50] (21.06.2019) Opencellid. , s Interneta, <https://www.opencellid.org/>

Pojmovnik

ANPR Automatic Number Plate Recognition. 6

CDR Call Data Records ili Charging Data Records. ix, 13, 15, 16, 26, 29–31, 33, 35, 37, 38, 54, 56, 58, 59

FCD Floating Car Data. 12

FPD Floating Phone Data (Floating Cellular Data). 12

GIS Geoprostorni Informacijski Sustav. 28

GNSS Global Navigation Satellite System. 12

GPS Global Positioning System. 59

GSM Global System for Mobile (Communications). 12

HW Home-Work. 54

ITS Intelligentnim Transportnim Sustavima (engl. *Intelligent Transportation Systems*). 8, 12

MSSIM Više-razinski indeks strukturalne sličnosti (engl. Mean Structural Similarity index). 21, 22

ODM Origin-destination Martix. 1

POIs Točke Interesa (engl. Points of Interest). 28, 29

POM Polazišno-Odredišna Matrica. v–vii, ix, 1–26, 28–31, 33–35, 38, 41, 53–59

RSI Road Side Interview. 35

SSIM Indeks strukturalne sličnosti (engl. Structural Similarity Index). 21–23

TAZ Traffic Analysis Zone. viii, 7–9, 20, 55

WH Work-Home. 54

Sažetak

Ovo je tekst u kojem se opiše sažetak vašega rada. Tekst treba imati duh rekapitulacije što je prikazano u radu, nakon čega slijedi 3-5 ključnih riječi (zamijenite dolje postavljene općenite predloške riječi nekim suvislim vlastitim ključnim riječima).

Ključne riječi — Polazišno-odredišna matrica, parametri kvalitete, usporedba

Abstract

This is a text where a brief summary of your work is outlined. The text should have a sense of recap of what was presented in the thesis, followed by 3-5 keywords (replace the general keyword templates below with some meaningful keywords of your own) .

Keywords —Origin-Destination Matrix, quality parameters, keyword 3

Dodatak A

Postojeće metrike za vrednovanje Polazišno-Odredišnih Matrica

A.0.1 Metrike za procjenu sličnosti Polazišno-Odredišnih Matrica s referentnom

Za procjenu kvalitete POM-a dobivenih isključivo anketranjem u radu [5] korištena je mjera Srednja Apsolutna Postotna Pogreška (engl. Mean Absolute Percentage Error, MAPE), te je prikazano da se zadovoljavajuća razina kvalitete takvih POM-a postiže ako uzorak obuhvaća 50% populacije. Autor Cools ističe važnost korištenja drugih izvora uz ankete za izradu POM-a.

U radu [10] navedene su statističke mjere Relativna Pogreška (engl. Relative Error, RE), Devijacija Ukupne Potražnje (engl. Total Demand Deviation, TDD), Srednja Apsolutna Pogreška (engl. Mean Absolute Error, MAE), Korijen iz Srednje Kvadratne Pogreške (engl. Root Mean Square Error, RMSE) te Najveća Moguća Relativna Pogreška (engl. Maximum Possible Relative Error, MPRE) i Razina Prometne Potražnje (engl. Travel Demand Scale, TDS) koji procjenjuju kvalitetu neovisno o referentnoj matrici (no MPRE ne dopušta pogreške u prebrojavanju prometa, dok TDS ovisi o topologiji mreže i odabiru ruta).[35]

U [25] korišten je *Pearsonov koeficijent korelacije* - r da bi se utvrdila **sličnost svakog retka POM-e** dobivene iz CDR s **retkom referentne** (ukupni izlazni

tok iz svake polazišne ćelije). Isti postupak korišten je za kontekstualizirane Home-Work (HW) i Work-Home (WH) POM-e dobivene iz CDR u usporedbi s referentnim POM-ma dobivenim anketiranjem.

U svome radu [31] navodi se nekoliko uobičajeno korištenih mjera - R^2 , Geoffrey E. Havers statistika (GEH), Korijen iz Postotak Srednje Kvadratne Pogreške (engl. Root Mean Squared Error percentage, %RMSE) te uvodi novu mjeru Eigenvalue-Based Measure (EBM) (temeljenu na svojstvenim vrijednostima matrica) i procjenjuje pouzdanost POM-e dobivene iz sustava automatskog prikupljanja podataka u javnom prijevozu (autobus, vlak i trajekt). Spominje i Wasserstein Metriku, mjeru koja se razlikuje po tome da ne uspoređuje samo vrijednosti parova istih ćelija (elementwise).

Spearmanov Koeficijent Korelacije Ranga korišten je u [46] za procjenu sličnosti POM-a dobivenih iz CDR sa tada aktualnim POM-ma dobivenim anketiranjem.

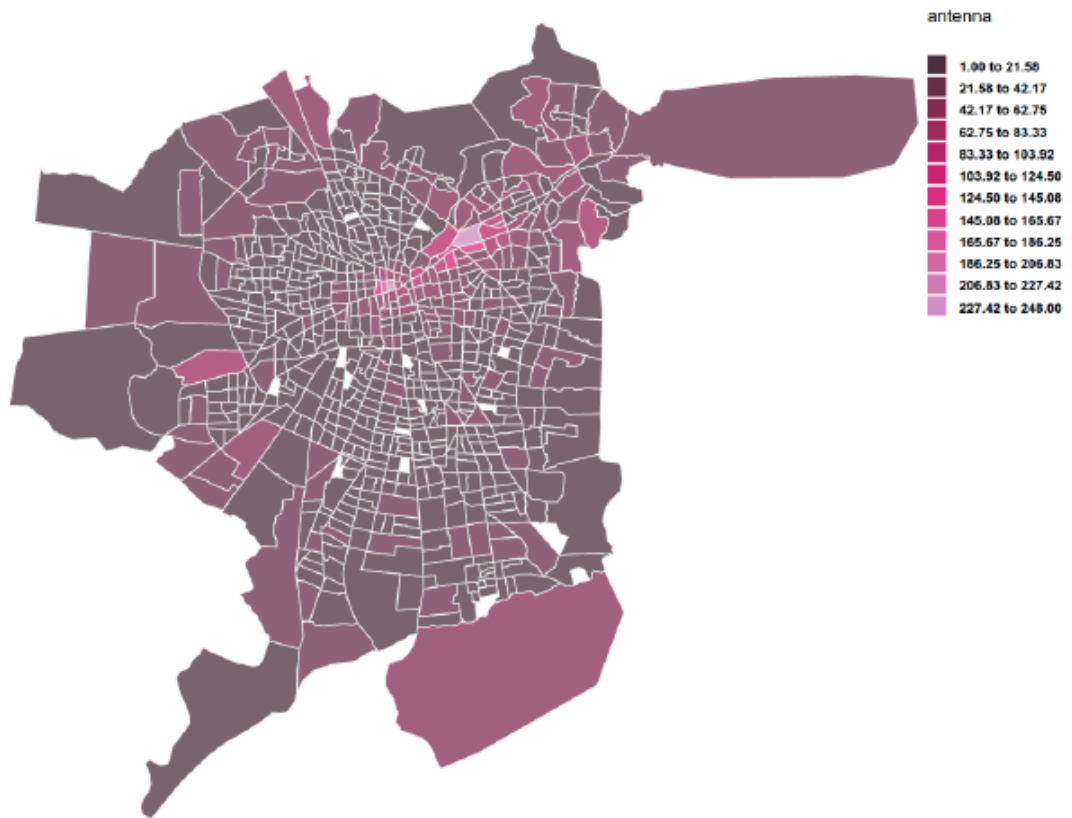
Dodatak B

Prilagodba prostorne podjele

Ako je točnost matrice definirana isključivo kao sličnost svakog njenog elementa s ekvivalentnim u elementom u *grand truth* matrici nužno je osigurati da prostorna podjela matrice odgovara prostornoj podjeli referentne ili je potrebno obje matrice svesti na zajedničku prostornu podjelu.

B.1 Agregacija baznih stanica

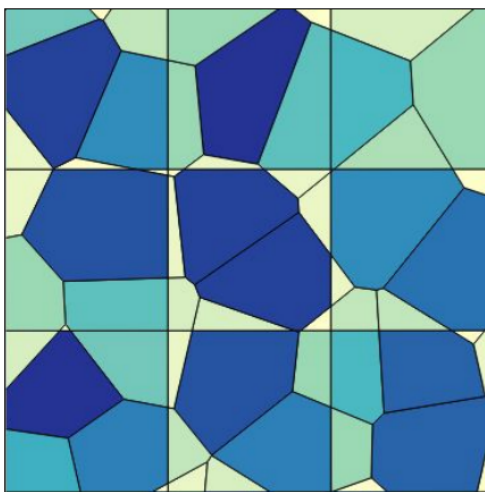
Pregledom literature utvrđeno je da se koriste *k-sredina* i *mean-shift* algoritam za grupiranje baznih stanica u broj grupa jednak broju TAZ na području o interesa [15] [8] [45] ili pak jedinicama samouprave (eng. *municipalities*) kako bi POM-e u konačnici bile usporedive s postojećima, dobivenim iz anketa [49] [25]. Nije utvrđivano koliko se greška uvodi ovim postupcima. Primjer gustoće tornjeva baznih stanica na područjima anketnih jedinica prikazan je na slici B.1



Slika B.1 Prostorna podjela na anketne jedinice u gradu Santiagu. Gustoća tornjeva baznih stanica po anketnim jedinicama. Gustoća varira od 1 do 250 tornjeva po anketnoj jedinici. [46]

B.2 Interpolacija

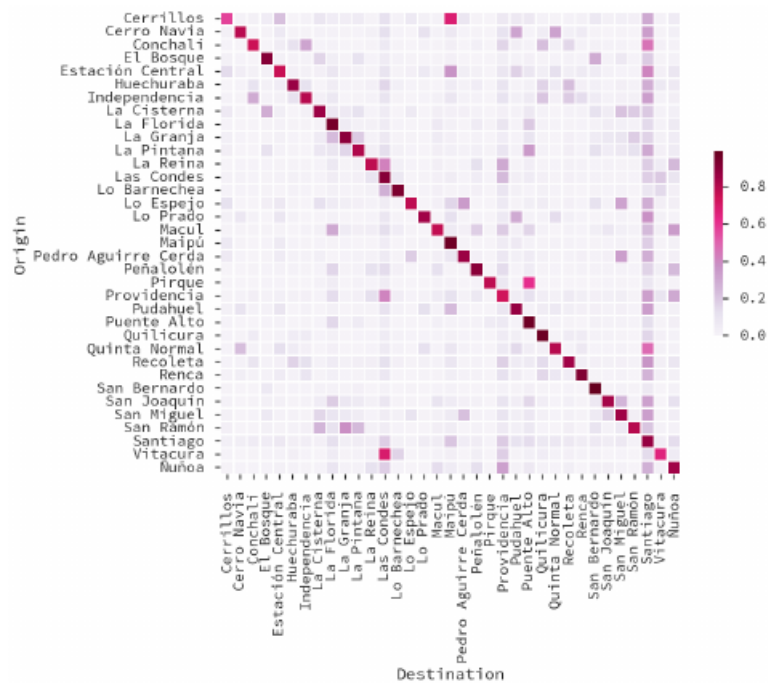
Postupak konverzije POM-a iz prostorne podjele na Voronoi ćelije u drugu prostornu podjelu, npr. uniformnu kvadratnu mrežu (ćelije $1km^2$), opisan u priručniku [15]. U postupku određivanja postotka ukupnog toka Voronoi ćelije koji će se dodijeliti novoj kvadratnoj ćeliji predlaže se uzeti u obzir: površinu preklapanja tih ćelija, broj zgrada ili ukupnu površinu zgrada (uključujući katove) na području preklapanja tih ćelija. Na slici B.2 je prikazan treći oblik interpolacije gdje su korišteni podaci o ukupnoj kvadratnoj površini zgrada.



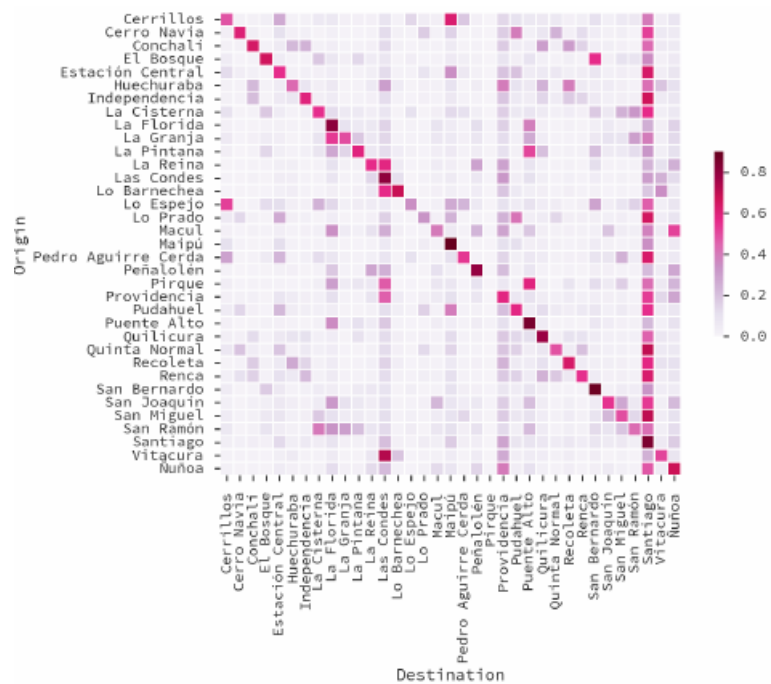
Slika B.2 Primjer mozaika *krhotina* za interpolaciju između Voronoi ćelija i kvadratne mreže. Nijansa odražava postotak površine zgrada u *krhotini* u odnosu na cijelu Voronoi ćeliju kojoj pripada. Rezultat se dobiva na osnovu tih *težinskih faktora* polazišne i odredišne *krhotine* te pripadajućih ("roditeljskih") Voronoi ćelija. Sumom svih rezultata *krhotina* jedne kvadratne ćelije izračuna se konačni tok. [15]

B.3 Posljedice agregacije prostornih ćelija

Voronoi ćelije također je moguće ugnijezditi (što se često radi kod usporedbe s POM-ma dobivenim drugim postupkom), a pritom se može u matricu bilježiti i aproksimacija internog toka, odnosno broja putovanja unutar nove ćelije (dijagonala matrice). Takva putovanja nazivaju se internim putovanjima. Studija [6] je pokazala da sa **smanjenjem rezolucije (visokim stupnjem agregacije Voronoi ćelija) stupanj korelacije s POM-ma iz ankete (za isti grad) raste** (POM-e dobivene iz CDR i iz anketa u 4 različita grada). Postotak eksternih putovanja koji se gubi u procesu agregacije Voronoi ćelija analiziran je u studijama [46] (Vidi sliku B.3) i [47] za POM-e iz regije Picardie u Francuskoj. **Agregacijom Voronoi ćelija na razinu *Urban Areas* (područja oko gradova) 85% svih putovanja postaje internim putovanjima, a na razini *Urban Cores* (područja oko većih gradova) čak 97% početno zabilježenih putovanja je interno, te ostaje samo 3% eksternih putovanja** [47].



(a) Izvor podataka anketa



(b) Izvor podataka CDR

Slika B.3 Distribucija CDR putovanja na području Santiago, Čile. POM je normalizirana prema redovima (polazištima) (*) [46]

Dodatak C

OpenCellID

Operatori u CDR zapisima ne otkrivaju položaje baznih stanica već samo njihov ID. Postoji nekoliko javno dostupnih baza kao što je OpenCellID [50] nastalih volonter-skim doprinosom koje pohranjuju podatke o procjenama položaja baznih stanica, snazi signala i sl. Položaji se prikupljaju pomoću aplikacije instalirane na mobilnom uređaju, registriranjem Global Positioning System (GPS) položaja i identifikacijske oznake bazne stanice na koju je uređaj trenutno spojen. Dakle, ove baze ne sadrže "najtočnije" položaje baznih stanica već usrednjene položaje mobilnih uređaja koji su u nekom trenutku bili spojeni na tu baznu stanicu. Tako procijenjeni položaj za potrebe izrada POM-a čak i više odgovara od stvarnog položaja. [45]

Dodatak D

OpenStreetMap

OpenStreetMap (OSM) je svjetski raširen projekt koji kreira i pruža slobodne zemljopisne podatke (zemljovide gradova i naselja) temeljen na volonterskom doprinosu zajednice. Projekt pruža detaljne i ažurne digitalne zemljovide kompatibilne s Geoprostorni Informacijski Sustav (GIS) aplikacijama. Započet je prije 15 godina, u Ujedinjenom Kraljevstvu, kao odgovor na tehnička ili pravna ograničenja postojećih "slobodnih" baza prostornih podataka kao što je GoogleMaps. Zajednički doprinos i međusobna kontrola unosa održavaju kartu na visokoj razini kvalitete i točnosti.

Struktura osnovnih elemenata na toj rasterskoj karti je hijerarhijska. Element u hijerarhiji može biti: čvor (engl. *node*), put (engl. *way*) ili relacija (engl. *relation*). Uz hijerarhijske postoji i opisni element koji se naziva oznakom (engl. *tag*), a njegova funkcija je opisati značajke hijerarhijskog elementa uz koji je vezan.

Čvor je jedinstvena točka u prostoru (sa identifikacijskom oznakom, zemljopisnom širinom i dužinom) koja najčešće predstavlja fizički objekt (zgrada, dio ceste...) te sadrži jedan ili više *ključ=vrijednost* oznaka (engl. *key=value tag*) koji definiraju razne značajke objekta.

Put je uređena lista 2 do 2,000 čvorova, koja također može biti opisana *ključ=vrijednost* oznakama.

Relacija je uređena lista čvorova, puteva i/ili relacija. Definira logičku ili geografsku povezanost članova.

Primjeri oznaka vezanih za čvorove: *office=company*, *building=residential*, *buil-*

ding=hotel, building=church, leisure=sports_centre , amenity=school. Primjer žiga vezanog uz put *highway=residential.*

OSM doprinosi istraživanjima u području analize kretanja stanovništva i kao izvor prometne infrastrukture i pripadajućih meta podataka (kolnik, mostovi, ograničenja brzine i dr.)[6] i kao izvor podataka o podnoj površini objekata (engl. floor area) [15].