

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

Diplomski rad

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

Rijeka, rujan 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, rujan 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, rujan 2019.

-----  
Ime Prezime

# Zahvala

# Sadržaj

Popis slika	ix
Popis tablica	xi
<b>1 Uvod</b>	<b>1</b>
<b>2 Polazišno-odredišna matrica</b>	<b>2</b>
2.1 Definicija Polazišno odredišne matrice . . . . .	2
2.2 Tradicionalni pristupi procjeni Polazišno-Odredišne Matrice . . . . .	3
2.2.1 Ankete . . . . .	3
2.2.2 Prebrojavanje vozila . . . . .	4
2.2.3 Matematičko modeliranje prometa . . . . .	5
2.2.4 Problematika i ograničenja tradicionalnih pristupa procjene Polazišno-Odredišna Matrica (POM)-e . . . . .	5
2.2.5 Pokretna osjetila . . . . .	6
2.3 Suvremeni pristupi procjeni Polazišno-Odredišne Matrice . . . . .	7
2.3.1 Anonimizacija podataka . . . . .	7
2.3.2 Polazišno-Odredišna Matrica iz zapisa o telekomunikacijskim aktivnostima u javnoj pokretnoj mreži . . . . .	8
2.4 Kontekst putovanja . . . . .	10

## Sadržaj

2.4.1	Uloga konteksta u postupku procjene Polazišno-Odredišnih Matrica . . . . .	10
2.4.2	Svrha putovanja kao kontekst putovanja dobivena iz izvora podataka o kretanju . . . . .	11
2.4.3	Način prijevoza kao kontekst putovanja razazan iz izvora podataka o kretanju . . . . .	12
2.5	Kontekst izveden iz vanjskih izvora . . . . .	12
2.5.1	OpenStreetMap . . . . .	13
2.5.2	Primjeri iskorištavanja konteksta izvedenog iz vanjskih izvora .	14
2.5.3	<i>Points of Interest</i> . . . . .	15
2.6	Geometrija prostorne podjele . . . . .	15
2.7	Vremenska interpretacija Polazišno-Odredišnih Matrica . . . . .	22
2.7.1	Vremenski zavisne Polazišno-Odredišne Matrice . . . . .	22
2.8	Diskusija . . . . .	25
2.9	Alternative izvora prostornih podataka za razvoj Polazišno-Odredišne Matrice . . . . .	26
2.9.1	Združena očitavanja prijamnika za satelitsku navigaciju . . . . .	26
2.9.2	Javni prijevoz i <i>pametne kartice</i> ( <i>Smart Card</i> sustavi) . . . . .	26
2.9.3	WiFi . . . . .	27
<b>3</b>	<b>Validacija Polazišno-Odredišne Matrice</b>	<b>28</b>
3.1	Točnost Polazišno-Odredišne Matrice . . . . .	28
3.1.1	Metrike za procjenu točnosti Polazišno-Odredišnih Matrica . .	29
3.2	Grafički prikaz (oblik) Polazišno-Odredišne Matrice . . . . .	30
3.3	Strukturalna sličnost Polazišno-Odredišnih Matrica . . . . .	31
3.3.1	MSSIM . . . . .	32

## *Sadržaj*

<b>4</b>	<b>Usporedba dviju Polazišno-Odredišnih Matrica</b>	<b>35</b>
4.1	Uvjeti usporedbe . . . . .	35
4.2	Kriteriji usporedbe . . . . .	35
4.2.1	Definicija putovanja . . . . .	36
4.2.2	Prostorna razlučivost (rezolucija) . . . . .	38
4.2.3	Vremenska razlučivost . . . . .	39
4.2.4	Širina toka . . . . .	39
4.2.5	Gustoća informacija - kontekst . . . . .	40
4.3	Diskusija . . . . .	46
<b>5</b>	<b>Zaključak</b>	<b>47</b>
	<b>Bibliografija</b>	<b>48</b>
	<b>Pojmovnik</b>	<b>53</b>
	<b>Sažetak</b>	<b>54</b>
<b>A</b>	<b>Naslov priloga</b>	<b>55</b>
A.1	Naslov sekcije . . . . .	55
A.2	Naslov sekcije . . . . .	55



# Popis slika

2.1	Mreža za analizu prometa sastoji se od 221,14 cesta - plave linije i 3,199 TAZ zona - granice označene iscrtanom crnom linijom [1] . . .	16
2.2	Jedinice lokalne samouprave za grad Shenzhen [2] . . . . .	17
2.3	Voronoi dijagram 1090 ćelija baznih stanica - Shenzhen . . . . .	19
2.4	Podjela na šesterokute Shenzhen - grupacija u različite boje na osnovu širine toka [3] . . . . .	20
2.5	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. [4]	21
2.6	Prijedlog prilagodbe administrativnih jedinica mreži kretanja - bojom su označene grupacije nastale na temelju POM-e, crnom su označene postojeće granice administrativnih jedinica [5] . . . . .	22
2.7	Broj putovanja javnim prijevozom zabilježenih <i>OV-chipkaart</i> pametnim karticama u Nizozemskoj izražen u u milijunima po tjednima u godini. Zelena linija označava prosječnu vrijednost radnog tjedna [6]	24
3.1	<i>Spaghetti-effect</i> - problem koji se može javiti kod grafičkog prikaza tokova POM-e visoke prostorne i niske vremenske rezolucije. Na slici su prikazani jednodnevni tokovi. [7] . . . . .	31

3.2	Narušenu strukturalnu sličnost lakše je uočiti nego narušen MSE. Structural Similarity index (SSIM) (i Mean Structural Similarity index (MSSIM)) u odnosu na referentnu POM-u može biti korišten kao dodatna informacija ( <i>goodness of fit</i> mjera) prilikom validacije POM-e [8] . . . . .	34
4.1	Dvije POM-e - produkti različitih definicija putovanja. Dobivene su iz potpuno istog izvora CDR zapisa. [9] . . . . .	41
4.2	Distribucija udaljenosti putovanja u kilometrima [9] . . . . .	42
4.3	Distribucija CDR putovanja na području Santiaga, Čile. POM je normalizirana prema redovima (polazištima) (*) [10] . . . . .	43
4.4	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. [10] . . . . .	44
4.5	Gustoća tornjeva baznih stanica po kvadratnom kilometru za pojedine države i grad-državu Singapur. [11] . . . . .	45

# Popis tablica

# Poglavlje 1

## Uvod

POM *eng. Origin-destination Matrix (ODM)* ili *Trip Table* alat je koji omogućuje opis i sustavnu statističku procjenu migracija stanovništva na nekom području u zadanom prostorno-vremenskom okviru. 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.

Za razliku od tradicionalnog pristupa prebrojavanja putovanja i putnika na licu mjesta, anketiranjem ili pomoću različitih strateški postavljenih osjetila, za procjenu POM-a 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. Kada daju informaciju o kontekstu - svrsi kretanja- POM-e su dodatno obogaćene i postaju izvor informacija za primjenu i van konteksta prometa.

(...)

## Poglavlje 2

# Polazišno-odredišna matrica

### 2.1 Definicija Polazišno odredišne matrice

**Definicija 2.1.1.** Polazišno-odredišna matrica definirana je tako da svaki njen element predstavlja broj putovanja između odgovarajućeg para zona unutar promatranog područja u promatranom vremenskom razdoblju  $T$ , u svakom smjeru zasebno. Redovi u POM-i  $i = 1, 2, 3 \dots m$  predstavljaju polazišta (izvore) prometnog toka, a stupci  $j = 1, 2, 3 \dots n$  njihova odredišta (ušća). Element na poziciji  $i, j$  u POM-i predstavlja broj putovanja iz pripadajuće polazišne zone  $i$  u pripadajuću ciljnu zonu  $j$ .

POM-a se može smatrati matričnim zapisom težinskog, usmjerenog grafa, gdje su težinski faktori usmjerenih veza širine prometnih tokova između čvorova mreže koji predstavljaju zone na koje je promatrano područje podijeljeno.

Postoje brojni postupci procjene POM o kojima je riječ u narednom odjeljku. 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...).

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

Kada kontekst (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.

## 2.2 Tradicionalni pristupi procjeni Polazišno-Odredišne Matrice

Treba li ovo poglavlje nositi naslov:

Tradicionalni pristupi procjene POM?

Ali tu se navode samo neki izvori ulaznih podataka.

Nema govora o **metodologiji** kojom se obavlja procjena.

Stari naslov bio je Tradicionalni pristupi generiranju POMa, na vaš prijedlog je izmjenjen u Generiranje POM-e, a potom u Procjena Polazišno-Odredišne Matrice

### 2.2.1 Ankete

Tradicionalni pristupi uključuju metode provođenja anketiranja. Postoji nekoliko vrsta anketa: u kućanstvima, presretanje vozila, anketiranje prijevoznih kompanija i tvrtki, tranzit na granici, turisti u hotelima i parking [12].

Podjela po subjektima anketa. Predlažem detaljniju razradu u metodološkom smislu, uz korištenje literature poput:

Hyndman, R J, and Athanasopoulos, G. (2018). Forecasting: Principles and Practice (2nd ed). (Chapter 4: Judgemental Forecast)). Preuzeto s: <https://otexts.com/fpp2/judgmental.html>

Ovakve metode su zahtjevne iz perspektive utrošenog vremena, uloženog truda i radne snage stoga se ne provode često. Anketiranje putovanja u kućanstvima provodi se na područjima brojnih svjetskih metropola tek jednom u deset godina. U većini svjetskih metropola na provođenje anketa odlazi najveći postotak budžeta planiranja, u procjeni oko 7.4 milijuna dolara godišnje izdvaja se za ankete po kućanstvima. [12]

Anketiranje vozača presretanjem vozila eng. Road Side Interview (RSI) u prosjeku košta čak 10 eura po jednom zabilježenom putovanju i ukupno obuhvati 10% prometa. Ova vrsta najčešće se provodi na autocestama. [13]

Provođenje ovih postupaka zahtjeva mnogo vremena. Pokrivaju malen dio populacije te je na taj način dobivena POM-a opterećena neizbježnom pristranošću.[14]

Volio bih vidjeti konkretniji prikaz koji će ukazati na suštinske prednosti i nedostatke, ne investicijske. Naime, cilj ovog poglavlja jest ukazati kako **suvremeni tehnološki sustavi nedostatke tradicionalnih pristupa uspješno prevladavaju.**

Studija iz 2010. [12] pokazala je da se zadovoljavajuća razina kvalitete POM-a procjenjenih isključivo iz anketa u kućanstvima postiže tek ako uzorak obuhvaća 50% populacije.

### 2.2.2 Prebrojavanje vozila

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 (ra-

darska, magnetna, video-senzori, zvučna..) i ručno prebrojavanje. Primjenom ovih postupaka nije moguće pouzdano odrediti stvarna polazišta i odredišta putovanja [15].

### 2.2.3 Matematičko modeliranje prometa

Matematičko 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 u mnogim 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 su primjerice da je ukupna dnevna potražnja podjednaka u oba smjera kod svakog para zona na području grada [16], da svi korisnici imaju saznanja o situaciji na cesti [4] te da je mobilnost općenito manja tijekom zime nego tijekom ljeta [15].

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. Izvedene verzije integriraju i neke druge (socio-ekonomske) faktore. [9]

### 2.2.4 Problematika i ograničenja tradicionalnih pristupa procjene POM-e

Širenje gradova i rast stanovništva rezultirali su rastućim brojem sve ozbiljnijih prometnih zastoja u velikim gradovima diljem svijeta. Prepoznata je potreba za strategijama upravljanja prometom koje će uvažiti dinamiku 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, poticati ekonomski rast i smanjiti zagađenje



okoliša. Suvremeni odnosi promijenili su prirodu kretanja ljudi i dobara zbog čega tradicionalni pristup postaje zastario i neprimjeren.

Redefiniranje metodologije koja će pružiti povećanu kvalitetu i transparentnost uz ispunjavanje suvremenih zahtjeva na tehnologije (cijena, zaštita privatnosti i podataka itd.) privlači pažnju istraživača i van područja prometne znanosti. Oni predlažu alternativne izvore podataka o kretanju za promatranje obrazaca putovanja i pristupe procjeni POM-a, no prema [13] nailaze na problem da struka zahtijeva da se njihova vrijednost dokazuje usporedbom s tradicionalnim pristupima, dobro poznatim u praksi.

Nije baš tako. Brkaju se znanost, istraživanje, tehnologija i korištenje. Naručitelj hoće usporediti s onim što je njemu poznati, ali to je usporedba krušaka i jabuka. Gledano iz perspektive validacije istraživačkog modela, usporedba se može raditi s drugom, neovisnom metodom koja daje približno jednako kvalitetne rezultate. A problem ostaje otvoren za raspravu, **koliki je udio populacije stvarno pokriven i je li model razvije na takvom podskupu populacije relevantan.**

### 2.2.5 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 sposoban opažati vrijednosti definiranih parametara gibanja vozila na koje je postavljeno te ih pohranjivati i prenositi radi naknadnog procesiranja. Podatke koje osjetilo prikuplja su položaj, brzina i smjer kretanja vozila.

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). Danas je FCD metoda zastupljena 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. [17]

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.

U čemu je razlika (kako tehnološka, tako i pravna)?

Je li vozilo s eCall-om (pa tako i sa SIM karticom) vozilo ili pametni telefon?

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 promatranom trenutku.

## 2.3 Suvremeni pristupi procjeni Polazišno-Odredišne Matrice

(...)

### 2.3.1 Anonimizacija podataka

Sve brojnije informacijske i telekomunikacijske usluge generiraju masivne skupove podataka vezanih za gibanje u prostoru (mobilnost). Podatci koje generiraju pametni telefoni omogućuju podatkovnim analitičarima razumjeti ponašanje pojedinaca u brojnim domenama, uključujući i njihov obrazac kretanja.

Operatori javne pokretne mreže moraju voditi računa o privatnosti pojedinaca kada je riječ o podacima koje oni prikupljaju. Ukoliko ustupe podatke za razvoj POM-a, takvi podaci moraju biti anonimizirani te ne smije postojati mogućnost povratnog inženjerstva kojim bi se povratio izvorni identitet. Podaci moraju biti anonimizirani prije pristupa da mogu biti korišteni bez razotkrivanja pojedinaca u

skladu s postavkama General Data Protection Regulation (GDPR).

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

Operatori javnih pokretnih mreža prikupljaju podatke o aktivnostima u javnoj pokretnoj mreži za potrebe pružanja i naplate usluga te za održavanje operativnosti same mreže. Pokretni uređaj (npr. pametni telefon) u javnoj pokretnoj mreži kontinuirano je 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 (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.

Operatoru su dodatno poznati 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 prikupljeni svakih 20 min do 2 sata o tome na koju skupinu baznih stanica je pokretna stanica spojena (prema [15] svaka LA skupina ima od 150 do 500 baznih stanica), te podatke o tome kada je pokretna stanica uključena u mrežu i isključena iz mreže.

U novije vrijeme su predloženi postupci u kojima se koriste upravo anonimizirane CDR kao izvor podataka za procjenu POM-a. [9] [14] [18] Metodologija korištenja odnosnih elemenata CDR-a za procjenu POM predložena je još 1994. godine.[9] U ostatku rada, na mjestima gdje nije izričito naveden izvor podataka

o kretanju, riječ je o postupku ili POM-i procijenioj iz CDR. Pri procjeni POM-e, ako su dostupni, i ranije navedeni dodatni podatci mogu se uzeti u obzir.

### Putovanje u kontekstu izrade polazišno-odredišne matrice iz zapisa o telekomunikacijskim uslugama

Priroda ovog pristupa dubinskoj analizi podataka za procjenu POM-a traži definiciju pojma zaustavljanja putovanja, polazišta i odredišta.

**Definicija 2.3.1.** Zaustavljanje (engl. *Stop*) je skup minimalno  $n$  događaja (telekomunikacijskih aktivnosti) u nizu koji su zabilježeni na istom položaju unutar definiranog minimalnog vremenskog perioda.

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. U slučaju kada je  $n > 1$  zaustavljanje ima 2 vremenske oznake, trenutak kada je zabilježen početak i trenutak kada je zabilježen kraj zaustavljanja.

Posve neprikladno, ovdje se razmatra algoritam razvrstavanja putovanja, umjesto da se precizno i egzaktno definira **putovanje kao informacijski objekt**. Ali opet, trebalo bi najprije definirati **putovanje u kontekstu izrade/procjene POM**

\* U definiciju 2.3.2 dodano je: Putovanje kao informacijski objekt...

**Definicija 2.3.2.** Putovanje je skup dva zaustavljanja koja zadovoljavaju unaprijed definirana minimalna i maksimalna ograničenja međusobne vremenske i prostorne udaljenosti. Putovanje kao informacijski objekt čine: polazišni položaj iz kojeg je putovanje krenulo, odredišni položaj u kojem se putovanje završava, vremenska oznaka kada je počelo i kada je završilo. Kada je poznat, u navedeno se ubraja i kontekst putovanja.

Definicija konteksta dana je u narednom odjeljku 2.4. Kraj polazišnog zaustavljanja postaje početak putovanja, a početak odredišnog zaustavljanja kraj putovanja. Niz kontinuiranog zaustavljanja na različitim položajima otkriva putanju kretanja.

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

## 2.4 Kontekst putovanja

Definirajte kontekst **iz elemenata putovanja** na egzaktn i jedinstven način, bez povezivanja na algoritme i procese koje još niste definirali!

Kontekst i što pod njim podrazumijevate, **a da je vezano za POM.**

*Transkript Putovanje treba definirati u kontekstu... Formalan opis koji bi obuhvatio kontekst... Svaki zastanak koji je sa svrhom bi bio putovanje... Putovanje nije ako sam ja stao na semaforu... Pokušaj definiranja zaustavljanja može biti po provedenom vremenu kad (koliko) sam stajao, ali to ne mora biti to - može biti gužva (na semaforu).*

**Definicija 2.4.1.** Kontekst putovanja je informacija vezana uz putovanje, njegovo polazište ili odredište koja procjenjuje ili upućuje na motiv (svrhu) kretanja, način prijevoza ili željeni faktor po kojem se putovanja mogu kategorizirati i time ih preciznije definirati (...u informacijskom smislu).

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

Poglavlje se treba baviti kontekstom i **uloži konteksta u postupku procjene POM**, ne POM kao takvom.

### 2.4.1 Uloga konteksta u postupku procjene Polazišno-Odredišnih Matrica

Svako zaustavljanje (2.3.1) ne mora nužno biti (postati) početak ili kraj putovanja. Postavljanjem uvjeta da zaustavljanje jest polazište ili odredište putovanja samo ako

mu se dokaže kontekst od interesa, eliminiraju se usputna zaustavljanja. Primjerice, ako je kontekst od interesa ograničen na isključivo "posao" i "prebivalište" s ciljem promatranja samo komutacijskih putovanja, zaustavljanje na benzinskoj crpki neće se gledati kao kraj putovanja. Komutacijska putovanja se ponavljaju na dnevnoj bazi te su stoga često upravo ona koje se nastoji obuhvatiti POM-ma.

Ograničavanjem na isključivo komutacijska putovanja, velik će dio putovanja \*koja su predvidiva na razini toka\* biti odbačen. Stoga uvođenjem nekoliko generalnih kontekstnih kategorija ...

Kontekst putovanja (...) [19]

#### 2.4.2 Svrha putovanja kao kontekst putovanja dobivena iz izvora podataka o kretanju

Kada su dostupni zapisi prikupljeni dulje vremensko razdoblje, tjednima ili mjesecima putovanje se može definirati na drugačije - kao isključivo komutacijsko. Pojednostavljeno, analizom zapisa može se utvrditi na kojem položaju neki pojedinci pretežito borave u tipičnom radnom vremenu te u kojoj tokom noći. Prvi položaj za njihova putovanja postaje *Work* destinacija, drugi *Home*, a sve preostale destinacije svrstaju se u zasebnu kategoriju *Other*. POM-e uključuju samo putovanja iz jednog tipa destinacije u drugi npr. Home-Work (HW) i Work-Home (WH). Kontekst, odnosno svrha svakog od putovanja koja čine POM-u je ili *odlazak na posao* ili *odlazak kući (s posla)*. [4][10][7][20][21]. POM-e koje sadrže ta putovanja usporedive su s onima dobivenim iz anketa stanovništva s istom podjelom putovanja prema svrsi. Komutacijska putovanja su ponavljajuća i u tom smislu predvidiva i zanimljiva u strateškom planiranju.

Studija A iz 2013. [22] pokazala je da su CDR zapisi u trajanju od 6 mjeseci bili su dovoljni za konačnu točnu procjenu položaja čak 98% kućanstava nešto više od 39 tisuća najaktivnijih korisnika na području Pariza i Chicaga, izdvojenih iz ukupnog skupa korisnika (milijuni korisnika). U podacima se ispitalo gdje svaki od korisnika provodi najviše vremena između 0:00 i 6:00 sati. Dvije godine kasnije studija B [21] definira kućanstvo kao mjesto gdje korisnik provodi najviše vremena između 20:00 i

7:00 na radne dane, a radno mjesto kao mjesto gdje korisnik provodi najviše vremena, barem 3 puta tjedno, između 7:00 i 20:00. Godinu dana kasnije studija C [9] koristi intervale od 22:00 do 7:00 te od 9:00 do 16:00. Ostale lokacije svrstava u kategoriju *Ostalo eng. Other*. Kod ovih vrsta POM-a najčešće naglasak nije stavljen na dobru vremensku rezoluciju kao što je to slučaj kod tranzitnih.[9]

### 2.4.3 Način prijevoza kao kontekst putovanja razazan iz izvora podataka o kretanju

Tradicionalni postupci ograničeni su samo na onaj dio infrastrukture gdje se provode intervjui ili gdje su postavljena osjetila za prebrojavanje prometa. Telekomunikacijski zapisi neovisni su o prometnoj infrastrukturi. Oni obuhvaćaju sve vrste vozila, pješački promet kao i svu prometnu infrastrukturu te područja bez razvijene prometne infrastrukture. Uklanjaju potrebu za različitim postupkom prebrojavanja za pojedine načine kretanja.

Autori studije [21] tvrde da isključivo CDR zapisi ne osiguravaju rezoluciju pogodnu za određivanje sredstva kretanja iako se u praksi [23] prema brzini kretanja pokretne stanice ili skupa njih istovremeno donosi zaključke o dominantnom sredstvu kretanja (pješak, automobil, javni prijevoz, avion). Iznimno se mogu izdvojiti bazen stanice dedicerane pružanju usluge isključivo u podzemnoj željeznici. Studija iz 2016. [10] pokazala je da je moguće izdvojiti taksiste/dostavljače uz pretpostavku da su to oni korisnici koji prijeđu preko 100 km dnevno unutar grada.

## 2.5 Kontekst izveden iz vanjskih izvora

### Opis metodologije?

Neka istraživanja spojila su ljudsku mobilnost s namjenom zemljišta (engl. *land use*) ili točkama interesa engl. *Points of Interest (POIs)* kako bi grupirala urbana područja prema njihovoj funkciji ili upotrebi. Izvori podataka o kretanju u tim istraživanjima obuhvaćaju zapise Global Positioning System (GPS) putanja, položaja ukrcaja i iskrcaja taxi vozila, zapisa koje generiraju pametne kartice u javnom pri-

jevozu, podatke o dijeljenim lokacijama na društvenim mrežama (engl. *check-in*) i zapise o telekomunikacijskim aktivnostima (CDR) [14].

Izvori podataka o kontekstu također variraju od istraživanja do istraživanja, a uključuju opise objekata u prostoru i relacija među tim objektima. Korišteni podaci su primjerice broj objekata, površina koju oni zauzimaju te, za kontekst kao svrhu putovanja najbitnija, socio-ekonomska svrha (funkcija) pojedinog objekta.

### 2.5.1 OpenStreetMap

OpenStreetMap (OSM) je svjetski raširen projekt koji kreira i pruža slobodne zemljopisne podatke (zemljovide gradova i naselja) temeljen na volonterskom doprinosu zajednice. 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 kojeg nazivamo 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=hotel*, *building=church*, *leisure=sports\_centre*, *amenity=school*. Primjer žiga vezanog uz put *highway=residential*.



## 2.5.2 Primjeri iskorištavanja konteksta izvedenog iz vanjskih izvora

### Model raspodjele toka

Inovativni pristup kontekstualizaciji toka korištenjem OpenStreetMap baze predstavljen je u radu [24]. Na osnovu pripadajućih opisnih podataka, prostorni objekti na području interesa kategorizirani su prema tipu socio-ekonomske aktivnosti u koju su uključeni. 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 i promatranog vremenskog okvira, ukupni ulazni tok u odredišnu ćeliju dijeli na 6 tokova usmjerenih prema objektima tih kategorija. Iz jedne POM-e tako se dobije 6 POM-a, po jedna za svaku od kategorija. U demonstraciji modela korištene su POM-e dobivene iz javno dostupnih, anonimiziranih telekomunikacijskih zapisa na području kineskog grada Shenzhena.[25][26] **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 dobivamo nove informacije o kretanju unutar ćelije, odnosno o konačnoj destinaciji nakon ulaska u ćeliju. U tom smislu imamo informaciju o internom kretanju.

### Izvor infrastrukture

OSM doprinosi istraživanjima u ovom području i kao izvor prometne infrastrukture i pripadajućih meta podataka (kolnik, mostovi, ograničenja brzine i dr.). Raspodjela toka po prometnoj infrastrukturi nadilazi područje ovog rada no jest dio problematike modeliranja prometa. [21] Još jedna moguća primjena spominje u potpoglavlju 2.6 koje opisuje interpolaciju Voronoi ćelija.

### 2.5.3 *Points of Interest*

U jednom od svojih radova grupa autora kvantizira vezu između ukupno 23 namjenske kategorije POIs iz službene gradske baze (npr. tvornice, bolnice, javne škole, religijski objekti, hoteli, knjižnice, sportski objekti) i onime što nazivaju 3 tipa privlačenja. Dolaze do zaključka da postoje 3 tipa privlačenja te da 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. [14] U radu koriste podatke iz službene baze s 12,000 POIs i CDR zapisima iz razdoblja od mjesec dana. Cilj projekta bio je ispitati mobilnost na području grada Riyadha, u Saudijskoj Arabiji za planiranje izgradnje podzemne željeznice. Njihovi rezultati mogu se primijeniti u planiranju pozicioniranja budućih objekata u gradu.

## 2.6 Geometrija prostorne podjele

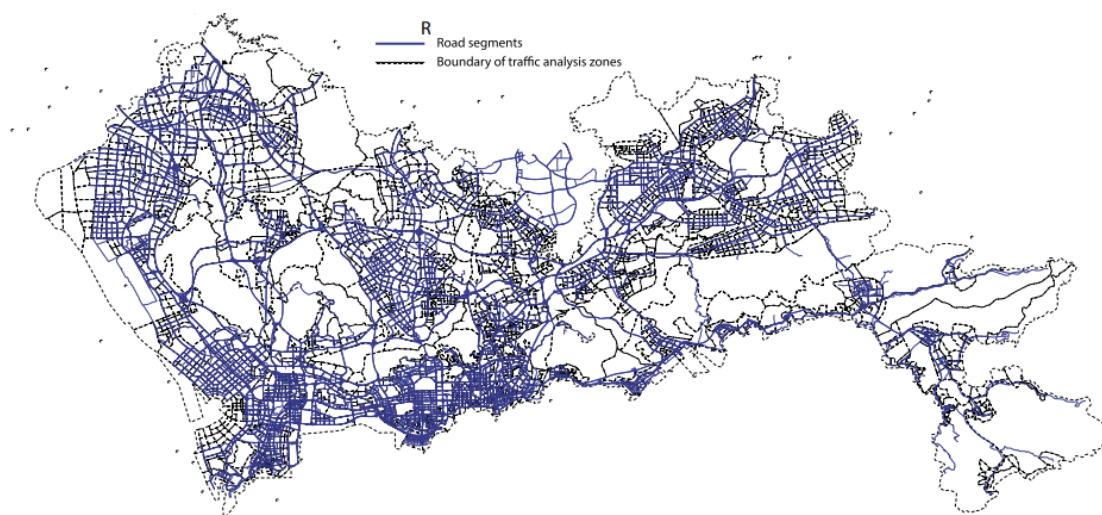
**Definicija 2.6.1.** Geometrija prostorne podjele je način na koji je područje od interesa podijeljeno u manje površine koje predstavljaju zone (polazišta/odredišta). Različite podjele razlikuje ponajprije njihov broj i (ne)uniformnost površine.

Različiti izvori podataka za procjenu POM-e rezultirali su i različitim prostornim podjelama promatranih područja. Prostorna podjela uvjetovana je postupkom prikupljanja podataka i/ili infrastrukturom. Nije moguće u startu osigurati primjerice uniformnu pravokutnu podjelu.

Konačna geometrija prostorne podjele alternativnih izvora ovisi o prostornoj podjeli *grand truth* POM-e. To je POM-a koja se uzima kao najvjerniji prikaz stvarnog kretanja i u odnosu na koju se procjenjuje točnost procijenjene POM-e. Obično se radi o jedinici prostorne podjele naziva Traffic Analysis Zone (TAZ) kod analize prometa, o anketnim jedinicama ili jedinicama samouprave. Najprihvaćenija prostorna podjela kod POM-a izvedenih iz CDR je ona na Voronoi ćelije.

### Traffic Analysis Zone (TAZ)

Traffic Analysis Zone (TAZ) je strani naziv za jedinicu prostorne podjele (ćeliju) kod analize prometa. Kod programskih podrška za analizu prometa veličina jedinica varira i nije uniformna. Podjela se može temeljiti na populaciji ili broju zaposlenih na nekom prostoru, primjerice svaka ćelija pokriva prostor na kojem obitava otprilike tri tisuće ljudi (propisanih preporučenih vrijednosti nema), ili podjela u kojoj ćelije generiraju podjednak broj izlaznih putovanja (jednaka širina izlaznih tokova) [4]. U hijerarhijskoj podjeli kod ankete stanovništva općenito postoje manje i veće jedinice podjele, a nema podjele koja bi precizno odgovarala TAZ. Do 2000. godine zastupljen je bio jedan model predviđanja prometa, a veličina TAZ bila je uvijek približno jednaka anketnom bloku od 600 do 3,000 ljudi. Pojavom kompleksnijih modela predviđanja prometa (i modela upotrebe zemljišta) 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.[27]



Slika 2.1 Mreža za analizu prometa sastoji se od 221,14 cesta - plave linije i 3,199 TAZ zona - granice označene iscrtanom crnom linijom [1]

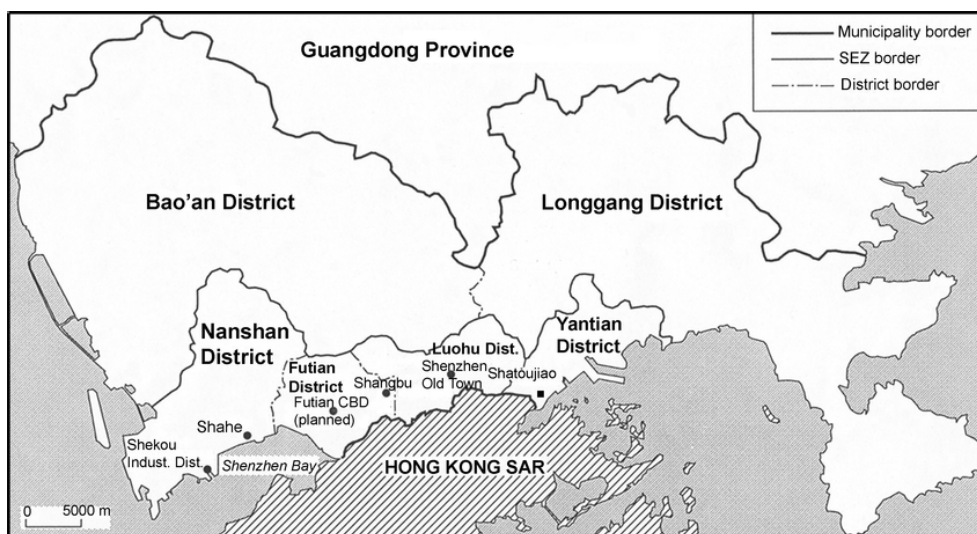
Prema podacima korištenim u istraživanju iz 2018. [28] Shenzhen je podijeljen na 996 TAZ od kojih 844 zone imaju barem jednu autobusnu ili stanicu podzemne

željeznice. Godinu dana ranije objavljeno istraživanje spominje više zona (Vidi sliku 2.1).

Kod analize na području cijele države [4] koristi *k-sredina* algoritam za grupiranje baznih stanica u broj grupa jednak broju TAZ na području Senegala. [14] na području grada Riaydha agregira bazne stanice u 1492 TAZ-e.

### Jedinice samouprave i anketne jedinice

[29] ne definira unaprijed broj grupa već za grupaciju koristi *mean-shift* algoritam koji sam određuje broj grupa na području Švedske, potom koristeći OSM grupe veže uz imena gradova i mjesta. U istraživanju [30] bazne stanice na području Kolumbije grupiraju po jedinicama samouprave (eng. *municipalities*). Isti postupak je korišten i u istraživanju na području Madrida, kako bi POM-e u konačnici bile usporedive s onima dobivenim iz anketa [18]. Prikaz jedinica samouprave za grad Shenzhen možemo vidjeti na slici 2.2. Podjela na gradske četvrti obično je relativno uniformna [14].



Slika 2.2 Jedinice lokalne samouprave za grad Shenzhen [2]

## Voronoi teselacija

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

**Definicija 2.6.2.** 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.

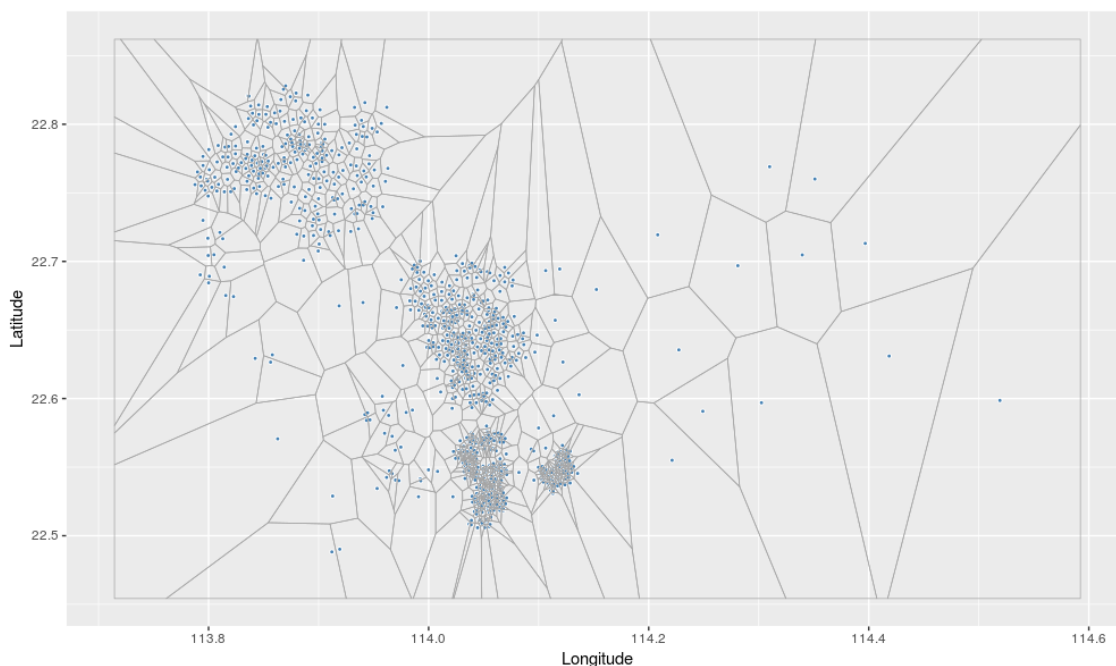
Točke skupa su položaji baznih stanice, a ćelije oko njih aproksimacija površine koju ta bazna stanica pokriva. U stvarnosti se te površine prekrivanja preklapaju jer područje prekrivanja nije sasvim fiksno i ovisi o trenutnoj aktivnosti svake bazne stanice [15]. (Vidi sliku 2.3)

Operatori u CDR zapisima ne otkrivaju položaje baznih stanica već samo njihov ID. Postoji nekoliko javno dostupnih baza kao što je OpenCellID [31] nastalih volonterskim doprinosom koje pohranjuju podatke o procjenama položaja baznih stanica, snazi signala i sl. Položaji se prikupljaju pomoću aplikacije instalirane na mobilnom uređaju, registriranjem 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. [29]

Gustoća Voronoi ćelija snažno je korelirana s gustoćom stanovništva te stoga bolju prostornu rezoluciju ima na područjima veće gustoće populacije. Mjerna nesigurnost određivanja položaja iznosi od približno stotinu metara u gusto naseljenom, do nekoliko kilometara u ruralnom području [20]. Gundlegard [9] smatra da neuniformnost ćelija nije problem, već preklapanje područja pokrivanja u stvarnosti, posljedica čega su stalne oscilacije. Uređaj mijenja baznu stanicu na koju je spojen iako svo vrijeme zadržava isti položaj. Učinak koji oscilacija ima na definiranje putovanja može ublažiti je određivanjem minimalne udaljenost putovanja ili potpu-

## Poglavlje 2. Polazišno-odredišna matrica

nim zanemarivanjem putovanja u susjedne ćelije. Važno je kod definiranja putovanja uzeti u obzir postojanje oscilacija.



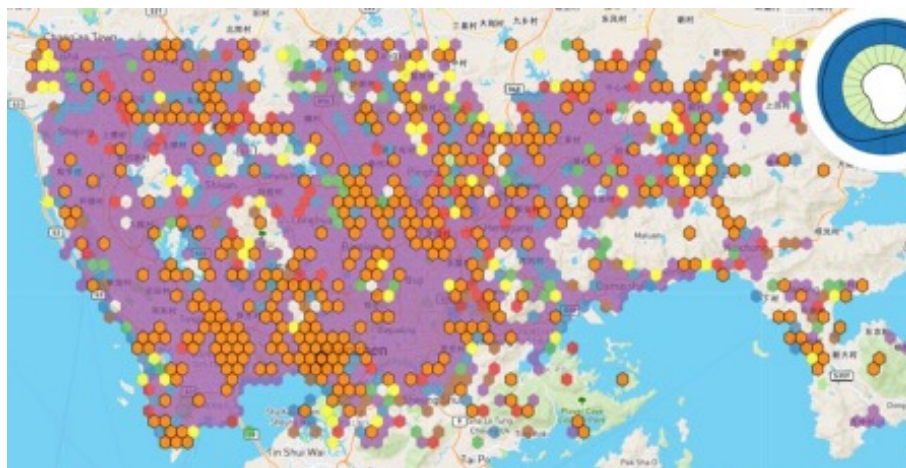
Slika 2.3 Voronoi dijagram 1090 ćelija baznih stanica - Shenzhen

### Kvadratna mreža i mreža šesterokuta

Uniformna podjela na kvadratne blokove od 1 kvadratni kilometar spominje se kao dobra praksa u britanskom priručniku za izradu POM-a iz CDR [4].

Područje pokrivanja bazne stanice na grafikama se standardno prikazuje kao šesterokut (heksagon). Terralytics u praksi [23] između ostalih podjela koristi i šesterokute koje po potrebi ugnježđuje i tako mijenja rezoluciju u kojoj promatra gibanja. Tvrde da su se u nekim segmentima pokazali i 13% bolji od kvadratne mreže. Prikaz šesterokuta za grad Shenzhen možemo vidjeti na slici 2.4

## Poglavlje 2. Polazišno-odredišna matrica



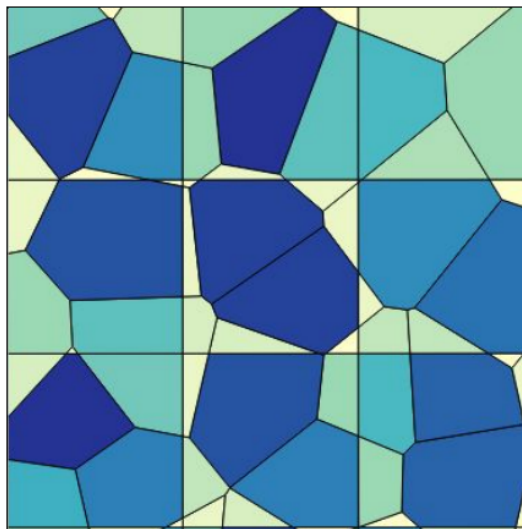
Slika 2.4 Podjela na šesterokute Shenzhen - grupacija u različite boje na osnovu širine toka [3]

### Interpolacija i usporedba

Postupak konverzije broja putovanja iz prostorne podjele na Voronoi ćelije u drugu prostornu podjelu, npr. uniformnu kvadratnu mrežu (ćelije  $1\text{km}^2$ ), opisan u [4]. 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 2.5 je prikazan treći oblik interpolacije gdje su korišteni podaci o ukupnoj kvadratnoj površini zgrada.

### Redefiniranje prostorne podjele na temelju Polazišno-Odredišnih Matrica

U sklopu projekta Data for Development (D4D) francuski operator *Orange* ustupio je na korištenje CDR višemjesečne zapise s područja Obale Bjelokosti. Različita istraživanja u ovom području rađena su upravo na tim podacima. Jedno od njih [30] ističe da je dobar broj istraživanja dokazao da administrativne regije u zemlji faktor koji definira ljudsku mobilnost i interakcije na području države. Da bi testirali tu tvrdnju oni, nakon što procjene matricu kretanja, ćelije prostorne podjele grupiraju prema *community detection* algoritmu. Dio jedinica preklapa sa postojećim



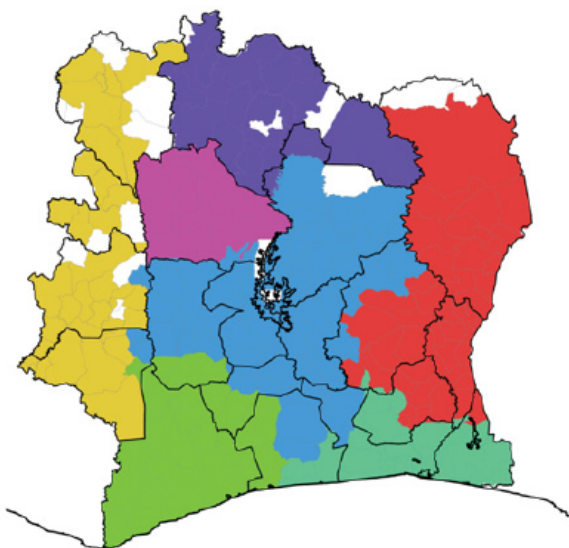
Slika 2.5 Primjer mozaika *krhotina* koje dozvoljavaju interpolaciju između Voronoi ćelija i kvadratne mreže. Nijansa odražava postotak površine zgrada u *krhotini* u odnosu na cijelu Voronoi ćeliju kojoj pripada. Rezultat se dobiva na osnovu tih *težinskih faktora* polazišne i odredišne *krhotine* te pripadajućih ("roditeljskih") Voronoi ćelija. Sumom svih rezultata *krhotina* jedne kvadratne ćelije izračuna se konačni tok. [4]

administrativnim jedinicama. Grupacije nastale na temelju mobilnosti drugačije su kod onih administrativnih jedinica u kojima se nalaze veliki gradovi. Zaključak je da potencijalno postoji potreba da se granice administrativnih jedinica izmjene (Slika 2.6)

Ovo otvara mogućnost analize prometa s dinamičkom prostornom podjelom. Primjerice sezonske izmjene jedinica u razdobljima kada su neka područja zemlje manje aktivnija.

Izmjene na dnevnoj bazi spominju se u [22] gdje su dva tornja spojena u jedan ako je tokom čitavo dana zabilježeno više od 3 prespajanja između njih. (Također uklanjaju se tornjevi na kojima su zabilježena sam tranzitna zaustavljanja, a takvih je manje od 0.5 posto tijekom ukupnog vremena promatranja od 154 dana).





Slika 2.6 Prijedlog prilagodbe administrativnih jedinica mreži kretanja - bojom su označene grupacije nastale na temelju POM-e, crnom su označene postojeće granice administrativnih jedinica [5]

## 2.7 Vremenska interpretacija Polazišno-Odredišnih Matrica

### 2.7.1 Vremenski zavisne Polazišno-Odredišne Matrice

Procjena POM-a može biti vremenski nezavisna ili vremenski zavisna. [16] Vremenski nezavisna POM-a prikazuje ukupan broj putovanja za promatrano razdoblje u cijelosti i nije poznato kada se putovanja odvijaju. Prilikom vremenski zavisne (dinamičke) procjene POM-a promatrano vrijeme dijeli se na vremenske okvire, a rezultat je po jedna matrica za svaki vremenski okvir. U vremenski zavisnu procjenu ubraja se i procijenjena u stvarnom vremenu. Sposobnost vremenski zavisnih modela procjene (i matrica kao rezultat) da uhvate razlike u dnevnim, tjednim, mjesečnim, sezonskim kretanjima je od izuzetne vrijednosti za planiranje prometa, ITS... Tradicionalne metode procjene nisu pogodne za procjenu vremenski zavisnih POM-a visoke rezolucije (npr. vremenski okvir 3 sata) i dugog vremena proma-

tranja (u mjesecima). Baze CDR iz kojih su procjenjivane POM-e u dosadašnjim istraživanjima obuhvaćaju podatke za razdoblja promatranja od nekoliko mjeseci te uključuju različite nacionalne praznike, meteorološke uvjete, značajne sportske događaje i slične faktore koji mogu utjecati na pojavu iznimka u dnevnim uzorcima kretanja.

### Dani u tjednu i praznici

Istraživanja modernih pristupa procjeni POM-a ukazuju na smanjenu mobilnost za neradnih dana u tjednu, najviše u nedjelju. [20] Dio istraživanja ističe da je vikendima također zabilježen i manji broj telekomunikacijske aktivnosti. [7] Osim očekivane razlike između radnih i neradnih dana, uočeno je da je petak, kao zadnji radni dan u tjednu, drugačiji od ostalih radnih dana. [32] [20] [5]. Na temelju POM-a izrađenih iz CDR uočena je pojačana mobilnost petkom na razini regije (Boston Metropolitan Area) [20] i na razini cijele države (Obale Bjelokosti) s povećanjem od 35% u odnosu na nedjelju [5]. U drugom istraživanju 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). Ponovno, petkom je također uočena i najviša razina telekomunikacijske aktivnosti [7] [15] što ukazuje na moguću pristranost kada su u pitanju POM-e procijenjene iz CDR. 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 jest u skladu sa zaključcima modernih postupaka.[20]

Državni praznici i izvanredni događaji kao što su vremenske nepogode unose odstupanja od ustaljenih uzorka kretanja. Privremene turističke atrakcije i sportski događaji mogu također bitno utjecati na izmjenu uobičajenog uzorka kretanja. [15] Primjerice odbrojavanje uoči Nove godine na New Yourk Times Square. U prosjeku je na festivalske dane gustoća poziva 10-15% veća od uobičajene. [33]

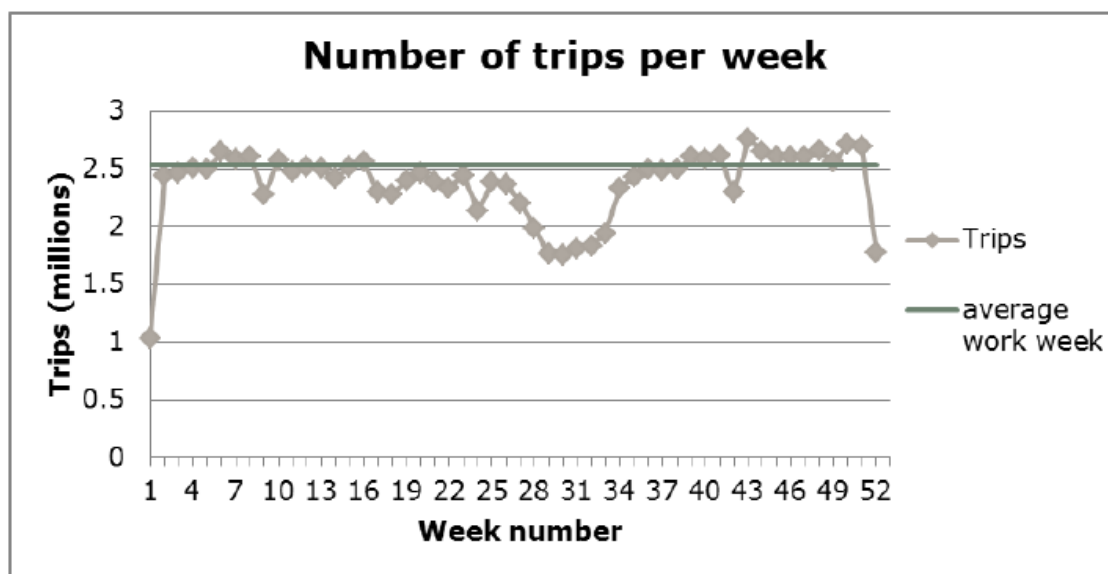
Za kraj valja spomenuti da je prilikom analize zapisa u sklopu projekta D4D operatora Orange je u kasno-večernjim satima bio zamijećen porast aktivnosti [9].

## Poglavlje 2. Polazišno-odredišna matrica

Kao uzrok porasta sumnja se na određene popuste koje moguće nudi operator u tim satima. [?] Ovo otvara teoretsku mogućnost da se za potrebe istraživanja na određeni ciljani period potpuno ukine naplata usluge korisnicima i razmotri kakav učinak to ima na broj i gustoću aktivnosti.

### Sezonske razlike

Velika prednost novih postupaka u odnosu na tradicionalne je da oni mogu uhvatiti tjedne i sezonske uzorke [20] Istraživanje iz 2013. [33] pokazalo je da su u Obali Bjelokosti najveća odstupanja u gustoći poziva u prvom i posljednjem mjesecu godine, gdje je 2. tjedan u prosincu bila najveća, a sredinom siječnja najniža zabilježena gustoća (u periodu promatranja od prosinca do travnja). Zabilježenu varijaciju po tjednima u godini iz jednog istraživanja procjene POM-a iz podataka o javnom prijevozu na području nizozemske možemo vidjeti na slici 2.7.



Slika 2.7 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 [6]

## 2.8 Diskusija

Novi moderni postupci procjene POM-a postaju sve više razmatrana opcija s obzirom na fleksibilnost, cijenu i ažurnosti koju nude u odnosu na tradicionalne. Prepoznat je potencijal ne samo za analizu urbanih i državnih razmjera već i za analizu mobilnosti planetarnog razmjera dijeljenjem i kombiniranjem različitih izvora podataka koji nose podatak o položaju [5][34], gdje se CDR ističe kao izvor gotovo jednako zastupljen u zemljama u razvoju i razvijenim zemljama, s penetracijom oko 40%-50% stanovništva za pojedine operatore [15]. U Francuskoj 2008. 80% stanovništva starijeg od 12 godina posjedovalo je mobilni uređaj [15]. Istraživanje iz 2016. provedeno na podacima u Čileu iznosi podatak da Čile ima čak 132 mobilne pretplate na 100 stanovnika [10].

Važno je osigurati anonimnost pojedinaca čiji se "tragovi" analiziraju i/ili osigurati pristanak pojedinaca da se podaci koriste u jasno definirane svrhe. Idealno je da operator ustupi već anonimizirane podatke stranci koja iz njih procjenjuje POM-e. Kako bi dodatno zaštitili identitet pojedinaca, dobra praksa je osigurati i  $k$ -anonimnost (engl. *k-anonymity*).  $K$  predstavlja najmanju dozvoljenu širinu toka koja se dozvoljava, u suprotnom se tok u potpunosti zanemaruje. Razlog je zaštita od eventualne mogućnosti identificiranja kretanja pojedinaca u vrlo rijetko naseljenim područjima. [4] Teralytics predlaže odabrati  $k$  u rasponu od 5 - 10 [23].

Jedna od malobrojnih kritika koju ističu zagovaratelji očuvanja statusa tradicionalnih postupaka u modeliranju prometa je činjenica da je u pojedinim zemljama zabranjeno ili nije preporučeno služiti se mobitelima tijekom upravljanja vozilom.

## 2.9 Alternative izvora prostornih podataka za razvoj Polazišno-Odredišne Matrice

Ako su to tradicionalne vs telekom POM, onda bi i za jedne i za druge, u posebnim poglavljima, trebalo opisati postupke prikupljanja i **vrste ulaznih podataka**. **Postoje li i treće vrste POM, izvedene iz taxi i lorry GPS podataka? Ili su one integrirane u telekom**, uvažavajući korištenje eCall-a, telekom usluga koje prikupljaju LBS podatke, 5G i IoT?

Predlažem izmjenu naslova 2.3. "...zapisi o telekomunikacijskim aktivnostima u javnoj pokretnoj mreži" u naslov koji bi bio prikladniji CDR i isključio ostale navedene izvore (*integrirane u telekom*)

GPS podatke bih svakako odvojila jer se u konačnici POM-e mogu razlikovati po prostornoj podjeli i rezoluciji te biti (nepotrebno?) zrnati.

Podjelu svih izvora koji uključuju automatsko prikupljanje (a prvenstveno se ne prikupljaju s ciljem izrade POM-a) trebalo bi možda napraviti po preciznosti određivanja položaja koju je moguće postići.

Nekadašnji nacrt poglavlja:

2.7 Drugi primjeri automatskog prikupljanja

2.7.1 Dodatni telekomunikacijski podatci

2.7.2 Združena očitavanja prijamnika za satelitsku navigaciju

2.7.3 Javni prijevoz i pametne kartice (Smart Card sustavi)

### 2.9.1 Združena očitavanja prijamnika za satelitsku navigaciju

(...)

komercijalna vozila, Taxi [35]

### 2.9.2 Javni prijevoz i *pametne kartice* (*Smart Card* sustavi)

(..)

82% putovanja javnim prijevozom naprave korisnici javnog prijevoza sa pametnim

## *Poglavlje 2. Polazišno-odredišna matrica*

karticama. [32]

- [6] [36]

### **2.9.3 WiFi**

(...)

- ITrack.CTrack,VTrack

## Poglavlje 3

# Validacija Polazišno-Odredišne Matrice

Nekadašnji naslov "Postojeće metrike za validaciju POM-a"

Postupak validacije je objektivna procjena kvalitete, vrijednosti objekta koji se validira. Validacija 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 kao što su: točnost, *prostorno obuhvaćanje*, *zrnatost*, definicija putovanja, prostorna rezolucija, vremenska rezolucija i širina toka.

\*\* Pojedinačno navesti *metrike za određivanje vrijednosti svakog od tih indikatora*  
\*\*\*

Transkript: *...linearna regresija (neka bude) u svakom slučaju za referentni model s kojim uspoređujete ostale.*

### 3.1 Točnost Polazišno-Odredišne Matrice

Točnost procijenjenih POM-a gotovo uvijek se definira u odnosu na referentnu matricu (eng. *grand truth matrix*) koja je dobivena tradicionalnim postupcima (anketiranje i/ili prebrojavanje vozila). Statističke mjere kvantiziraju razliku procijenjenih

i “istinitih” vrijednosti, ako su nam one poznate.

Često se u literaturi (jednoznačno) koriste pojmovi *točnost*, *pouzdanost* i *kvaliteta*. Za njihovu procjenu koriste se mjere koje opisuju razinu sličnosti odnosno razlike (greška) s referentnom matricom. **Tradicionalni postupak validacije POM-e svodi se na određivanje njene točnosti u odnosu na *grand truth* matricu.**

### 3.1.1 Metrike za procjenu točnosti Polazišno-Odredišnih Matrica

Za procjenu kvalitete POM-a dobivenih isključivo anketranjem u radu [12] korištena je mjera Mean Absolute Percentage Error (MAPE), te je prikazano da se zadovoljavajuća razina kvalitete takvih POM-a postiže tek ako uzorak obuhvaća 50% populacije. Istaknuta je važnost korištenja dodatnih izvora za izradu POM-a.

U radu [37] navedene su statističke mjere Relative Error (RE), Total Demand Deviation (TDD), Mean Absolute Error (MAE), Root Mean Square Error (RMSE) te Maximum Possible Relative Error (MPRE) i 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). [8]

U [18] 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 HW i WH POM-e dobivene iz CDR u usporedbi s referentnim POM-ma dobivenim anketiranjem.

Travassoli u svome radu [32] navodi nekoliko uobičajeno korištenih mjera -  $R^2$ , Geoffrey E. Havers statistics (GEH), 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 metric, mjeru koja se razlikuje po tome da ne uspoređuje samo vrijednosti parova istih ćelija (elementwise).

Spearmanov koeficijent korelacije ranga korišten je u [10] za procjenu sličnosti

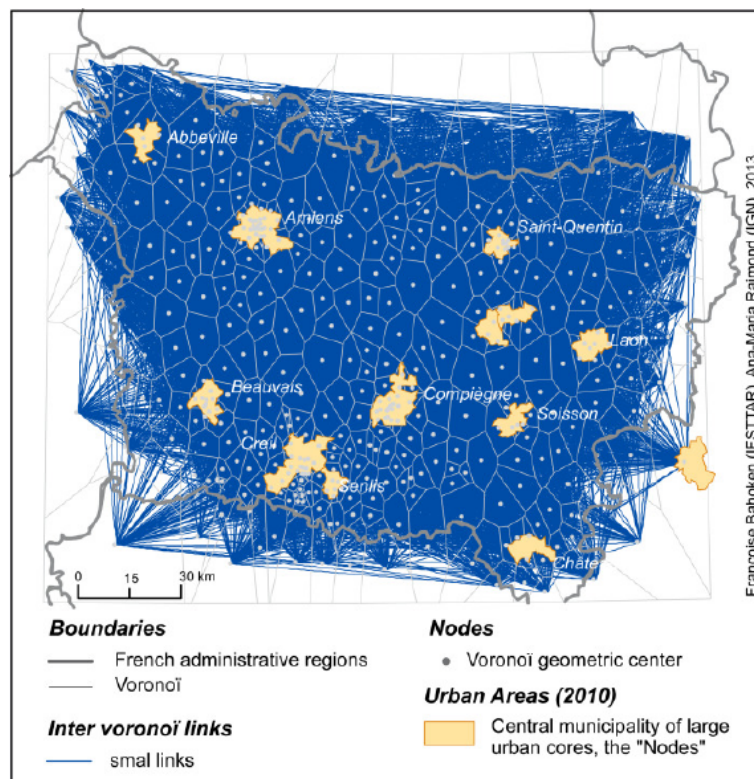


POM-a dobivenih iz CDR sa tada aktualnim POM-ma dobivenim anketiranjem.

## 3.2 Grafički prikaz (oblik) Polazišno-Odredišne Matrice

Na slici 3.1 može se vidjeti jedan oblik grafičkog prikaza POM-a, odnosno njenih tokova. Svaki par centara ćelija spojen je ravnom linijom čija debljina odgovara širini toka između te dvije linije. Međutim, riječ je o primjeru kada ovaj oblik prikaza nikako nije primjeren, zbog visoke prostorne rezolucije (velikog broja zona) kao i velikog broja zabilježenih tokova. Dodatno, da bi se u ovom obliku prikazali svi podaci iz POM-e, potrebno je prikazati jednu sliku za ulazne te jednu za sve izlazne tokove.

Postoji i još jedan često korišten grafički oblik POM-e. Vrijednosti u matrici mogu se normalizirati na opseg vrijednosti koji je moguće prikazati u obliku slike tako da svaka vrijednost predstavlja vrijednost piksela. Koriste se različite palete, od raspona 0-100 koji predlažu [25], prikazom nijansama jedne boje 0-255 engl. *grayscale* [38], paletom nijansa 2 boje, ili kojom drugom proizvoljnom paletom s više boja (vidi paletu boja na slici 4.1).



Slika 3.1 *Spaghetti-effect* - problem koji se može javiti kod grafičkog prikaza tokova POM-e visoke prostorne i niske vremenske rezolucije. Na slici su prikazani jednodnevni tokovi. [7]

### 3.3 Strukturalna sličnost Polazišno-Odredišnih Matrica

Dosada spomenute mjere neće uhvatiti strukturalnu sličnost POM-a. 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.**

### 3.3.1 MSSIM

MSSIM dolazi iz područja računalne obrade slike i koristi se kao mjera usporedbe digitalnih slika (*eng. measure of comparison*). U prometu ideja o korištenju MSSIM za mjerenje sličnosti POM-a prvi se puta spominje i demonstrira na simuliranim POM-ma dobivenim iz referentne POM-e dodavanjem šuma. [8]

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.

SSIM bazira se 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

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  $\sigma_{xy}$

$$l(x, y) = (2\mu_x\mu_y + C1)/(\mu_x^2 + \mu_y^2 + C1) \quad (3.1)$$

$$c(x, y) = (2\sigma_x\sigma_y + C2)/(\sigma_x^2 + \sigma_y^2 + C2) \quad (3.2)$$

$$s(x, y) = (\sigma_{xy} + C3)/(\sigma_x\sigma_y + C3) \quad (3.3)$$

$l(x, y)$  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

$$SSIM(x, y) = [l(x, y)^\alpha][c(x, y)^\beta][s(x, y)^\gamma] \quad (3.4)$$

### Poglavlje 3. Validacija Polazišno-Odredišne Matrice

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

$$SSIM(x, y) \leq 1 \quad (3.5a)$$

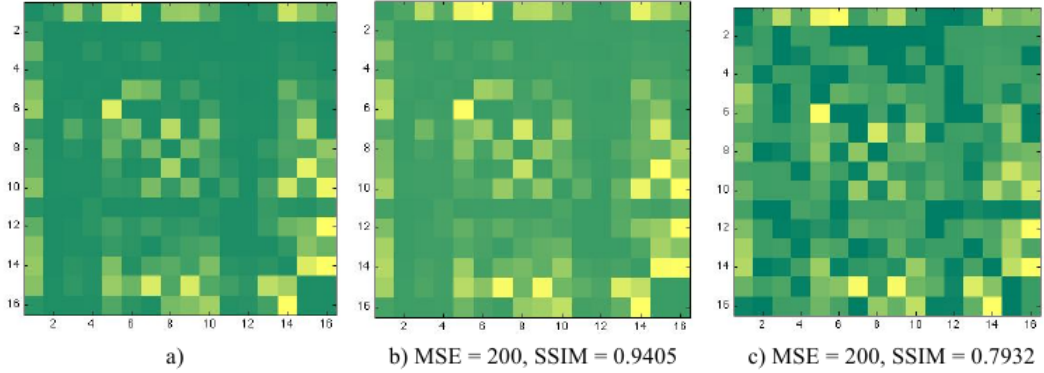
$$SSIM(x, y) = SSIM(y, x) \quad (3.5b)$$

$$SSIM(x, y) = 1 \iff x = y \quad (3.5c)$$

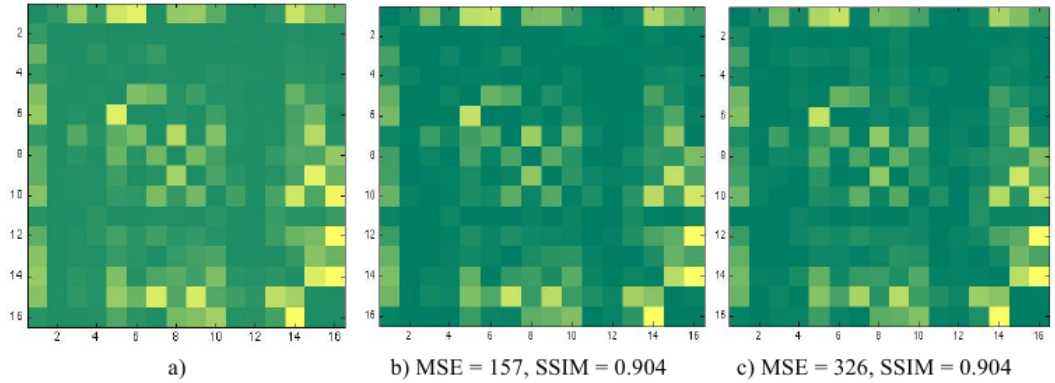
$$MSSIM(X, Y) = \frac{1}{M} \sum_{m=1}^M SSIM(x_m, y_m) \quad (3.6)$$

#### Poboljšani

Nekoliko godina nakon prvog spominjanja  $MSSIM$  kao metrike usporedbe POM-a [38] 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  $4D-MSSIM$  gdje u izračun dodaju stvarne euklidske udaljenosti prostornih zona.



- (a) U odnosu na referentnu *grand 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 (3.5). 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 validacije POM-e [8]

## Poglavlje 4

# Usporedba dviju Polazišno-Odredišnih Matrica

### 4.1 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
- ako se radi o jednom danu - idealno isti dan u tjednu
- ista kategorija putovanja - tranzitna ili *commuting*
- obje POM-e napravljene objektivno na cijeloj populaciji (reprezentativnom uzorku)

### 4.2 Kriteriji usporedbe

stari naslov Komparacijski Indikatori, stariji naslov Odnosni parametri kvalitete

## Poglavlje 4. Usporedba dviju Polazišno-Odredišnih Matrica

*Iz teksta zadatka: ...U ovom radu potrebno je definirati **odnosne parametre kvalitete POM-e** ...*

### 4.2.1 Definicija putovanja

Kod procjene POM-a treba definirati hoće li se putovanja koja se protežu kroz više perioda dodijeliti vremenskom okviru u kojem započinju ili u kojem završavaju. Iako ponekad nije specificirano, autori [37] [25] [9] sva putovanja dodjeljuju intervalu u kojem je putovanje započeto. Teralytics u praksi procjenjuje obje vrste POM-a [23]. Autori [9] kako bi uvrstili što više putovanja, ne zanemaruju one zapise iz kojih nije jasno vidljiv početak putovanja već takva putovanja dodjeljuje na osnovu vjerojatnosti da su započela u određenom periodu. Sličan postupak spominje se i u [21] gdje za izračun vjerojatnosti posežu za nacionalnim anketama ili frekvencijom poziva.

Usporedba POM-a iz različitih izvora (ankete i CDR) s 2 različite definicije putovanja (samo putovanja HW i WH te sva putovanja) pokazala se kao loš pristup, POM-e su imale mali stupanj sličnosti i strukturalna sličnost bila je vrlo mala. [15] Isti zaključak kasnije iznosi i [9] čije grafičke prikaze POM-a možemo vidjeti na slici 4.1 . **O definiciji putovanja ovisit će njihov konačan broj, odnosno širina toka.**

Već je spomenuta podjela na POM-e koje sadrže isključivo HW i WH putovanja i one koje u obzir uzimaju i ostala putovanja, bilo pod kategorijom *Ostalo* (engl. *Other*), preciznijom podjelom ili bez posebne podjele po svrsi. Prva vrsta obično je povezana s tradicionalnim postupcima procjene (osobito ankete) dok su ostale obično prisutne u vremenski zavisnim POM-ma, najčešće procijenjenim modernim postupcima.

1. *commuting trips* - isključivo Home-Work i Work-Home putovanja
2. *general trips* -bilo koje kretanje koje zadovoljava uvjete putovanja

#### Poglavlje 4. Usporedba dviju Polazišno-Odredišnih Matrica

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
5. putovanja ograničena minimalnom duljinom trajanja zaustavljanja (\*)
6. putovanja ograničena minimalnom i/ili maksimalnom duljinom trajanja
7. putovanja ograničena bojem događaja koji potvrđuju zaustavljanje i maksimalnim vremenskim razmakom između dva uzastopna događaja

**Glavni razlog svih ograničenja kod definicije putovanja je izbjeći bilježenje lažnog kretanja uzrokovanog *load balancing*-om i lažnog zaustavljanja uzrokovanog neuniformnom gustoćom zapisa pojedinca u vremenu.**

U istraživanju iz 2016. kod definiranja putovanja iz CDR eliminirani su zapisi onih korisnika koju su radili dulja "neprekidna putovanja" od preko 100 km dnevno (kategorija 4) pod napomenom da se radi o taksistima i sl. te da ih neće uzeti u obzir [10].

Za 2 različite definicije putovanja, za POM-e iz identičnih podataka na području Snegala generirano je 0.7 putovanja dnevno po korisniku u jednom, a 0.06 u drugom. U potonijoj definiciji ograničenja su bila: vremensko na 1 sat minimalnog trajanja (kategorija 6) i prostorno na udaljenost od barem 5 udaljenosti baznih stanica (kategorija 4). Pokazuje da je najveća zastupljenost upravo kratkih putovanja. [9] Distribuciju možemo vidjeti na slici 4.2

*Commuting* putovanja također mogu biti pod nekim ograničenjima, osim što su



najčešće već ograničena vremenskim periodom u kojem se odvijaju. Primjerice [9] definira minimalnu udaljenost od 3 km između prebivališta i mjesta rada (kategorija 4) i izvan susjednih ćelija (kategorija 2).

#### 4.2.2 Prostorna razlučivost (rezolucija)

Prema hrvatskoj enciklopediji definicija razlučivosti (rezolucije) glasi: mjera za razaznavanje sitnih pojedinosti na nekom prikazu (npr. televizijskoj slici). U računalstvu se odnosi na finoću rasterske slike iskazanu ukupnim brojem slikovnih elemenata (relativna razlučivost) ili brojem slikovnih elemenata po inču (stvarna razlučivost).

Kod POM-a razvijenih iz CDR rezolucija će ovisiti o gustoći tornjeva bazni stanica. Gustoća tornjeva snažno je korelirana s gustoćom populacije [9]. Distribucija tornjeva nije homogena, niti na području grada, niti na bilo kojem drugom području. Kod izrazito visoke gustoće [4] predlaže grupiranje aktivnosti tornjeva koji se nalaze na udaljenosti manjoj od 75 m (dodjeljivanje aktivnosti najaktivnijem od tornjeva) kako bi se izbjegla prividna kretanja koja uzrokuje preusmjeravanje na bližnje bazne stanice (engl. *load balancing*). Median udaljenost baznih stanica 2013. na razini francuske pokrajine je 75 km, minimum 600 m i maksimum 208 km [7]. U gradu Santiagu 2016. srednja vrijednost svih udaljenosti polazišta i odredišta svih putovanja je 732 m, minimalna 45 m, maksimalna 14.1 km. Postignuta je zadržavanjem 0.8 kvartil udaljenosti svih zabilježenih kretanja [10]

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

**Definicija 4.2.1.** Interno putovanje je ono putovanje čije se polazište i odredište nalaze unutar iste prostorne zone. Kao dodatno ograničenje može se postaviti da ni putanja putovanja ne izlazi iz prostora te zone - odnosno da se putovanje u cijelosti odvija unutar prostora jedne prostorne zone. Suprotno internom je eksterno putovanje.

Postotak eksternih putovanja koji se gubi u procesu agregacije Voronoi ćelija analiziran je 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. [7]. Sličnu situaciju opisuje i [10] (Vidi sliku 4.3)

Na temelju usporedbi POM-a kretanja u 4 različita grada, dobivenih iz CDR i iz anketa, gdje je svaki grad u anketnim POM-ma imao svoju prostornu podjelu odnosno rezoluciju, Toole u svom radu [21] (očekivano) zaključuje da sa **smanjenjem rezolucije (visokim stupnjem agregacije Voronoi ćelija) stupanj korelacije s POM-ma iz ankete (za isti grad) raste.**

(\*) macro - meso - micro level ? César A. Segovia

### 4.2.3 Vremenska razlučivost

Neki autori predlažu da kraj odnosno početak jednog dana bude u 3:00 kako bi putovanja započeta prethodnog dana, a koja prelaze u novi dan, u analizi spadala u raniji dan, odnosno kako dobar dio putovanja ne bi bio izostavljen samo zato jer se proteže kroz 2 dana [22][21]. Najmanji spominjani interval je 15 min [37], drugi najmanji 30 min [22], kod POM-a kreiranih iz podataka generiranih pametnim karticama u javnom prijevozu također 15 min [32], 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.

### 4.2.4 Širina toka

**Definicija 4.2.2.** Ukupna širina toka podrazumijeva ukupan broj odlazaka/dolazaka po vremenskom okviru za cijelu matricu. Suma svih njenih elemenata.

Na širinu toka utjecat će svi prethodno spomenuti kriteriji usporedbe/parametri kvalitete. Osim njih, utjecat će veličina i reprezentativnost uzorka stanovništva te k-anonimizacija.

## Poglavlje 4. Usporedba dviju Polazišno-Odredišnih Matrica

Neka istraživanja izdvajaju samo najaktivnije korisnike (generalno one s najviše generiranih događaja) za kreiranje POM-a [22] [21].

2015., istraživanje kreiranja POM iz CDR na području jedne francuske regije je podijelilo korisnike u 3 skupine po količini podataka koju generiraju te se pokazalo da frekvencija korištenja nije uvela pristranost [15]. Podsjetimo se ovdje da je u poglavlju 2.7.1 rečeno da broj aktivnih korisnika, a posljedično i širina toka, značajno varira iz dana u dan unutar tjedna. Za razliku od izvlačenja najaktivnijih korisnika [20] uzima samo slučajno odabranih 25% korisnika (njih 1 milijun) da pojednostavi analizu, a [10] samo 100,000, kako bi u konačnici POM bila usporediva s onom dobivenom iz ankete provedene na uzorku iste veličine, što je pomalo apsurdno ograničenje.

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. [15]. [4] ističe da se operatori mogu svoje usluge tržiti različitim demografskim skupinama, a najčešće su izuzetak skupine slabijeg socio-ekonomskog stanja. Kod tržišta gdje veliki postotak stanovništva posjeduje mobilni uređaj (npr. 92% za urbana područja Tanzanije) eventualno postojanje pristranosti može se smatrati zanemarivim.

Nije rijetkost da su POM-e dobivene iz anketa rijetko popunjene, odnosno da se većinski radi o Nul-ćelijama (eng. *Null-cells*, *Zero-cells*). U istraživanju kreiranja POM-a iz ankete, 2010. [12] zasebno je analizirana POM-a sa i bez takvih ćelija jer je od  $589 \times 589$  ćelija, čak 77.8% njih bilo Nul-ćelija. Ista situacija je s POM-ma kreiranim iz RSI, dok će POM-e kreirane iz GPS ili CDR zapisa umjesto rijetko biti gusto ispunjene, najčešće vrlo malim vrijednostima umjesto nula. [38].

### 4.2.5 Gustoća informacija - kontekst

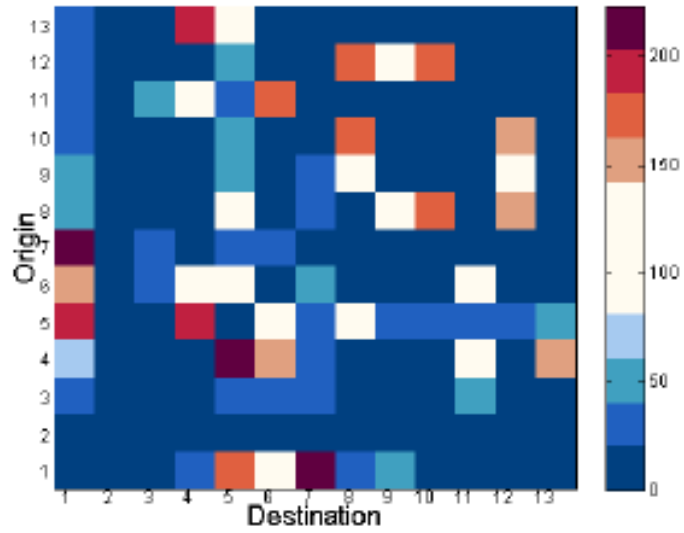
Informacija na razini putovanja - izrazito detaljno

Informacija na razini toka (postotka toka) - zlatna sredina

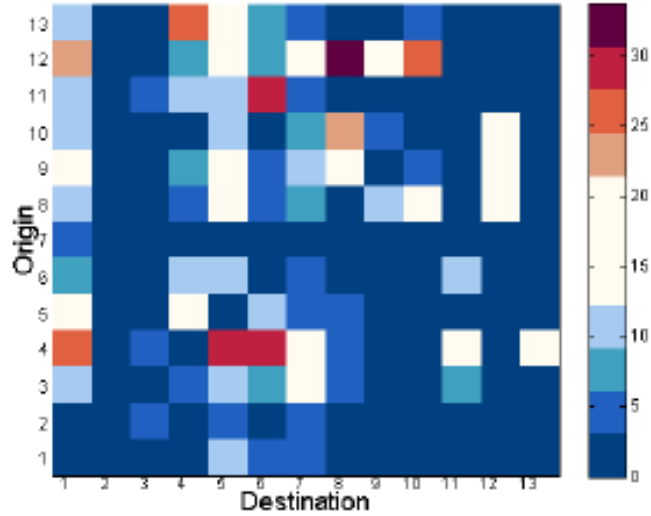
Informacija na razini cijele matrice - "nedovoljno" detaljno

(...)

Poglavlje 4. Usporedba dviju Polazišno-Odredišnih Matrica

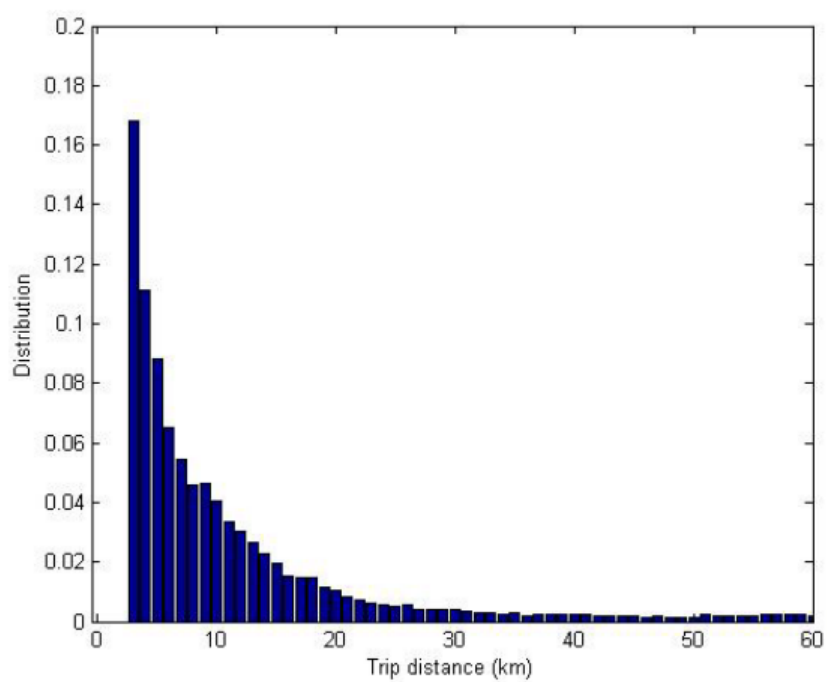


(a) POM-a u kojoj je putovanje definirano kao zabilježeni pomak između 2 antene - *kratka putovanja*



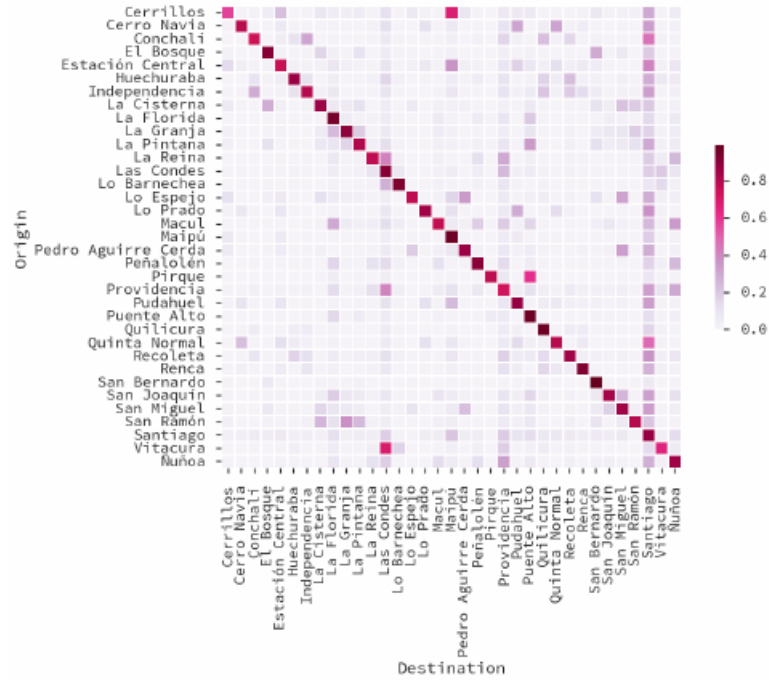
(b) POM-a u kojoj je putovanje definirano kao promjena položaja između 2 prethodno definirana perioda, odnosno između položaja na kojima korisnik pretežito boravi u ta 2 perioda - definiranjem tih perioda pokušavaju se opisati putovanja od prebivališta na posao i s posla nazad na prebivalište engl. *commuting trips*

Slika 4.1 Dvije POM-e - produkti različitih definicija putovanja. Dobivene su iz potpuno istog izvora CDR zapisa. [9]

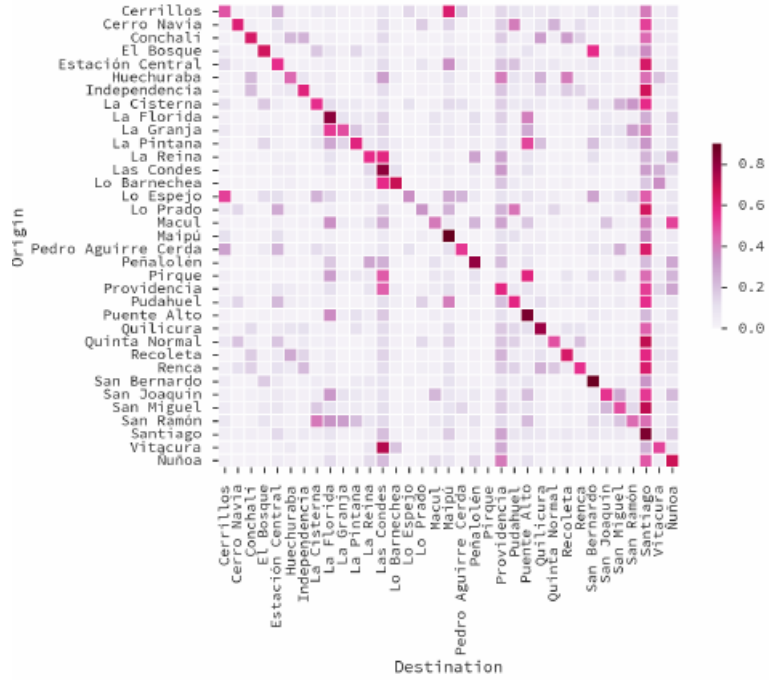


Slika 4.2 Distribucija udaljenosti putovanja u kilometrima [9]

Poglavlje 4. Usporedba dviju Polazišno-Odredišnih Matrica

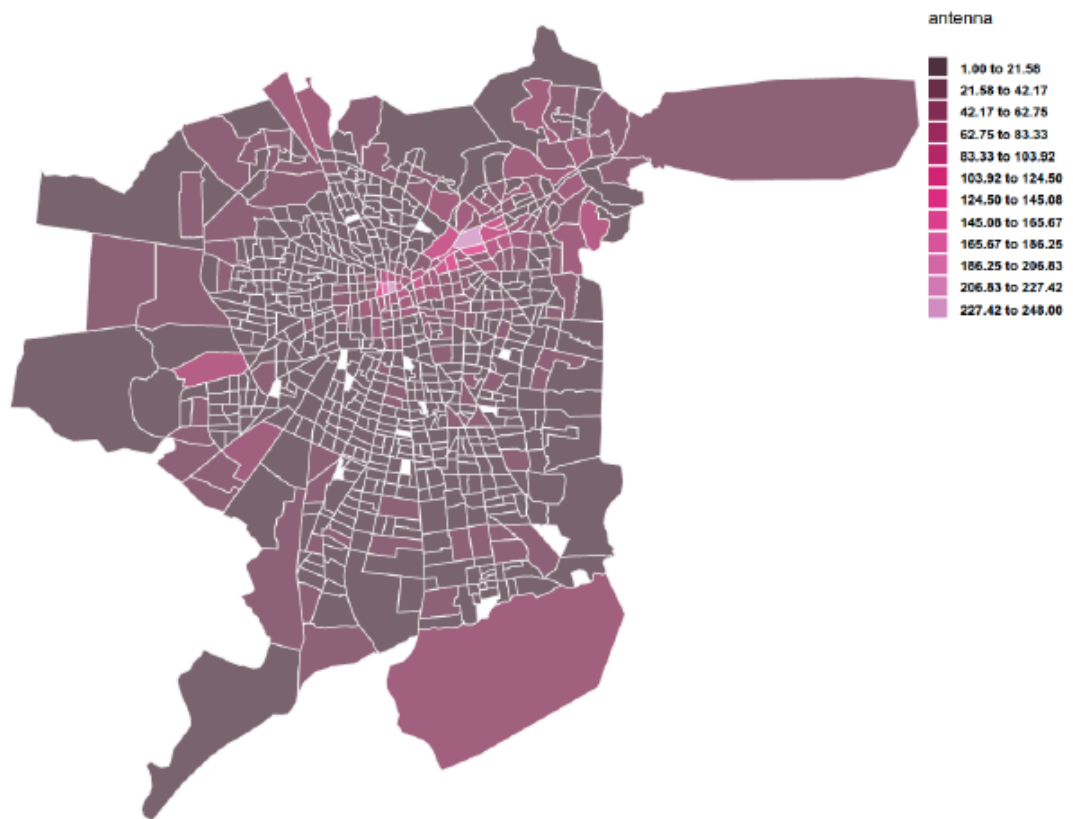


(a) Izvor podataka anketa



(b) Izvor podataka CDR

Slika 4.3 Distribucija CDR putovanja na području Santiago, Čile. POM je normalizirana prema redovima (polazištima) (\*) [10]



Slika 4.4 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. [10]

Poglavlje 4. Usporedba dviju Polazišno-Odredišnih Matrica

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 4.5 Gustoća tornjeva baznih stanica po kvadratnom kilometru za pojedine države i grad-državu Singapur. [11]



## 4.3 Diskusija

(...)

- Ukoliko je rezolucija mala (velike ćelije) nema potrebe za preciznim definiranjem kraja putovanja
- O definiciji putovanja ovisit će i zero-cells ? - [9] str. 9 definicija putovanja i prostorna i vremenska rezolucija, DEFINICIJA PUTOVANJA JE KLJUČNA
- Za analizu Peak hour - intervali od 5 minuta! [9] str. 10
- 30 min vremenski okvir [22] str. 6
- short trips are less likely to use public transport [10]

*stvari koje stvaraju neodređenost jer ne mogu biti objektivno definirane. Sve navedite tu što nije jednoznačno određeno, ima nekakvih preklapanja, što onda izaziva neodređenost.*

**\*\*zasebni kriterij - dodatak\*\***

- takve matrice su subjektivne, "nisu objektivne"
  - faktor skaliranja spominje se u najmanje 2 istraživanja, u jednom se radi o  $\times 10$
- Predlaže skaliranje na osnovu prebrojavanja prometa na ključnim mjestima izvučenim iz postupka kreiranja POM-a prije skaliranja i interpolacije na kvadratnu mrežu - zadnjeg koraka [4].

**Poglavlje 5**

**Zaključak**

# Bibliografija

- [1] J. Xie, Y. Nie, and X. Liu, “Testing the proportionality condition with taxi trajectory data,” *Transportation Research Part B Methodological*, May 2017.
- [2] M. K. Ng and W. S. Tang, “The Role of Planning in the Development of Shenzhen, China: Rhetoric and Realities,” *Eurasian Geography and Economics*, July 2014.
- [3] Z. Zhou, J. Yu, Z. Guo, and Y. Liu, “Visual exploration of urban functions via spatio-temporal taxi OD data,” *Journal of Visual Languages & Computing*, vol. 48, pp. 169–177, October 2018.
- [4] J. Goulding, *Best Practices and Methodology for OD Matrix Creation from CDR data*, N/LAB, University of Nottingham, 2016.
- [5] 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.
- [6] W. Kuhlman, “The conconstruct of purpose specific od matrices using public transport smart card data,” Tech. Rep., 2015.
- [7] 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.
- [8] T. Djukić, “Reliability assessment of dynamic OD estimation methods based on structural similarity index,” *92nd meeting of the Transportation Research Board*, 2013.
- [9] D. Gundlegard, C. Rydergren, N. Breyer, and B. Rajna, “Travel demand estimation and network assignment based on cellular network data,” *COMPUTER COMMUNICATIONS*, 2016.

- [10] E. Graells-Garrido and D. Saez-Trumper, “A Day of Your Days: Estimating Individual Daily Journeys Using Mobile Data to Understand Urban Flow,” 2016.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] A. Peterson, “The Origin-Destination Matrix Estimation Problem- Analysis and Computations,” Ph.D. dissertation, Linköping Studies in Science and Technology, 2007.
- [17] N. Jelusic, “Telematicka sucelja (nastavni tekst),” 2016.
- [18] V. Frías-Martínez, E. Frías-Martínez, and C. S. Ruiz, “Estimation of urban commuting patterns using cellphone network data,” 2012.
- [19] 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.
- [20] 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.
- [21] 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.

## Bibliografija

- [22] 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.
- [23] (03.06.2019) Teralytics. , s Interneta, <https://thenewstack.io/teralytics-takes-big-data-approach-human-movement/>
- [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] 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.
- [26] (03.06.2019) Data description for urbancps. , s Interneta, <https://www-users.cs.umn.edu/~tianhe/BIGDATA/>
- [27] 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.
- [28] Z. Zheng, Z. Haung, F. Zhang, and P. Wang, “Understand coupling dynamic of public transport network,” *EPJ Data Science*, 2018.
- [29] E. Mellegard, “Obtaining origin-destination- matrices from cellular network data,” Master’s thesis, Chalmers University of Technology, Department of Mathematical Sciences, 2011.
- [30] M. Coscia and R. Hausmann, “Evidence that Calls-based Mobility Networks are Isomorphic,” *Working Papers - Center for International Development at Harvard University*, 2015.
- [31] (21.06.2019) Opencellid. , s Interneta, <https://www.opencellid.org/>
- [32] 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.
- [33] J. Varun, H. Sunil, R. Vasisht, S. Saragur, and A. Lele, “On the Analysis and Visualisation of Anonymised Call Detail Records,” 2013.

## Bibliografija

- [34] P. Hui, R. Mortier, M. Piorkowski, T. Henderson, and J. Crowcroft, “Planet-scale human mobility measurement,” in *2ND ACM International Workshop on Hot Topics in Planet-scale Measurement. Hot Planet*, 2010.
- [35] C. Krause and L. Zhang, “Short-term travel behaviour predictions with gps, land use and point of interest data,” *Elsevier*, 2018.
- [36] N. S. Aslam, T. Cheng, and J. Cheshire, “Adata percision heuristic model to detect home and work llocation from smart card data,” *Geo-spatial Information Science*, 2018.
- [37] 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.
- [38] T. van Vuren, “256 shades of gray - comparing OD matrices using image quality assessment tehniques,” *Scottish Transport Applications Research (STAR)*, 2015.

# Pojmovnik

**ANPR** Automatic Number Plate Recognition. 4

**CDR** Call Data Records ili Charging Data Records. x, 6–11, 13, 16–18, 21, 23, 26, 33–38, 40

**D4D** Data for Development. 18, 21

**FCD** Floating Car Data. 5

**FPD** Floating Phone Data (Floating Cellular Data). 6

**GDPR** General Data Protection Regulation. 8

**GIS** Geoprostorni Informacijski Sustav. 3, 11

**GNSS** Global Navigation Satellite System. 5–7

**GPS** Global Positioning System. 11, 16

**GSM** Global System for Mobile (Communications). 6, 7

**HW** Home-Work. 9, 26, 33

**ITS** Inteligentnim Transportnim Sustavima (engl. *Intelligent Transportation Systems*). 4, 5, 20

**MSSIM** Mean Structural Similarity index. x, 29, 31

**ODM** Origin-destination Martix. 1

**OSM** OpenStreetMap. 11, 12, 15

**POIs** Points of Interest. 10, 13

**POM** Polazišno-Odredišna Matrica. vi, vii, 1–10, 13, 16, 20, 21, 23, 25–43

**RSI** Road Side Interview. 3, 37

**SSIM** Structural Similarity index. x, 29, 31

**TAZ** Traffic Analysis Zone. ix, 13–15

**WH** Work-Home. 9, 26, 33



# 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

## Naslov priloga

A.1 Naslov sekcije

A.2 Naslov sekcije