

SVEUČILIŠTE U RIJECI  
**TEHNIČKI FAKULTET**  
Diplomski studij računarstva

Diplomski rad

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

Rijeka, studeni 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, studeni 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, studeni 2019.

-----  
Ime Prezime

# Sadržaj

Popis slika	viii
Popis tablica	x
<b>1 Uvod</b>	<b>1</b>
<b>2 Polazišno-Odredišna Matrica</b>	<b>3</b>
2.1 Definicija Polazišno odredišne matrice . . . . .	3
2.1.1 <i>Traffic Analysis Zone</i> . . . . .	7
2.1.2 Model prometa . . . . .	7
2.1.3 Vremenski nezavisno i vremenski zavisno modeliranje . . . . .	9
2.2 Suvremeni pristupi procjeni	
Polazišno-Odredišnih Matrica . . . . .	10
2.2.1 Pokretna osjetila . . . . .	11
2.2.2 Polazišno-Odredišna Matrica iz zapisa o telekomunikacijskim aktivnostima u javnoj pokretnoj mreži . . . . .	12
2.2.3 Diskretizacija Vremena . . . . .	16
<b>3 Pregled postupaka i istraživanja o vrednovanju Polazišno-Odredišnih Matrica</b>	<b>18</b>
3.0.1 Linearna regresija - mjera $r^2$ . . . . .	19

## Sadržaj

3.0.2	Vrednovanje suvremenih Polazišno-Odredišnih Matrica . . . .	20
3.0.3	Grafički oblik i sličnost strukture Polazišno-Odredišnih Matrica . . . . .	21
<b>4</b>	<b>&lt; Naslov &gt;</b>	<b>25</b>
4.1	Uvodna razmatranja i pregled istraživanja . . . . .	25
4.2	Kontekstualizacija Polazišno-Odredišnih Matrica . . . . .	28
4.2.1	Primjeri korištenja konteksta u analizi kretanja . . . . .	28
4.3	Teza . . . . .	30
4.3.1	Alternativni pristup vrednovanju Polazišno-Odredišnih Matrica	31
4.4	Odnosni parametri kvalitete . . . . .	31
4.4.1	Prostorno obuhvaćanje . . . . .	31
4.4.2	Prostorna razlučivost (rezolucija) . . . . .	31
4.4.3	Zrnatost . . . . .	33
4.4.4	Vremenska razlučivost (rezolucija) . . . . .	33
4.4.5	Gustoća informacija . . . . .	34
4.4.6	Putovanje kao parametar . . . . .	34
4.4.7	Širina toka . . . . .	36
<b>5</b>	<b>Metodologija</b>	<b>42</b>
5.1	Uvjeti usporedbe . . . . .	42
5.2	Brzina kretanja . . . . .	44
<b>6</b>	<b>Rezultati</b>	<b>45</b>
<b>7</b>	<b>Diskusija</b>	<b>46</b>
<b>8</b>	<b>Zaključak</b>	<b>47</b>

## *Sadržaj*

<b>Bibliografija</b>	<b>48</b>
<b>Pojmovnik</b>	<b>54</b>
<b>Sažetak</b>	<b>55</b>
<b>A Postojeće metrike za vrednovanje Polazišno-Odredišnih Matrica</b>	<b>56</b>
A.0.1 Metrike za procjenu sličnosti Polazišno-Odredišnih Matrica s referentnom . . . . .	56
<b>B Prilagodba prostorne podjele</b>	<b>58</b>
B.1 Agregacija baznih stanica . . . . .	58
B.2 Interpolacija . . . . .	58
B.3 Posljedice agregacije prostornih ćelija . . . . .	59
<b>C OpenCellID</b>	<b>62</b>
<b>D OpenStreetMap</b>	<b>63</b>

# Popis slika

2.1	Polazišno-Odredišna Matrica. Prema: [1] . . . . .	4
2.2	Usmjereni graf kretanja i pripadajuća Polazišno-Odredišna Matrica. (Zone su označene slovima abecede). Prema: [2] . . . . .	4
2.3	Prikaz prosječnog volumena na vezama (prometnicama) sa stanicama za prebrojavanje radnim danom 2005. Orlando, Florida. [3] . . . . .	6
2.4	Mreža za analizu prometa sastoji se od 221,14 cesta - plave linije i 3,199 TAZ zona - granice označene iscertanom crnom linijom [4] . . . .	8
2.5	Voronoi dijagram 1090 ćelija baznih stanica - kineski grad Shenzhen [5] . . . . .	15
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. $a = 0.0071273$ , $b = 0.631128$ , $r^2 = 0.4026$ . . . . .	20
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 [6] . . . . .	23
4.1	Širina vrhunca prometne potražnje - vremenski zavisno modeliranje [7]	27
4.2	Distribucija CDR putovanja na području Santiaga, Čile. Polazišno- Odredišna Matrica (POM) je normalizirana prema redovima (po- lazištima) (*) [8] . . . . .	38



4.3	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. [8] . . . . .	39
4.4	Gustoća tornjeva baznih stanica po kvadratnom kilometru za pojedine države i grad-državu Singapur. [9] . . . . .	40
4.5	Distribucija udaljenosti putovanja u kilometrima, Senegal 2016. [10]	40
4.6	Dvije POM-e - produkti različitih definicija putovanja. Dobivene su iz potpuno istog izvora CDR zapisa. [10] . . . . .	41
5.1	Broj putovanja javnim prijevozom zabilježenih <i>OV-chipkaart</i> pametnim karticama u Nizozemskoj izražen u milijunima po tjednima u godini. Zelena linija označava prosječnu vrijednost radnog tjedna [34]	43
B.1	Primjer mozaika <i>krhotina</i> koje dozvoljavaju interpolaciju između Voronoi ćelija i kvadratne mreže. Nijansa odražava postotak površine zgrada u <i>krhotini</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") Voronoi ćelija. Sumom svih rezultata <i>krhotina</i> jedne kvadratne ćelije izračuna se konačni tok. [11]	59
B.2	Distribucija CDR putovanja na području Santiaga, Čile. POM je normalizirana prema redovima (polazištima) (*) [8] . . . . .	61

# Popis tablica

2.1	Primjeri zapisa o telekomunikacijskim aktivnostima (pozivi) [12]	. . . 13
2.2	Primjeri zapisa o telekomunikacijskim aktivnostima [13]. 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)	. . . . . 13

# Poglavlje 1

## Uvod

Polazišno-Odredišna Matrica *eng. Origin-destination Matrix (ODM)* ili *Trip Table* 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.

## *Poglavlje 1. Uvod*

Suvremeni izvori podataka o položaju i načini procjene POM-a koji uključuju kontekstualizaciju 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

### 2.1 Definicija Polazišno odredišne matrice

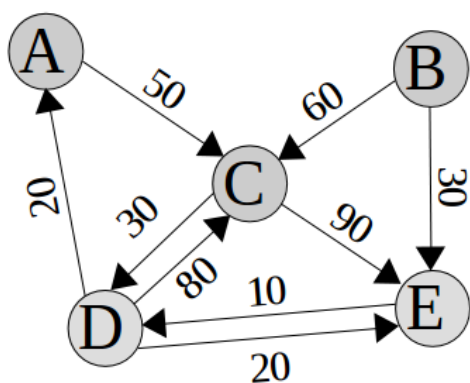
**Definicija 2.1.1.** Polazišno-odredišna matrica je matrica čiji svaki element predstavlja broj putovanja između odgovarajućeg para zona unutar promatranog područja u promatranom vremenskom razdoblju  $\tau$ , u svakom smjeru zasebno. Redovi u POM-i  $i = 1, 2, 3 \dots n$  predstavljaju polazišta (izvore) 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)

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 sjevera prema jugu. [?]

Poglavlje 2. Polazišno-Odredišna Matrica

Polazišta (Origins)	Odredišta (Destinations)							$\sum_{j=1}^n f_j = O_i$
	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	$f_{n1}$	$f_{n2}$	$f_{n3}$	...	$f_{nj}$	...	$f_{nn}$	$O_n$
$\sum_{i=1}^n f_i = D_i$	$D_1$	$D_2$	$D_3$	...	$D_j$	...	$D_n$	$\sum_{i=1}^n \sum_{j=1}^n f_{ij} = F$

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



odredišta polazišta	odredišta					
	A	B	C	D	E	
A	0	0	50	0	0	50
B	0	0	60	0	30	90
C	0	0	0	30	0	30
D	20	0	80	0	20	120
E	0	0	90	10	0	100
	20	0	280	40	50	390

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

### 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 [14], već se ona određuju matematički modeliranjem uz uvažavanje pretpostavki** (odlomak 2.1.2). 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. [15] 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 [15] [16]. 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 [17].

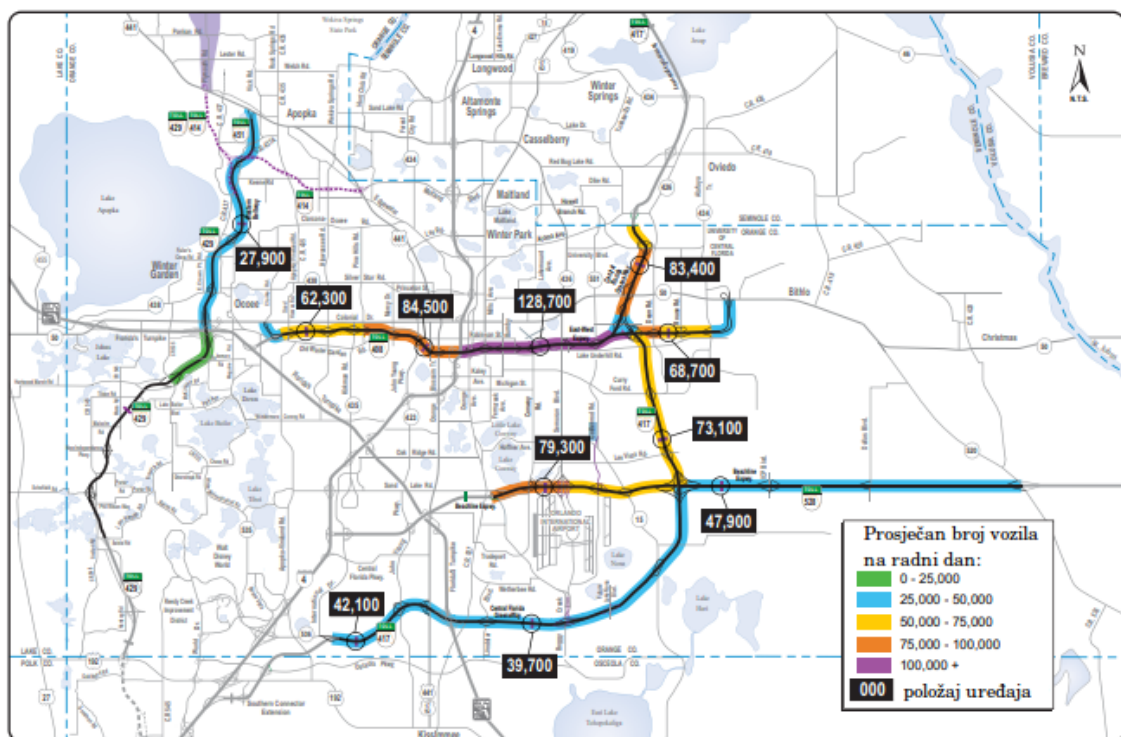
### Raspodjela stanica za prebrojavanje

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 [15]. 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

## Poglavlje 2. Polazišno-Odredišna Matrica

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 [15]. 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 [18].** Na slici 2.3 prikazana je karta grada Orlando s istaknutim važnim prometnicama (sivom) i s prosječnim prometnim tokovima zabilježenim uređajima za prebrojavanje vozila. [3]



Slika 2.3 Prikaz prosječnog volumena na vezama (prometnicama) sa stanicama za prebrojavanje radnim danom 2005. Orlando, Florida. [3]



### 2.1.1 *Traffic Analysis Zone*

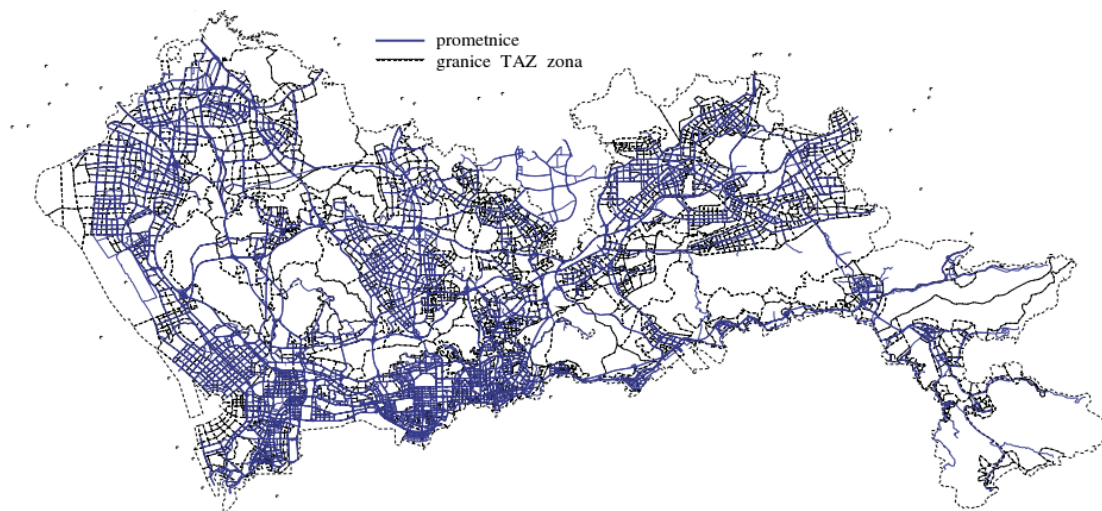
Traffic Analysis Zone (TAZ) je strani naziv za jedinicu prostorne podjele (zonu) kod analize prometa. Površina tih jedinica varira i nije uniformna. Podjela se može temeljiti na populaciji ili broju zaposlenih na nekom prostoru, primjerice svaka zona pokriva prostor na kojem obitava otprilike tri tisuće ljudi, ili zone generiraju podjednak broj izlaznih putovanja (jednaka širina izlaznih tokova) [11].

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.[19]

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.

### 2.1.2 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 strateških planiranja prometne infrastrukture i u Inteligentnim Transportnim Sustavima (engl. *Intelligent Transportation Systems*) (ITS) koji su zauzvrat ključni za



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 [4]

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 [20], da svi korisnici imaju saznanja o situaciji na cesti [11] te da je mobilnost općenito manja tijekom zime nego tijekom ljeta [14] i da je broj putovanja po kućanstvu proporcionalan dohotku [21]. 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)* [22]. Izvedene verzije integriraju i neke druge (socio-ekonomske) faktore [10].

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 [18]:

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*. [18]

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. [18]

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).

### 2.1.3 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.** [18]

#### 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  $\tau$  od 15 minuta [15][21].

Neki *Peak-hour* vremenski modeli koji opisuju potražnju za vrijeme dnevnog vrhunca prometne potražnje, procjenjuje potražnju za svaku minutu ( $\tau = 1$  min). [7]. 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. [15]

## 2.2 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 pojedinaca 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 prikupljeni za analizu mobilnosti stanovništva već za pružanje usluge zasnovane na lokaciju. **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.

### 2.2.1 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. [23]

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. [23]

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 izjednačavanjem 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 rang 0-15 m. [24][25]

### 2.2.2 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, [14]), te podaci kada je pokretna stanica uključena u mrežu i isključena iz mreže i opterećenje pojedinog tornja [12].

## Poglavlje 2. Polazišno-Odredišna Matrica

Tablica 2.1 Primjeri zapisa o telekomunikacijskim aktivnostima (pozivi) [12]

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 [13]. 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

Novi postupci koriste upravo anonimizirane, često višemjesečne CDR, kao izvor podataka za procjenu POM-a [26][10][27]. Metodologija korištenja odnosnih elemenata CDR-a za procjenu POM predložena je još 1994. godine [10].

**Definicija 2.2.1.** Skup minimalno  $n$  događaja (telekomunikacijskih aktivnosti) u nizu, istog korisnika, koji su zabilježeni na istom položaju unutar definiranog minimalnog 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 2.2.2.** 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

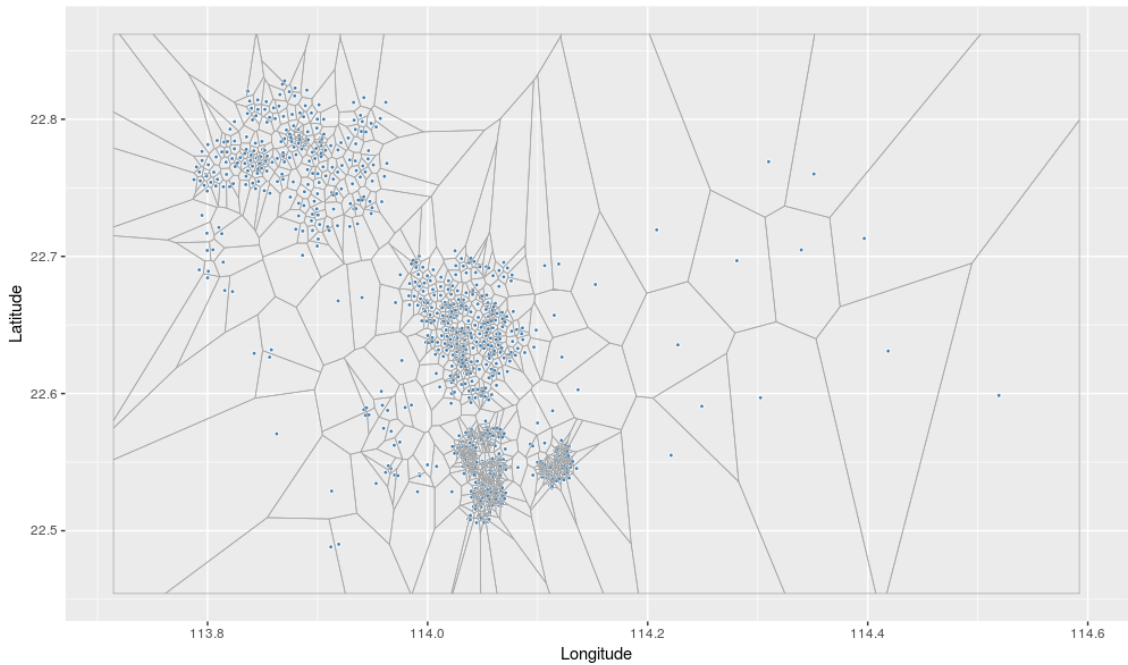
*Najfinija* prostorna rezolucija koju je moguće postići kada je riječ o POM-i dobivenoj iz CDR postiže se podjelom prema Voronoiovoj teselaciji.

**Definicija 2.2.3.** 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 ikojoj 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. [28]

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 [14].

**Gustoća Voronoi ćelija dobro je korelirana s gustoćom stanovništva te stoga bolju prostornu rezoluciju ima na područjima veće gustoće popu-**





Slika 2.5 Voronoi dijagram 1090 ćelija baznih stanica - kineski grad Shenzhen [5]

lacije [10]. 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 [29].** Problem predstavlja preklapanje područja pokrivanja u stvarnosti [10]. 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 [10] ili agregacijom područja pokrivanja baznih stanica čija je udaljenost manja od 75 m (s dodjelom događaja najaktivnijem tornju). [11]**

Analiza snage signala i trenutnih opterećenja okolnih baznih stanica mogu dati jasniji uvid u opterećenje mreže u vremenu početka (i trajanja) telekomunikacijske aktivnosti. [12]

### 2.2.3 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 [12]. U literaturi je korištena različita vremenska rezolucija (razlučivost) - podjela na vremenske periode  $\tau$  od 15 min, 30 min [30], 1 sat, 2 sata, 3 sata [31][12]. 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* [30][32][10] 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 2.2.4.** Potpunost putanje definirana je izrazom

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

gdje je  $n_L$  broj zabilježenih položaja (zaustavljanja),  $\tau$  vremenski period, a  $\mathcal{T}$  duljina ukupnog vremenskog razdoblja za koje su dostupni zapisi. Idealna potpunost je postignuta ako postoji barem jedan zabilježeni položaj (zaustavljanje) po periodu. [12]

Veća potpunost putanja postiže se ako se osim poziva kao komunikacijskih aktivnosti u obzir uzmu SMS poruke i prijenos mobilnih podataka [12]. **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 [12].**

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

## *Poglavlje 2. Polazišno-Odredišna Matrica*

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. [12]

## 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** [33]. 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 2.1.2) u postupku procjene nove POM-e iz rezultata prebrojavanja vozila [18]. 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 matičnu 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  $\epsilon$  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). [34] (vidi sliku 3.1)

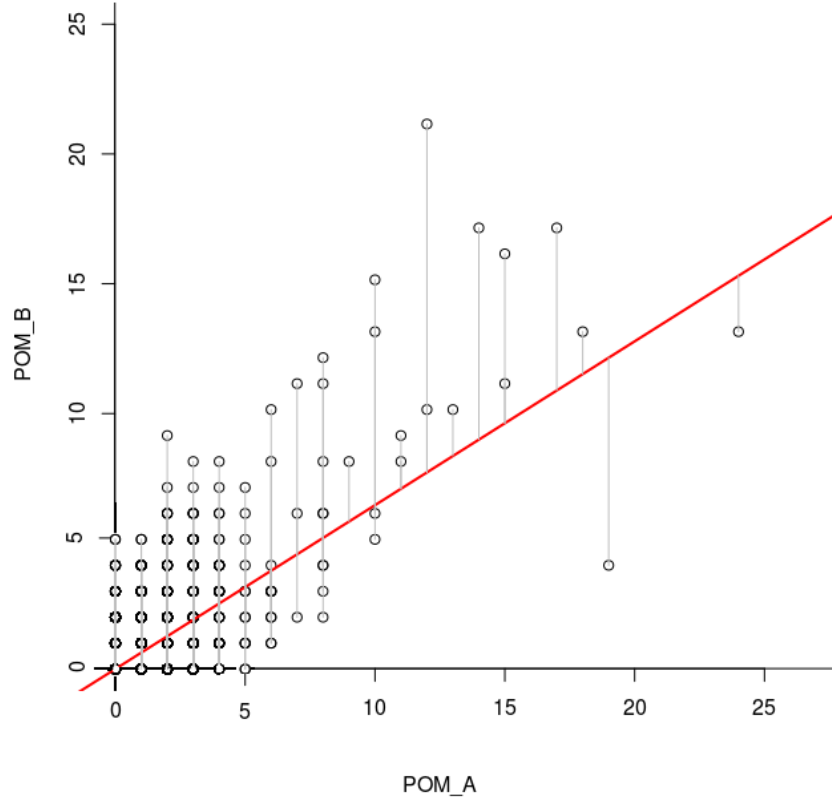
Korelacijski koeficijent određenja ili obuhvaćanja varijance  $r^2$  je mjera koja pokazuje koliko je slaganje između vrijednosti izračunatih modelom procjene i izmjerenih vrijednosti (*ground truth*). Općenito, to je mjera dobrote pristajanja (engl. *goodness of fit*) koja govori koliko uspješno model 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 [34]. Kod usporedbe 2 POM-e u odnosu na referentnu *ground truth*, ona koja je sličnija referentnoj bolje opisuje kretanje. Ova mjera ne smatra se praktičnom za usporedbu POM-a dobivenih

različitim pristupima procijeni [34] što dolazi kod izražaja i kada se radi o POM-i iz suvremenih izvora.



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 [35] 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 [31], prikazom nijansama jedne boje 0-255 engl. *grayscale* [36], 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. [6]

Informacija o strukturi slike definira se kao atributi slike koji predstavljaju struk-

turu 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  $\mu_x$  i  $\mu_y$  sa standardnim devijacijama  $\sigma_x$  i  $\sigma_y$  i kovarijancom  $\sigma_{xy}$

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

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

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

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  $\alpha$ ,  $\beta$  i  $\gamma$  parametri relativne važnosti svake komponente. Za  $SSIM$  vrijedi slijedeće:

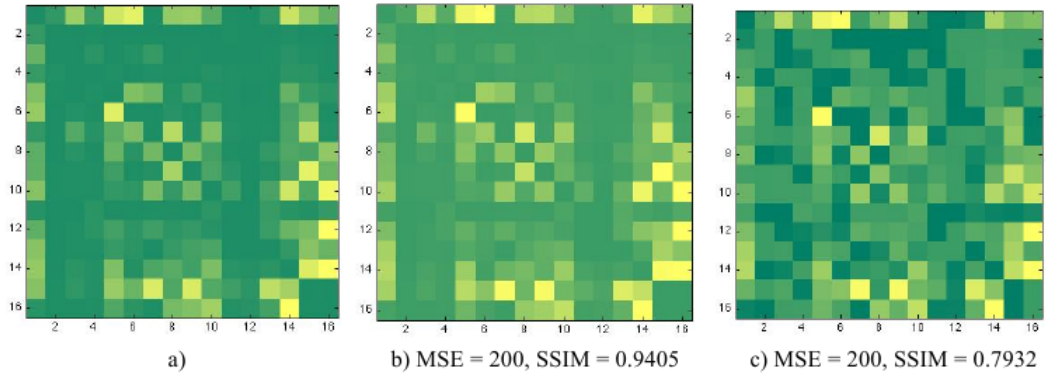
$$SSIM(x, y) \leq 1 \quad (3.7a)$$

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

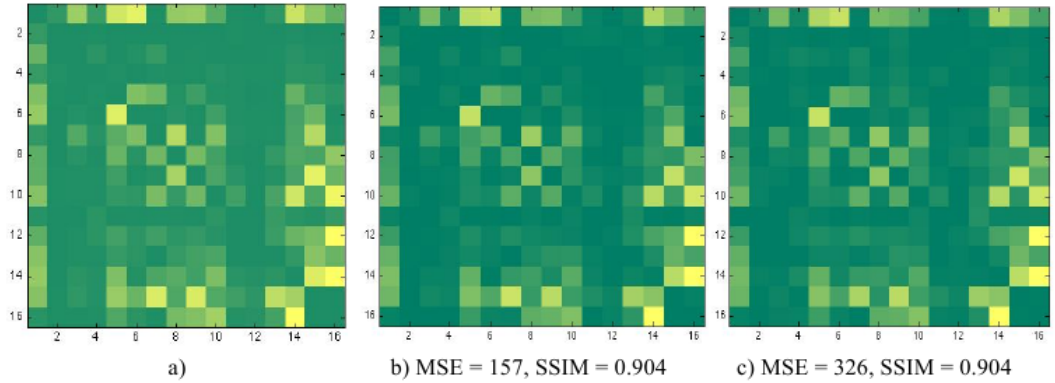
$$SSIM(x, y) = 1 \iff x = y \quad (3.7c)$$



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



- (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 [6]

### **Poboljšani višerazinski indeks strukturalne sličnosti**

Nekoliko godina nakon prvog spominjanja *MSSIM* kao metrike usporedbe POM-a [36] 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. [35][37][36]

# 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.

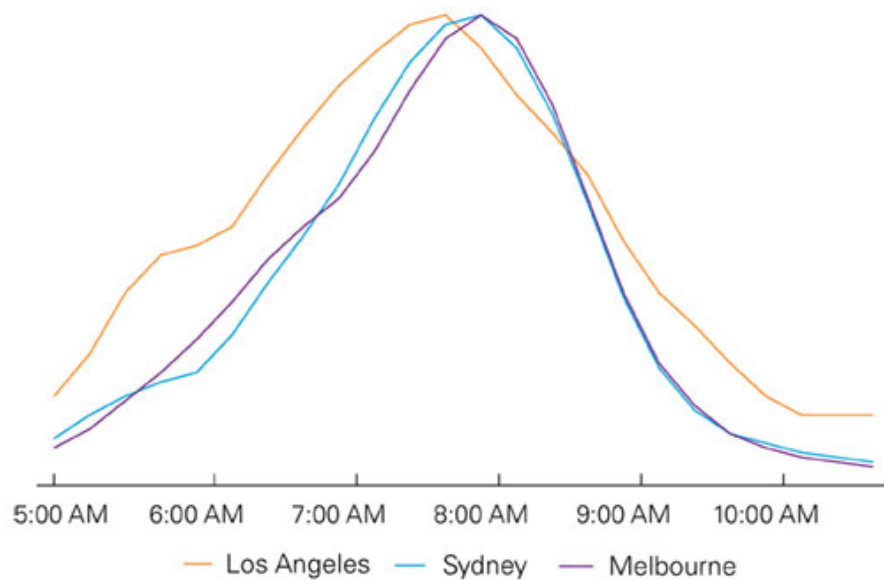
Analiza prometne potražnje tijekom dnevnih vrhunaca i pokazatelja kako će se taj vrhunac mijenjati u bliskoj i dalekoj budućnosti zahtjeva odabir dobre vremenske rezolucije. Tradicionalni modeli procjene pretpostavljaju da se širina vremenskog vrhunca u budućnosti neće mijenjati iako se ona već sad razlikuje od grada do grada (Slike 4.1) [7].

**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 *grand 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-

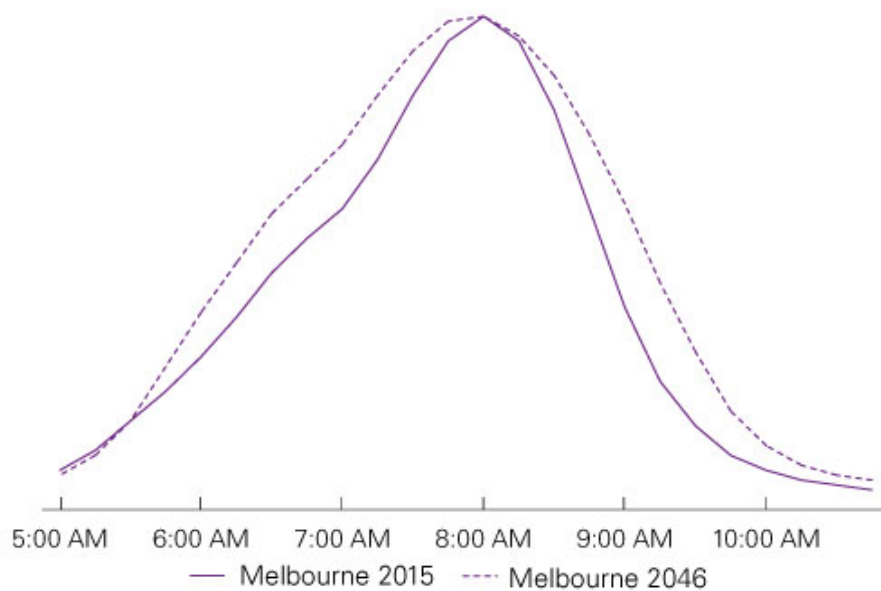
litete.

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 [14]. **Praksa odbacivanja ogromnog dijela uzorka zbog niske potpunosti putanja može se zamijeniti rekonstrukcijom putanja uvažavajući njihov kontekst [12].**

Analiza kretanja stanovništva koja uzima u obzir i kontekst proširuje područje primjene zaključaka analize kretanja stanovništva van prometa [38]. 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 vrhunca prometne potražnje - vremenski zavisno modeliranje [7]

## 4.2 Kontekstualizacija Polazišno-Odredišnih Matrica

**Definicija 4.2.1.** Putovanje kao informacijski objekt je

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

**Određivanje konteksta putovanja proširuje područje pokrivanja u fizičkom i informacijskom smislu te povećava točnost POM-e putem unosa relevantnih atributa vezanih za način i svrhu putovanja, čime je putovanje, kao informacijski objekt, točnije opisano.** 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. [27] 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 [13]. Prostorni objekti na području interesa kategorizirani su na osnovu pripadajućih opisnih podataka 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 ćeliji (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 [31][39]. 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 [38]

#### **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. [40]

#### **Studija E - Rekonstrukcija rijetkih trajektorija**

[12]

### **4.3 Teza**

Poboljšanje kvalitete procjene omogućeno je preslikavanjem POM-e na kontekst. Određivanje konteksta proširuje područje pokrivanja u fizičkom i informacijskom



smislu te povećava točnost POM-e unosom relevantnih atributa vezanih za način i svrhu putovanja - čime je putovanje točnije opisano.

### 4.3.1 Alternativni pristup vrednovanju Polazišno-Odredišnih Matrica

Vrednovanje POM-e je proces koji ima za cilj procjenu obilježja kvalitete te POM-e i znači identifikaciju obilježja kvalitete POM-e, njenog sadržaja, njene smislenosti, a izražava se slijedećim indikatorima: ...

Točnost, rezolucija i obujam podataka su kritični parametri geografske informacije i znanje o čemu oni zavise je nužna baza za veće pothvate procjene vrijednosti geografskih informacija kao društvenog dobra. [41]

ažurnost?!

## 4.4 Odnosni parametri kvalitete

### 4.4.1 Prostorno obuhvaćanje

**Definicija 4.4.1.** Prostorno obuhvaćanje je geografsko područje gdje su prikupljeni podatci, mjesto koje je predmet opažanja ili lokacija u fokusu aktivnosti. [42]

1. prometna infrastruktura (ceste, željeznica, javni prijevoz)
2. neovisno o prometnoj infrastrukturi (CDR, GNSS)

### 4.4.2 Prostorna razlučivost (rezolucija)

Degbelo u radu "Pet općih svojstava rezolucije" [41] kao jedno od (uvijek točnih) svojstava navodi da je **rezolucija (geografskih podataka) pozitivno korelirana s točnošću**. Razmatra suprotnu tvrdnju - da je *za očekivati* da se točnost povećava sa smanjenjem rezolucije. Njegova interpretacija te tvrdnje je da je točnost veća, ako je zahtjev za specifičnosti manji [41].

Taj drugi pristup točnosti zastupljen je u evaluaciji POM [34].

Točnost matrica u odnosu na referentnu provjerava se na 3 različite razine granularnosti (broja zona). Rezultat je veća točnost/sličnost iskazana mjerom  $r^2$  kod usporedba nižeg stupnja granularnosti (niže rezolucije u smislu broja zona). Odabirom nižeg stupnja granularnosti zapravo se smanjuje zahtjev za specifičností.

Zaključak: Povećanje točnosti/sličnosti matrica iskazane mjerom  $r^2$  koje je postignuto uzimanjem nižeg stupnja granularnosti posljedica je smanjenja zahtjeva za specifičností.

**Definicija 4.4.2.** Prostorna razlučivost je najmanja pređena udaljenost koju je moguće identificirati.

Rezolucija POM-e procijenjene iz CDR ovisit će o gustoći tornjeva bazni stanica. Gustoća tornjeva dobro je korelirana s gustoćom populacije [10]. Distribucija tornjeva nije homogena, niti na području grada, niti na bilo kojem drugom području.

## Kontekstualizacija

Objekt u ćeliji je najmanja prostorno razlučiva jedinica

Ispitati:

- podijeliti u kategoričke vrijednosti:
- 0.2, 0.5, 1 km - što je previše, što je idealno, a što je premalo i opet ne valja (prema literaturi ne valja manje od 75m, a idealno je 1 km)
- Na temelju čega grupirati BS- klasteri? Na osnovu blizine? Na osnovu broja objekata u ćeliji kod određivanja konteksta? Aglomerative clustering bez zadanog broja k?
- Na temelju čega u konačnici zaključiti da premalo ne valja, gubi smisao?
- Brzina kretanja

### 4.4.3 Zrnatost

Rezolucija se odnosi na količinu detalja u reprezentaciji (stvarnosti) dok se zrnatost odnosi na spoznajni aspekt uključen pri odabiru značajki [41]. Pojam zrnate podjele (engl. granular partition) uključuje ne samo karte (podatke) već kategorizaciju koja nadilazi mjerenja i obzervaciju i zalazi u područje konceptualnog modeliranja [41]. Za razliku od rezolucije prema [41], zrnatost i točnost su neovisne. Odgovori na pitanje "Gdje se nalazite?" "U Hrvatskoj", "U Rijeci", "U svojoj kući" i "Za svojim stolom" svi točni u smislu da 'govore istinu'.

**Definicija 4.4.3.** 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. [43]

### 4.4.4 Vremenska razlučivost (rezolucija)

**Definicija 4.4.4.** Vremenska razlučivost je najmanji vremenski odmak koji je moguće identificirati.

Najmanji spominjani period u literaturi je 15 min [15], drugi najmanji 30 min [30], kod POM-a kreiranih iz podataka generiranih pametnim karticama u javnom prijevozu također 15 min [33], a zastupljena je i dinamičkoj podjela na kratke intervale za najprometnija razdoblja u danu: jutarnji vrhunac (engl. *morning peak*), poslijepodnevni vrhunac (engl. *afternoon peak*) i duže intervale za ostala razdoblja.

Ispitati:

periode od

- 15 min, 1h, 3h, 8h → pokazati da 8 gubi smisao, preveliki periodi da gube svrhu

-kako to pokazati?

### Kontekstualizacija

-Kontekst vezan za vrijeme u kojem se odvija putovanje

#### 4.4.5 Gustoća informacija

(...)

-Informacija na razini putovanja

-Informacija na razini toka (postotka toka) - metoda koja se oslanja na vanjske izvore konteksta može se primijeniti na matrice iz različitih izvora podataka.

- nemam implementaciju funkcija s kampa za kontekstualizaciju objekata i tokova, imam za izvlačenje objekata iz voronoi ćelija

Što je gušće opisan prostor u ćeliji to je moguće ostvariti bolju rezoluciju

Ispitati:

#### 4.4.6 Putovanje kao parametar

Putovanja koja se protežu kroz više perioda dodjeljuju se vremenskom okviru u kojem započinju ili u kojem završavaju. Autori [15] [31] [10] sva putovanja dodjeljuju intervalu u kojem je putovanje započeto ili procjenjuje obje vrste POM-a [44].

Putovanja kod vremenski zavisnih POM-a kreiranih iz CDR moguće je svrstati u nekoliko kategorija na osnovu prihvatanja ili odbacivanja određenih kretanja kao putovanja.

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
4. putovanja ograničena minimalnom i/ili maksimalnom udaljenosti [8]
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 maksi-

malnim vremenskim razmakom između dva uzastopna događaja

**Iskazivanje putovanja kao parametra ne smije obuhvatiti lažno kretanje uzrokovano uravnoteženjem opterećenja i lažnog zaustavljanja uzrokovnog neuniformnom gustoćom zapisa pojedinaca u vremenu, što se postiže postavljanjem navedenih ograničenja.**

Za 2 različite definicije putovanja, za POM-e iz identičnih podataka na području Senegala generirano je 0.7 putovanja dnevno po korisniku u jednom, a 0.06 u drugom. Ograničenja su bila: 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. [10] Distribuciju možemo vidjeti na slici 4.5

**O definiciji putovanja ovisit će njihov konačan broj, odnosno širina toka.**

Ispitati:

- koliki se postotak putovanja gubi izbacivanjem putovanja u susjedne ćelije?!  
→ ovisit će o veličini ćelija
- koliki se postotak putovanja gubi izbacivanjem putovanja koja prelaze iz perioda u period?! → ovisit će o veličini perioda
- u demonstraciji ovisit će o simulaciji

**Definicija 4.4.5.** 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.

### Kontekstualizacija

-interna putovanja između objekata - možemo ih samo pretpostaviti?!

### 4.4.7 Širina toka

-samo tok?!

**Definicija 4.4.6.** 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 širinu toka. Veličina i reprezentativnost uzorka stanovništva te k-anonimizacija također će utjecati na širinu toka.

Izdvajanje najaktivnijih korisnika (onih s najviše zapisa) zastupljeno je u istraživanjima procjene POM iz CDR [30] [32]. Podjelom korisnika u 3 skupine po količini podataka koju generiraju i analizom svake skupine zasebno, za područje francuske regije Picardie), studija [14] pokazuje da frekvencija korištenja ne uvodi pristranost [14].

Calabrese [29] uzima samo slučajno odabranih 25% korisnika (njih 1 milijun) da pojednostavi analizu, a [8] samo 100,000, kako bi procijenjena POM bila usporediva s onom dobivenom iz ankete provedene na uzorku iste veličine.

#### Poglavlje 4. < Naslov >

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 općenito. [14]. Kod tržišta gdje veliki postotak stanovništva posjeduje mobilni uređaj (npr. 92% za urbana područja Tanzanije) eventualno postojanje pristranosti može se smatrati zanemarivim [11].

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*) [45]. POM-e kreirane iz GPS ili CDR zapisa gusto su ispunjene, najčešće vrlo malim vrijednostima umjesto nula [36].

Broj aktivnih korisnika, a posljedično i širina toka, značajno varira iz dana u dan unutar tjedna [14] [8].

#### Kontekstualizacija

- Razdioba toka prema objektima pojedine kategorije

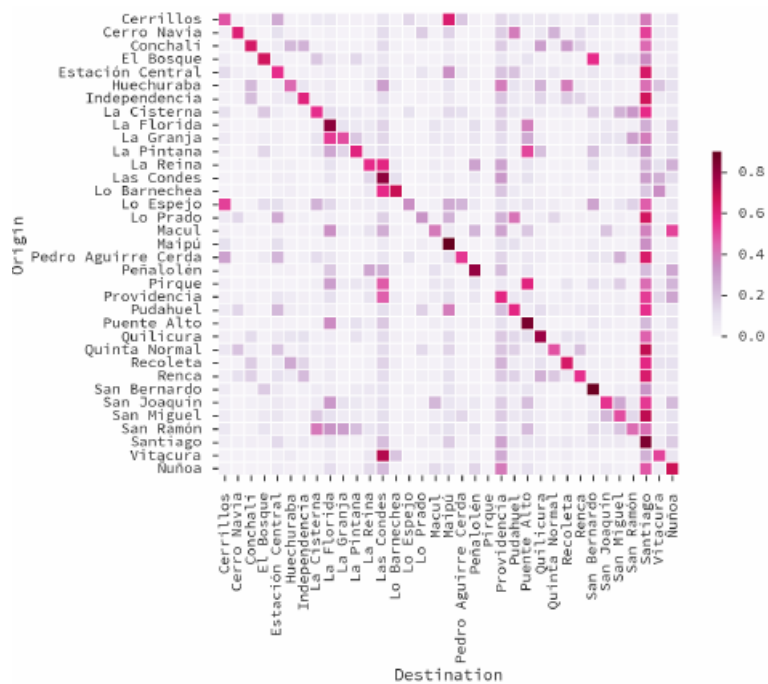
Ispitati:

- kako sve prethodno utječe na širinu toka/konačan broj putovanja?!
- kontekstualizacija na osnovu analize više dana ( tjedan - 3 mjeseca) pomaže da se rekonstruiraju putanje, nema potrebe za odbacivanjem velikog broja

Poglavlje 4. < Naslov >



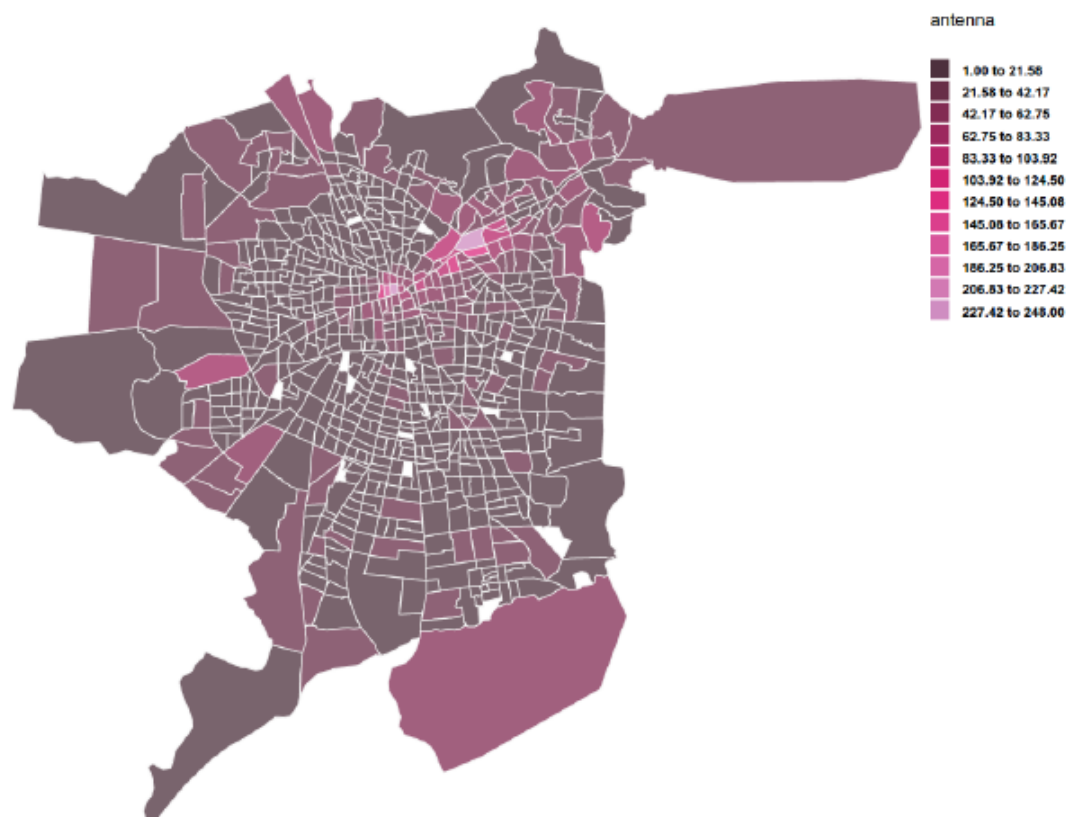
(a) Izvor podataka anketa



(b) Izvor podataka CDR

Slika 4.2 Distribucija CDR putovanja na području Santiago, Čile. POM je normalizirana prema redovima (polazištima) (\*) [8]



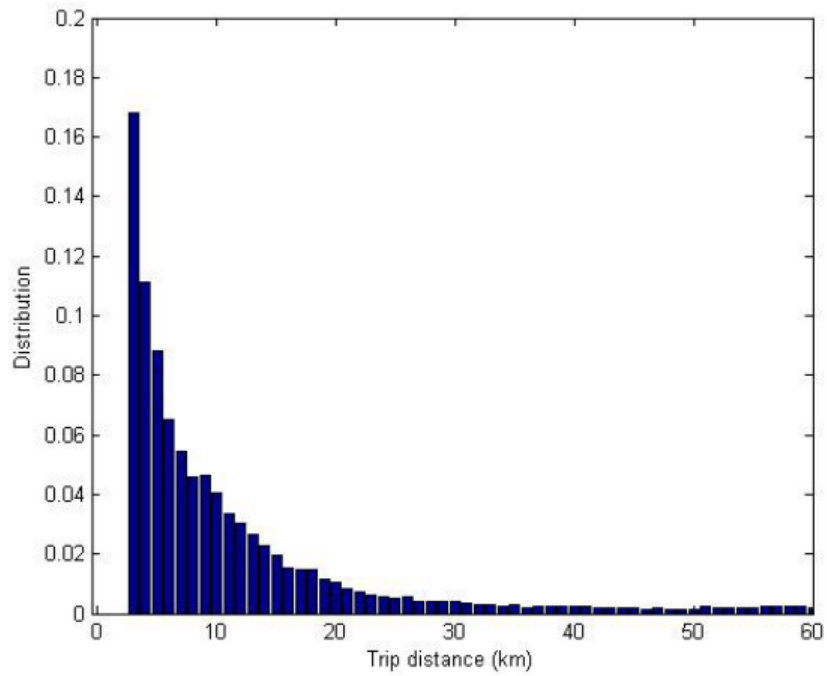


Slika 4.3 Prostorna podjela na anketne jedinice u gradu Santiagu. Gustoća tornjeva baznih stanica po anketnim jedinicama. Gustoća varira od 1 do 250 tornjeva po anketnoj jedinici. [8]

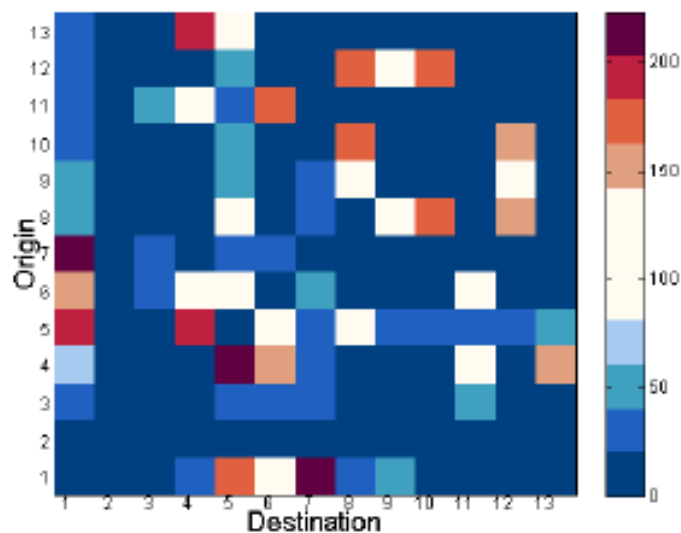
Poglavlje 4. < Naslov >

Country		Area( /km <sup>2</sup> )	No. of BSs	Density of BSs( 1/km <sup>2</sup> )
Europe	Poland	312685	136905	0.4378
	UK	244100	160827	0.6589
	Germany	357376	276408	0.7734
	France	553965	273279	0.4933
	Italy	301338	180867	0.6002
	Netherlands	41864	83891	2.0040
Asia	Singapore	719.1	13396	18.6288
	South Korea	100210	38845	0.3876
	Japan	377972	316977	0.8386
	China	9634057	193764	0.0201
	Tailand	513120	62410	0.1216
	India	2980000	178834	0.0600

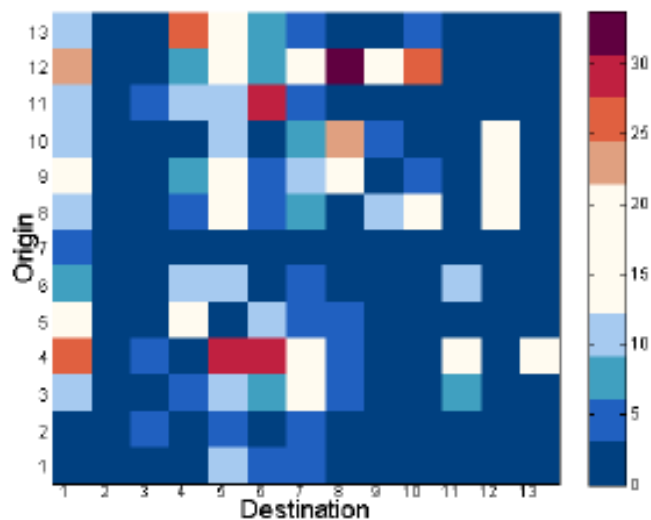
Slika 4.4 Gustoća tornjeva baznih stanica po kvadratnom kilometru za pojedine države i grad-državu Singapur. [9]



Slika 4.5 Distribucija udaljenosti putovanja u kilometrima, Senegal 2016. [10]



(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 4.6 Dvije POM-e - produkti različitih definicija putovanja. Dobivene su iz potpuno istog izvora CDR zapisa. [10]

# Poglavlje 5

## Metodologija

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

### 5.1 Uvjeti usporedbe

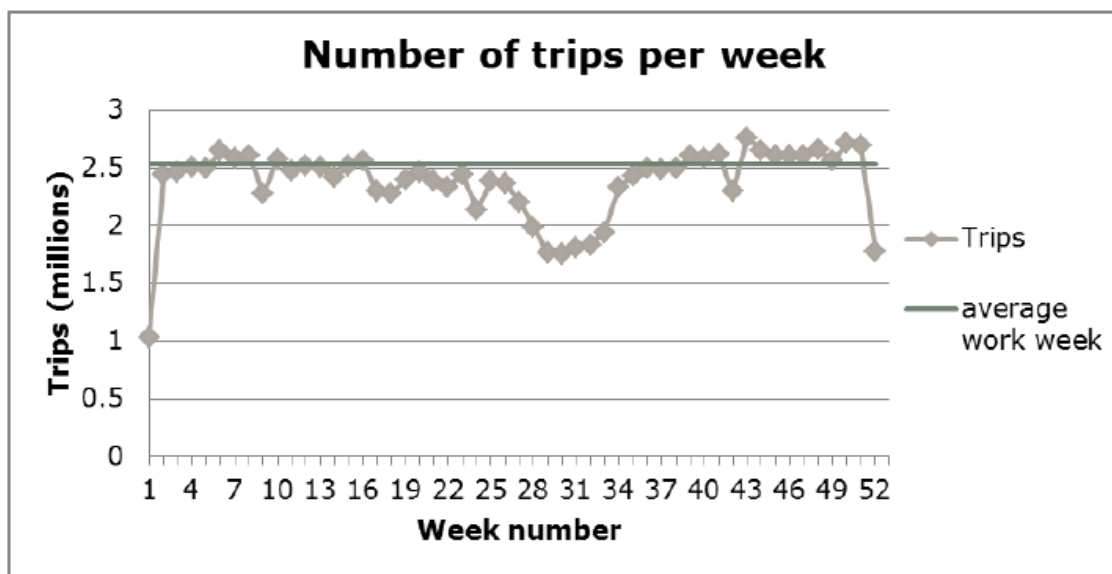
POM-e koje se uspoređuju trebaju zadovoljavati uvjete usporedbe. Riječ je o slijedećim zajedničkim obilježjima:

- isto područje/grad/država
- isto doba godine
- vremensko razdoblje iste duljine

Prednost novih postupaka procjene POM iz CDR u odnosu na ankete stanovništva je da mogu uhvatiti tjedne i sezonske uzorke [29] 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.

- 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.[29][46]



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 [34]

Osim očekivane razlike između radnih i neradnih dana, uočeno je da je petak, kao zadnji radni dan u tjednu, drugačiji od ostalih radnih dana [29] [47][33]. Uočena je pojačana mobilnost petkom na razini regije (Boston Metropolitan Area) [29] i na razini cijele države (Obale Bjelokosti) s povećanjem od 35% u odnosu na nedjelju [47] na temelju POM-a izrađenih iz CDR. Učena 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 bitno utjecati na izmjenu uobičajenog uzorka kretanja. [37]

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

Dio putovanja započinje u jednom a završava u drugom danu. Autori [30][32] zastupaju definiranje 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.

## 5.2 Brzina kretanja

- Analiza prikladnosti odabira perioda i dimenzije ćelija s obzirom na brzinu kretanja agenata ?!
- loše postavke one s kojima će putovanja nužno završiti isključivo u istim/susjednim ćelijama?! - `avemove()` funkcija iz `trajectories`?

## Poglavlje 6

### Rezultati

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

## Poglavlje 7

## Diskusija



## 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] O. S. T. D. . S. Manual. (2005.) Central Florida Expressway Authority. , s Interneta, <http://files.meetup.com/515268/orlando%20Highway%20traffic.pdf> (1.9.2019.).
- [4] J. Xie, Y. Nie, and X. Liu, “Testing the proportionality condition with taxi trajectory data,” Transportation Research Part B Methodological, May 2017., scientific Figure on ResearchGate. , s Interneta, [https://www.researchgate.net/figure/GIS-data-of-the-city-of-Shenzhen\\_fig3\\_316912924](https://www.researchgate.net/figure/GIS-data-of-the-city-of-Shenzhen_fig3_316912924)
- [5] , s Interneta, <https://www-users.cs.umn.edu/~tianhe/BIGDATA/>
- [6] T. Djukić, “Reliability assessment of dynamic OD estimation methods based on structural similarity index,” 92nd meeting of the Transportation Research Board, 2013.
- [7] 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.
- [8] E. Graells-Garrido and D. Saez-Trumper, “A Day of Your Days: Estimating Individual Daily Journeys Using Mobile Data to Understand Urban Flow,” 2016.
- [9] Y. Chen, R. Li, Z. Zhao, and H. Zhang, “Study on Base Station Topology in Cellular Networks: Take Advantage of Alpha Shapes, Betti Numbers, and Euler Characteristics,” 2018.

## Bibliografija

- [10] D. Gundlegard, C. Rydergren, N. Breyer, and B. Rajna, "Travel demand estimation and network assignment based on cellular network data," COMPUTER COMMUNICATIONS, 2016.
- [11] J. Goulding, Best Practices and Methodology for OD Matrix Creation from CDR data, N/LAB, University of Nottingham, 2016.
- [12] 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>
- [13] 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.
- [14] 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.
- [15] 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.
- [16] 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/>
- [17] 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.
- [18] 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](http://www.fsutmsonline.net/images/uploads/fsutms_training/FSUTMS_Webinar_Series_Origin_and_Destination_Skims_and_Matrix_Information_20150917.pdf)
- [19] 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.

## Bibliografija

- [20] A. Peterson, "The Origin-Destination Matrix Estimation Problem- Analysis and Computations," Ph.D. dissertation, Linköping Studies in Science and Technology, 2007.
- [21] 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.).
- [22] prof. dr. sc. Ljupko Simunovic. (2019.) Prometno planiranje. , s Interneta, [https://www.weboteka.net/fpz/Osnove%20prometnog%20in%C5%BEEnerstva/OPI-predavanje\\_6.-Lj\\_Simunovic.pdf](https://www.weboteka.net/fpz/Osnove%20prometnog%20in%C5%BEEnerstva/OPI-predavanje_6.-Lj_Simunovic.pdf) (5.10.2019.).
- [23] N. Jelusic, "Telematicka sucelja (nastavni tekst)," 2016.
- [24] 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>
- [25] "Materijali kolegija usluge zasnovane na lokaciji," 2018,.
- [26] V. Frías-Martínez, E. Frías-Martínez, and C. S. Ruiz, "Estimation of urban commuting patterns using cellphone network data," 2012.
- [27] 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.
- [28] Geomatematicki rjecnik. Struna- Hrvatsko Strukovno Nazivlje. , s Interneta, <http://struna.ihjj.hr/naziv/voronoiov-dijagram/33840/> (20.1.2019.).
- [29] 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.
- [30] C. M. Schneider, V. Beik, T. Couroneé, Z. Smoreda, and M. C. González, "Unraveling daily human motifs," J. R. Soc. Interface, no. 10, May 2013.
- [31] M. Filić, R. Filjar, and K. Vidović, "Graphical Representation of Origin-Destination Matrix in R Statistical Environment," 36. skup o prometnim sustavima s međunarodnim sudjelovanjem AUTOMATIZACIJA U PROMETU, November 2016.

## *Bibliografija*

- [32] J. L. Toole, S. Colak, B. Sturt, L. P. Alexander, A. Evskoff, and M. C. González, “The path most traveled: Travel demand estimation using big data resources,” Transport Research Part C, 2015.
- [33] A. Travassoli, A. Alsger, 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.
- [34] W. Kuhlman, “The construct of purpose specific od matrices using public transport smart card data,” Tech. Rep., 2015.
- [35] 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.
- [36] T. van Vuren, “256 shades of gray - comparing OD matrices using image quality assessment techniques,” Scottish Transport Applications Research (STAR), 2015.
- [37] J. Varun, H. Sunil, R. Vasisht, S. Saragur, and A. Lele, “On the Analysis and Visualisation of Anonymised Call Detail Records,” 2013.
- [38] E. Barbour, C. C. Davila, S. Gupta, C. Reinhart, J. Kaur, and M. C. González, “Planning for sustainable cities by estimation building occupancy with mobile phones,” Nature Communications, 2019.
- [39] (03.06.2019) Data description for urbancps. , s Interneta, <https://www-users.cs.umn.edu/~tianhe/BIGDATA/>
- [40] C. Krause and L. Zhang, “Short-term travel behaviour predictions with gps, land use and point of interest data,” Elsevier, 2018.
- [41] A. Degbelo and W. Kuhn, “Five general properties of resolution,” GIO 2014 - Workshop on Geographic Information Observatories 2014, 2014.
- [42] , s Interneta, <https://documentation.ands.org.au/display/DOC/Spatial+coverage> (4.10.2019.).
- [43] , s Interneta, <http://wiki.gis.com/wiki/index.php/Granularity> (4.10.2019.).
- [44] (03.06.2019) Teralytics. , s Interneta, <https://thenewstack.io/teralytics-takes-big-data-approach-human-movement/>
- [45] 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.

## *Bibliografija*

- [46] 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.
- [47] 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.
- [48] E. Mellegard, “Obtaining origin-destination- matrices from cellular network data,” Master’s thesis, Chalmers University of Technology, Department of Mathematical Sciences, 2011.
- [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 [45] 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 [15] 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).[6]

U [26] 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 [33] 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 [8] 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.

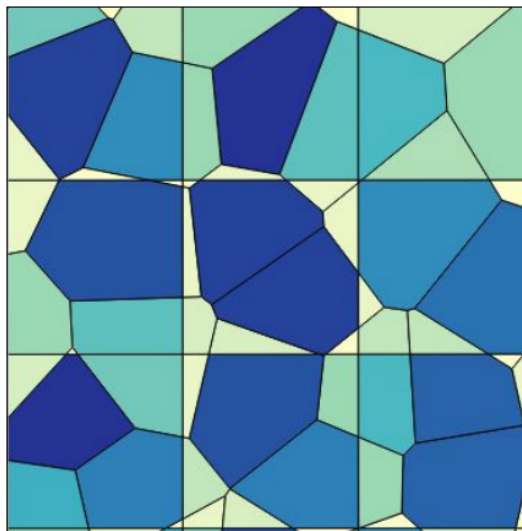
### B.1 Agregacija baznih stanica

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 [11] [27] [48] ili pak jedinicama samouprave (eng. *municipalities*) kako bi POM-e u konačnici bile usporedive s postojećima, dobivenim iz anketa [49] [26]. Nije utvrđivano kolika se greška uvodi ovim postupcima.

### 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 [11]. 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.1 je prikazan treći oblik interpolacije gdje su korišteni podaci o ukupnoj kvadratnoj površini zgrada.



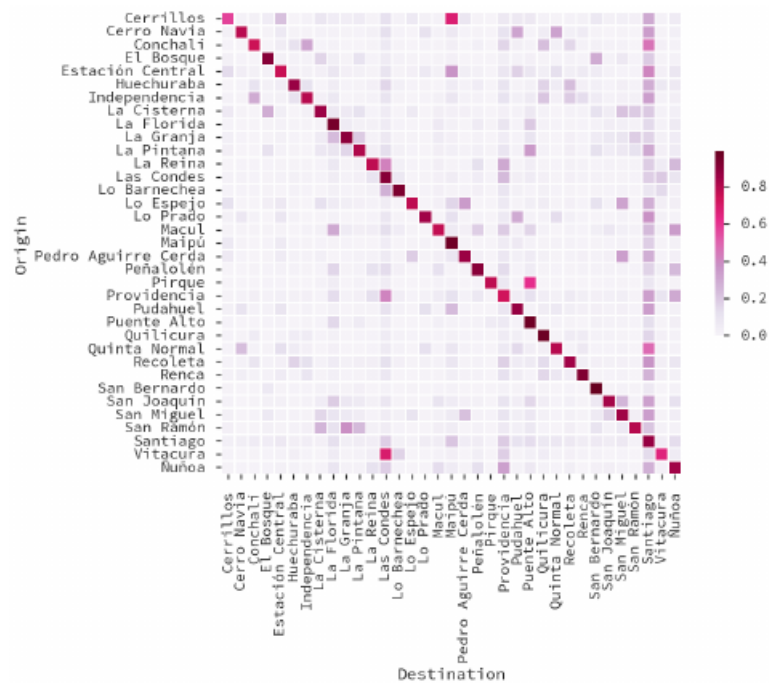
Slika B.1 Primjer mozaika *krhotina* koje dozvoljavaju interpolaciju između Voronoi ćelija i kvadratne mreže. Nijansa odražava postotak površine zgrada u *krhotini* u odnosu na cijelu Voronoi ćeliju kojoj pripada. Rezultat se dobiva na osnovu tih *težinskih faktora* polazišne i odredišne *krhotine* te pripadajućih ("roditeljskih") Voronoi ćelija. Sumom svih rezultata *krhotina* jedne kvadratne ćelije izračuna se konačni tok. [11]

### B.3 Posljedice agregacije prostornih ćelija

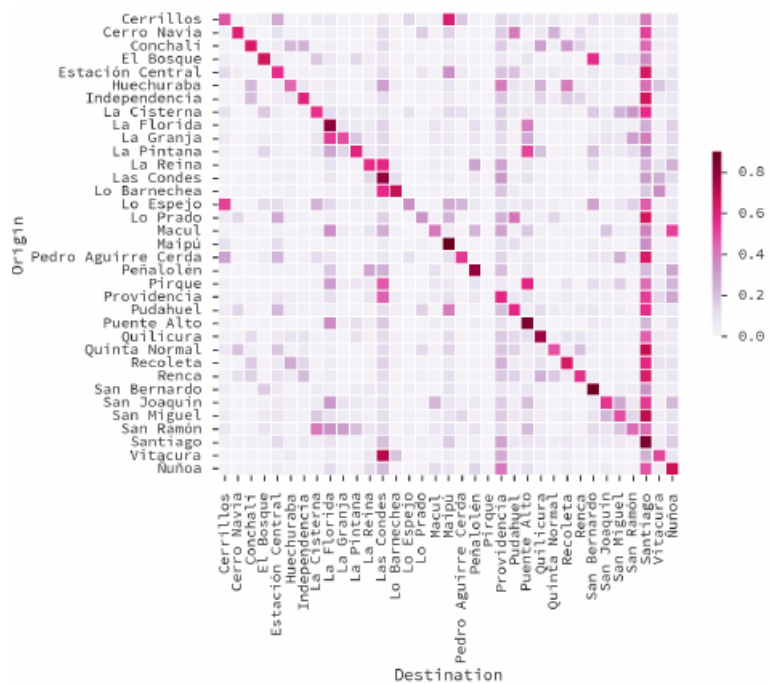
Voronoi ćelije također je moguće ugnijezditi (što se često radi kod usporedbe s POM-ma dobivenim drugim postupkom), a pritom se može u matricu bilježiti i aproksimacija internog toka, odnosno broja putovanja unutar nove ćelije (dijagonala matrice). Takva putovanja nazivaju se internim putovanjima.

Na temelju usporedbi POM-a dobivenih iz CDR i iz anketa u 4 različita grada, studija [32] je pokazala da sa **smanjenjem rezolucije (visokim stupnjem agregacije Voronoi ćelija) stupanj korelacije s POM-ma iz ankete (za isti grad) raste**. No postotak eksternih putovanja koji se gubi u procesu agregacije Voronoi ćelija ana-

liziran je u studijama [8] (Vidi sliku B.2) i [46]. 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. [46].



(a) Izvor podataka anketa



(b) Izvor podataka CDR

Slika B.2 Distribucija CDR putovanja na području Santiaga, Čile. POM je normalizirana prema redovima (polazištima) (\*) [8]

# 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. [48]



## 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*, *building=*

*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.)[32] i kao izvor podataka o podnoj površini objekata (engl. floor area) [11].