



ROMÂNIA

Acord de servicii de asistență tehnică rambursabile privind Consolidarea Sistemului Statistic Național (P167217)

REZULTAT Nr. 7c

Raport privind serviciile de asistență tehnică oferite Beneficiarului privind Recomandări și bune practici pentru implementarea unui sistem de management al datelor pentru datele geospațiale și sectorizarea efectivă din cadrul RPL2021

August 2021

Versiune revizuită noiembrie 2022



*Proiect cofinanțat din Fondul Social European prin Programul Operațional
Capacitate Administrativă 2014-2020*

Declinarea responsabilității

Acest raport este un produs elaborat de personalul Băncii Mondiale. Constatările, interpretările și concluziile exprimate în acest document nu reflectă neapărat punctele de vedere ale Directorilor Executivi ai Băncii Mondiale sau ale guvernelor pe care aceștia le reprezintă. Banca Mondială nu garantează acuratețea datelor incluse în prezentul document și nu își asumă responsabilitatea pentru orice erori, omisiuni sau inconsecvențe în ceea ce privește informațiile sau răspunderea cu privire la utilizarea sau neutilizarea informațiilor, metodelor, proceselor sau concluziilor prezentate. Granițele, culorile, denumirile și alte informații afișate pe orice hartă din acest document nu implică nicio apreciere din partea Băncii Mondiale cu privire la statutul juridic al oricărui teritoriu sau aprobarea sau acceptarea acestor granițe. Prezentul raport nu reprezintă neapărat poziția Uniunii Europene sau a Guvernului României.

Declarația privind drepturile de autor

Materialul din această publicație este protejat de drepturi de autor. Copierea și/sau transmiterea unor porțiuni din această lucrare fără permisiune poate constitui o încălcare a legilor aplicabile.

Pentru permisiunea de a fotocopia sau imprima orice parte din această lucrare, vă rugăm să trimiteți o solicitare cu informațiile complete către: (i) Institutul Național de Statistică din România (Bd. Libertății nr. 16, Sector 5, București, România) sau (ii) Grupul Băncii Mondiale România (str. Vasile Lascăr nr. 31, etaj 6, București, România).

Prezentul raport a fost predat în luna august 2021 și revizuit în luna noiembrie 2022 în cadrul Acordului de servicii de asistență tehnică rambursabile privind Consolidarea Sistemului Statistic Național (P167217), semnat între Institutul Național de Statistică din România și Banca Internațională pentru Reconstrucție și Dezvoltare la 17 septembrie 2019. Raportul face parte din Rezultatul 7 din acordul menționat mai sus.

Cuprins

Rezumat	6
1. Contextul sectorizării în RPL2021	8
1.1. Obiectivele sectorizării	9
1.2. Principiile sectorizării	10
2. Metodologia de sectorizare.....	11
2.1. Model	11
2.2. Informații necesare.....	12
3. Pregătirea datelor	13
3.1. Surse de date de intrare	13
3.2. Pregătirea datelor de intrare	14
3.3. Tratarea datelor.....	14
3.4. Eșantionul de localități.....	15
3.5. Exemple de localități propuse de INS	16
3.6. Localități rurale	16
3.7. Localități urbane.....	18
3.8. Zona rurală – localități de mari dimensiuni	21
4. Rezultate	22
4.1. Eșantionul rural.....	22
4.2. Eșantionul urban	25
4.3. Eșantion urban – număr mare de localități cu locuințe.....	31
4.4. Observații	33
5. Recomandări	36
5.1. Date de intrare	36
5.2. Metode	37
6. Implementare	38
6.1. Pornire	38
6.2. Etape suplimentare și specifice pentru tratarea datelor și model	39
7. Managementul datelor geospațiale în contextul RPL2021	42
7.1. Utilizarea datelor geospațiale în RPL2021	43
7.2. Armonizarea nomenclurii amplasărilor și străzilor.....	44
Anexe	46
7.3. Cod de segmentare.....	46
7.4. Anomalii WebGIS.....	46
7.4.1. Date complete	46
7.4.2. Actualitate și acoperire	46
7.5. Cadrul de informații geospațiale integrate.....	47
7.6. Cod pentru managementul datelor	49
7.7. Referințe.....	128

Figuri

Figura 1: Harta localităților cu zone rurale și delimitări (2019)	13
Figura 2: Harta unei localități cu zone nealăturate	13
Figura 3: Harta subeșantionului prezintă populația în funcție de localitate (stânga) și amplasările (dreapta)	15
Figura 4: Hărțile (panou A-D) ilustrează principalele etape de tratare a datelor pentru modelul din Sâncrăieni	17
Figura 5: Hărțile (panou A-D) ilustrează principalele etape de tratare a datelor pentru modelul din Perieți	18
Figura 6: Hărțile (panou A-D) ilustrează principalele etape de tratare a datelor pentru modelul din Onești	19
Figura 7: Hărțile (panou A-D) ilustrează principalele etape de tratare a datelor pentru modelul din Câmpia Turzii	20
Figura 8: Histograma numărului de locuințe dintr-o localitate pentru eșantionul urban de localități de mari dimensiuni	21
Figura 9: Străzile din Cluj-Napoca (stânga) sunt tratate pentru a conecta toate străzile (partea dreaptă)	22
Figura 10: Aceste histograme ilustrează diferența medie absolută (ponderea numărului de locuințe) cu o țintă de 100 (stânga) și 600 (dreapta)	22
Figura 11: Hărțile (panou A-B) ilustrează rezultatul modelului (A) și sectorului (B) din Sâncrăieni	23
Figura 12: Hărțile (panou A-B) ilustrează rezultatul modelului (A) și sectorului (B) din Perieți	24
Figura 13: Histograma afișează numărul de locuințe dintr-un sector situat peste și sub dimensiunea țintă în Sâncrăieni (stânga) și Perieți (dreapta)	24
Figura 14: Hărțile (panou A-B) ilustrează rezultatul modelului (A) și sectorului (B) din Onești	26
Figura 15: Hărțile (panou A-B) ilustrează rezultatul modelului (A) și sectorului (B) din Câmpia Turzii	27
Figura 16: Histograma afișează numărul de locuințe dintr-un sector care include o amplasare peste și sub dimensiunea țintă în Onești (stânga) și Câmpia Turzii (dreapta)	28
Figura 17: Hărțile (panou A-B) ilustrează modelul cu dimensiunea țintă de 600 (A) și sectoarele (B) din Onești	29
Figura 18: Hărțile (panou A-B) ilustrează modelul cu dimensiunea țintă de 600 (A) și sectoarele (B) din Câmpia Turzii	30
Figura 19: Histograma afișează numărul de locuințe dintr-un sector situat peste și sub dimensiunea țintă de 600 în Onești (stânga) și Câmpia Turzii (dreapta)	31
Figura 20: Numărul de locuințe din sector pentru Cluj-Napoca	32
Figura 21: Imagini din satelit care arată amplasarea clădirilor	35
Figura 22: Diagrama tratării locuințelor	40
Figura 23: Diagrama tratării străzii cu locuințe	40
Figura 24: Diagrama tratării zonei cu localități	40
Figura 25: Diagrama pașilor de model	41
Figura 26: Cele trei etape principale ale unui recensământ	43
Figura 27: Structura imbricată a unităților de recensământ	45
Figura 28: Harta limitei unui grup mic (linia albastră) din limitele tuturor clusterelor (linia galbenă) care are o stradă în imagini (săgeata roșie) care nu este prezentă în străzile din (linia violetă) baza de date webgis. Punctele reprezintă amplasarea cu locuințe cu numărul de locuințe etichetat. ...	47

Listă de acronime

GIS	Sistemul Informațional Geografic
INS	Institutul Național de Statistică
RPL2021	Recensământul Populației și Locuințelor 2021
RAS	Servicii de Asistență Tehnică Rambursabile
SuSo	Survey Solutions
WebGIS	Aplicație software de localizare geografică

Rezumat

Scopul prezentului raport este în primul rând să prezinte metodologia de sectorizare pentru RPL2021 în contextul pregătirii recensământului pentru etapa de derulare și producție din 2022 (martie-iunie), în baza cadrului legal european și național pentru runda 2021. Acesta face parte din livrabilele din cadrul Acordului de servicii de asistență tehnică rambursabile (RAS) privind *Consolidarea Sistemului Statistic Național (P167217)*. Proiectul este implementat de Institutul Național de Statistică (INS) cu sprijinul Băncii Mondiale (BM).

Acest raport are legătură directă cu rapoartele livrate în cadrul aceluiași Proiect și a fost elaborat în urma discuțiilor cu conducerea și cu personalul INS, inclusiv cu Președintele și cu cei care reprezintă actorii principali în organizarea și implementarea recensământului din 2022.

Sectorizarea în contextul unui recensământ al populației și al locuințelor se referă la procesul de creare a unor sectoare ale zonei de recensământ, cu un număr similar de unități de recensământ (și anume, gospodării, locuințe) în fiecare dintre aceste sectoare. Pentru RPL2021, aceasta înseamnă că întreaga zonă de recensământ, care este definită plecând de la ***un cadru al zonei intravilane, excluzând zona extravilană***, trebuie să fie segmentată în zone geografice exclusiv reciproc, cu un număr țintă aproximativ de unități de **locuințe** (și anume, de 100 sau 600), cu o abatere dorită de +/- 20%. Elementele esențiale în acest proces sunt reprezentate de calitatea datelor de intrare privind amplasările de clădiri, harta corespunzătoare a drumurilor, precum și de limitele administrative care sunt utilizate în procesul de sectorizare, care nu numai că trebuie să fie **complete** în ceea ce privește informațiile, **cuprinzătoare** în ceea ce privește acoperirea, dar și suficient de **precise** în ceea ce privește numărul și amplasarea gospodăriilor. De asemenea, baza de date cu drumuri este esențială pentru a completa baza de date cu locuințe pentru a construi sectoare care au în vedere navigația de-a lungul drumurilor. În special, lipsa nivelului minim de calitate al datelor de intrare solicitat permanent a făcut ca procesul de sectorizare să fie inutil de lung și dificil, cu timp de procesare crescut și funcționalitate suplimentară pentru metoda automată. Acest aspect este motivat de faptul că dezvoltarea unei metode automate necesită o metodă și date consistente. Dacă în cazul unor procese realizate manual se pot soluționa abateri ale datelor și metodelor, metodele automate necesită pași consecvenți pentru a crea un sector.

Setul de criterii stabilit de INS pentru sectorizare are atât asemănări, cât și devieri de la liniile directoare ale Națiunilor Unite privind sectorizarea (ONU 2017). Asemănările între criteriile INS și liniile directoare ale ONU includ dimensiunea țintă a unui sector, precum și utilizarea străzilor și a zonelor intravilane pentru navigație. De asemenea, zona intravilană respectă geografia exclusivă reciproc (care nu se suprapune) în care locuiesc majoritatea persoanelor, urmând liniile directoare ale ONU; cu toate acestea, aceste zone intravilane sunt disjunctive și nu asigură o acoperire geografică completă a întregii țări, ceea ce deviază de la liniile directoare ale ONU. Datele despre străzi oferă informații pentru navigare; cu toate acestea, metoda necesită o acoperire geografică completă a străzilor pentru a integra perfect datele împreună, dar străzile prezintă segmente lipsă. Rezultă deci, că aceste date necesită o funcționalitate suplimentară (a metodei automate) cu pași de tratare pentru a atribui amplasări cu locuințe în afara acestor zone.

Metodologia ia în considerare (profită de) disponibilitatea recentă a datelor digitale și progresele înregistrate privind capacitatea de procesare la INS și modelarea. INS a furnizat trei baze de date: dimensiunea și amplasările cu locuințe, drumurile, și limitele intravilane ale localităților la nivel

de 2019. Deoarece metodele automate necesită o structură de date și pași consecvenți, raportul oferă pași specifici și suplimentari pentru tratarea datelor, pași care sunt necesari pentru implementarea modelului. Acești pași suplimentari de tratare identifică oportunități specifice de îmbunătățire ex-ante a calității și acoperirii datelor de intrare, astfel încât, pentru activitățile viitoare de sectorizare, baza de date să îndeplinească cerințele pentru o abordare automată. Metoda funcționează atât în mediul rural, cât și în cel urban, unde a fost acordată o atenție deosebită funcționalității suplimentare a localităților urbane cu un număr mare de locuințe.

Raportul prezintă contextul (inclusiv pașii pentru îmbunătățirea bazei de date), documentația și rezultatele modelului dezvoltat și testat pe larg cu mediul INS de Tehnologia Informației (IT). Rezultatele sunt în concordanță cu bunele practici și raportul evidențiază câteva localități selectate din localități rurale și urbane, precum și rezultatele pentru o localitate urbană cu un număr mare de locuințe (Cluj-Napoca cu 54984 locuințe). Totodată, pe serverul INS dedicat rulării scriptului sunt disponibile rezultatele obținute pentru sectorizarea unui număr semnificativ de localități de diferite dimensiuni inclusiv municipii și sectoare ale capitalei, respectiv: Arad (9271), Bacău (20304), Brăila (42691), București Sectorul 4 (179178), București Sectorul 5 (179187), București Sectorul 6 (179196), Constanța (60428), Galați (75105), Ploiești (130543), Sibiu (143469). Rezultatele depind de calitatea datelor de intrare și de distribuțiile acestora, precum și de parametrii selectați în model. Abaterea medie de la dimensiunea țintă este mai mică în localitățile cu distribuții uniforme și cu un număr relativ scăzut de locuințe în interiorul clădirilor comparativ cu dimensiunea țintă (în general zone rurale și câteva urbane), în timp ce abaterea medie este mai mare în localitățile cu amplasări cu locuințe relativ mai mari decât dimensiunea țintă. În aceste cazuri, un număr mare de locuințe poate fi grupat spațial, astfel încât rezultatele să se îmbunătățească prin evaluarea unei toleranțe specifice pentru fiecare divizare a acestor localități mari.

Acest rezultat a fost dezvoltat pe parcursul a mai mult de doi ani și în mai multe iterații care au urmat procesul de pregătire și implementare a RPL. Pentru realizare a fost utilizată baza de date cu drumuri, actualizată în diferite etape pentru sectorizarea efectivă aplicată în colectarea datelor de recensământului în timpul etapei CAPI, respectiv în perioada mai - iulie 2022. Toată experiența privind sectorizarea, dobândită de echipele INS și BM în timpul pregătirii și implementării recensământului pilot și real (martie 2021 – august 2022), a contribuit la o mai bună documentare a procesului și a fost aplicată în pregătirea și dezvoltarea metodologiei și scriptului pentru producerea rezultatelor, astfel încât acestea să poată fi utilizate și actualizate/extinse ulterior de către echipele de specialiști R și GIS de la INS în anchete structurale viitoare.

Prezentul document conține șapte (7) secțiuni. Prima (1) secțiune oferă o prezentare generală a metodologiilor de sectorizare. A doua (2) secțiune include metodologia. Secțiunea a treia (3) furnizează documentația privind aspectele, etapele de curățare și analitice pentru pregătirea datelor. Secțiunea patru (4) oferă rezultatele dintr-un eșantion de localități, iar Secțiunea cinci (5) oferă recomandări pentru introducerea și metodologia cu privire la date. Secțiunea (6) oferă contextul implementării metodologiei cu script R pe serverul INS. Secțiunea (7) tratează în cele din urmă cerințele pentru introducerea, stocarea, organizarea și întreținerea datelor, conform cerințelor mediului RPL2021. Codul de sectorizare documentat pe larg și codul de management al datelor sunt furnizate ca anexe la prezentul raport și completează secțiunea șapte (7) alături de referințele pentru documentarea livrabilului.

1. Contextul sectorizării în RPL2021

Recensământul decenal este singura activitate de colectare a datelor care generează informații statistice complete și cuprinzătoare privind numărul, structura și alți factori covariabili socio-economici pentru populația totală de persoane, gospodării, locuințe și unități de locuințe din România.

Pe lângă furnizarea unui instantaneu al situației socio-economice actuale a populației României, recensământul servește și ca bază pentru toate anchetele viitoare ca un cadru de eșantionare, precum și pentru tranziția la un recensământ administrativ.

Acest livrabil a fost dezvoltat pe parcursul a mai mult de doi ani, în mai multe iterații, care au urmat procesul de pregătire și implementare al RPL actual și amânat cu 15 luni pentru livrare (incluzând reprogramarea în 2022 a RPL din cauza pandemiei de COVID19, pe baza deciziei CCRPL¹ și îndeplinirea declarației OUG 19/2020² privind organizarea și efectuarea RPL). Motivele amânării au constat în îmbunătățirea metodologiei, având în vedere situația reală a bazelor de date (locuințe și drumuri) și îmbunătățirea acestora, având în vedere modul în care RPL a aplicat sectorizarea în timpul operațiunilor de colectare a datelor în teren și ținând cont de rezultatele RPL runda 2021. Îmbunătățirile au abordat deficiențele datelor și necesitatea adaptării metodologiei de sectorizare în așa fel încât să poată prelucra datele care nu îndeplinesc cerințele specificate pentru un proces automatizat (prezentat în continuare în acest raport) observate în setul de date final înainte de a începe RPL (care a avut îmbunătățiri semnificative față de primul set de date testat, dar, totuși, a avut încă limitări, inerente unui proces dinamic de actualizare). De asemenea, dezvoltarea metodologiei de sectorizare și implicit a soluției (scriptului) de repartizare a locuințelor în cadrul sectoarelor de recensământ a evoluat în stabilirea criteriilor (ex: nr. locuințe în sector de la 100 la 600) și a depins de disponibilitatea resurse pentru recensământ (ex: nr. de recenzori, tablete). Astfel, odată cu implementarea recensământului (auto-recenzare) și prin activitățile recenzorilor în teren (numărul de locuințe alocate într-un sector, numărul de locuințe vizitate și/sau numărul a chestionarelor completate) s-au obținut detalii suplimentare pentru actualizarea bazelor de date clădiri și străzi. Prin aceasta, observațiile privind structura sectoarelor și volumul de muncă colectate de pe teren în faza CAPI (mai-iulie 2022) au fost utilizate în continuare pentru îmbunătățirea procesării datelor și a funcțiilor de parametrizare a metodologiei și scriptului de producere a sectoarelor de recensământ.

Metodele de sectorizare pot beneficia de progresele recente înregistrate în ceea ce privește algoritmii de grupare în clustere și colectarea de date digitale georeferențiate cu Survey Solutions sau alte programe software de colectare a datelor geospațiale. Cu toate acestea, aceste numeroase date noi prezintă o oportunitate pentru câștiguri în eficiență și reproductibilitate sistematică în comparație cu metodele manuale cu muncă intensivă folosite în trecut.

Totuși, acești algoritmi au și ei limitările lor (prezentate în continuare în acest raport), deoarece depind în mare măsură de acuratețea și distribuția datelor de intrare și au cerințe de calcul de

¹ CCRPL - Decizia nr. 1 din 4 noiembrie privind revizuirea Calendarului de pregătire și efectuare a recensământului rotund al populației și locuințelor 2021, modificarea datei de referință a recensământului și alte modificări legate de datele revizuite

² OUG 19/2020 privind organizarea și desfășurarea Recensământului populației și locuințelor din România în anul 2021, modificată și aprobată prin Legea nr. 178/2020.

efectuat. Deoarece orice algoritm din acest domeniu se bazează pe convergența unui proces de optimizare testat iterativ, orice date greșite pot duce la o lipsă a convergenței algoritmului sau eșuează ignorând eventuale restricții dorite (sau, în cel mai rău caz, toate).

1.1. Obiectivele sectorizării

Proiectarea zonelor de recensare ar trebui să ia în considerare diverse criterii. Manualul ONU 2017 (pagina 88, Organizația Națiunilor Unite, 2017) afirmă că, delimitate corect, zonele de recensare vor:

- (a) fi reciproc excluzive (nu se suprapun) și exhaustive (acoperire a întregii țări, nu numai a zonei intravilane actuale;*
- (b) avea limite care pot fi ușor identificate în teren (pentru navigație);*
- (c) fi în concordanță cu ierarhia administrativă (identificare și delimitare geografică);*
- (d) fi compacte din punct de vedere geografic și să nu aibă lipsuri sau secțiuni despărțite;*
- (e) avea populații de dimensiune aproximativ egală; și*
- (f) fi suficient de reduse și accesibile pentru a fi acoperite de un recenzor în perioada recensământului;*
- (g) fi suficient de mici și flexibile pentru a permite cea mai largă gamă de tabelări pentru diferite unități de raportare statistică; și*
- (h) fi utile pentru alte tipuri de recensăminte și activități de colectare a datelor (de exemplu, recensământ agricol, anchete etc.).*

Pe lângă cerințele minime pentru zonele de recensare din perspectiva unui practician, există și cerințe metodologice, care sunt necesare pentru o abordare automată după cum urmează:

- (a) Baza de date este completă și conține toate datele necesare (adică nu conține elemente de stradă lipsă sau unități în afara limitelor); și**
- (b) Numărările utilizate pentru valorile țintă (adică numărul de locuințe) sunt corecte și reprezintă corect distribuția geografică.**

Ulterior discuțiilor cu INS, au fost furnizate următoarele criterii în ordinea priorităților:

- Dimensiunea țintă a unui sector este de exact 100 sau 600 de locuințe (acolo unde este posibil). Sunt acceptate abateri de +/-20%, mai ales în zonele cu blocuri de apartamente (unde valoarea numărului de locuințe variază aleatoriu și poate depăși numărul țintă de locuințe). Aceste abateri nu trebuie să fie o regulă, ci mai degrabă o excepție în cazul în care nu este posibil altfel;
- Sectoarele sunt construite ținând cont de străzi (un sector trebuie să conțină codurile administrative care aparțin aceleiași străzi sau mai multor străzi, în cazul în care nu acoperă 100 sau 600 de locuințe);
- Sectorizarea trebuie să conțină toate numerele administrative chiar dacă acestea depășesc limita urbană oficială;
- Sectorizarea trebuie făcută la nivelul intravilan al localității (având în vedere nivelurile SIRUTA);
- Dacă într-o localitate (în special pentru mediul rural) ultimul sector este incomplet (mai puțin de 20% din numărul țintă de locuințe), atunci numărul total de sectoare poate fi redus cu 1 prin redistribuirea numărului de locuințe și regenerarea sectorizării într-o nouă iterație.

Liniile directe ale ONU, precum și cerințele metodologice și criteriile INS se abat de la câteva dintre criterii. În timp ce criteriile de dimensiune egală și exclusivitate reciprocă a sectoarelor sunt în conformitate cu liniile directe ale ONU, lipsa acoperirii complete a întregii țări, precum și prezența oricăror secțiuni disjuncte nu respectă liniile directe menționate. În plus, baza de date incompletă a străzilor, prezența oricăror unități de recensământ în afara limitelor administrative, precum și numărarea inexactă a locuințelor încalcă cerințele pentru o abordare automată și au contribuit semnificativ la întârziere. În timp ce aceasta din urmă este o problemă mai mică atunci când sectorizarea este efectuată manual, deoarece oamenii pot face față unor astfel de excepții, este o problemă considerabilă atunci când este efectuată automat.

În plus, metodologia trebuie să ia în considerare distribuția spațială a străzilor, pentru a ușura navigarea și ierarhia administrativă. Zona intravilană respectă geografia reciproc exclusivă (nu se suprapune) și abordează zonele în care locuiesc oamenii, ceea ce este exhaustiv la nivel de intravilan. Cu toate acestea, aceste zone intravilane nu oferă o acoperire geografică completă a întregii țări. Prin urmare, procedurile suplimentare atribuie amplasări cu locuințe care nu sunt situate, în limitele acestor zone intravilane, de cea mai apropiată localitate, iar zonele intravilane din aceeași localitate sunt conectate pentru a atribui vecinii unei locuințe.

1.2. Principiile sectorizării

Metodologia trebuie să ia în considerare trei principii principale ale sectorizării.

- (A) Primul principiu este de a respecta limitele administrative. Limitele administrative și sociale oferă credibilitate procesului de diseminare, deoarece informațiile de recensământ pot fi afișate cu precizie la diferite niveluri geografice, cum ar fi la nivel de regiune, unitate teritorială și localitate.
- (B) Al doilea principiu este de a respecta limitele navigabile, ceea ce facilitează o referință clară pentru munca recenzorilor și poate include drumuri, râuri, cărări și repere (puncte de interes).
- (C) Al treilea principiu este să adere la o zonă traversabilă. Luarea în considerare a terenului și a elevației oferă recenzorului o zonă practică de parcurs în timpul dat.

2. Metodologia de sectorizare

Folosim algoritmi despre redistribuire din literatura științelor politice pentru a determina sectorizarea pentru eșantioanele rurale și urbane restrânse la unitățile administrative teritoriale (UAT) și la amplasări cu locuințe atribuite unei localități. Folosim o metodă restrânsă din punct de vedere spațial pentru a ne asigura că toate sectoarele sunt alăturate spațial și luăm în considerare o dimensiune țintă a numărului de locuințe pentru fiecare sector (și anume, cluster). Cu alte cuvinte, algoritmul de grupare în clustere ar trebui să ia în considerare amplasarea fiecărei locuințe în raport cu toate celelalte locuințe. În plus, metodologia trebuie să ia în considerare paritatea numărului de locuințe în sectoarele unei localități în cea mai mare măsură posibilă pentru a putea asigura echilibrul populației per sector. În realitate, metoda trebuie să efectueze o reconciliere între aceste două obiective în ceea ce privește parametrizarea și ponderile. În plus, distribuția spațială și intervalul numărului de locuințe pot restrânge soluțiile dincolo de o anumită toleranță țintă de +/- 20%.

Eșantionul de localități este divizat atât la nivelurile UAT, cât și la nivel de localitate, pentru a se asigura că sectorizarea este conformă cu aceste limite statistice și principii de sectorizare³. De asemenea, sectoarele ar trebui numerotate secvențial pentru fiecare localitate.

2.1. Model

Folosim pachetele de software disponibile în mediul software-ului R statistic și aplicate în cercetările care încadrează redistribuirea ca problemă de divizare grafică (Deford și colab. 2019; Fifield și colab. 2020b). Folosim funcția de îmbinare și divizare care utilizează algoritmul Markov Chain Monte Carlo, inclusiv constrângerile privind contiguitatea spațială, toleranța la paritatea populației între segmente și compactitatea geografică (a se vedea Carter și colab. 2019). Contiguitatea spațială este estimată din amplasările disponibile cu drumuri și locuințe dintr-o localitate. Toleranța la paritatea populației este un parametru care asigură toleranța maximă a diferenței în ceea ce privește numărul de locuințe pentru fiecare sector sau sub-sector din localitate. Executăm mai multe simulări și, ulterior, selectăm soluția cu cea mai mică constrângere a toleranței la paritatea populației. În cazul în care numărul de sectoare estimate este mare (adică peste 50 într-o localitate), folosim un algoritm de schimb pentru a grupa toate locuințele într-un grup. Ulterior, algoritmul de fuziune și divizare procesează fiecare grup, iar algoritmul agregă rezultatele grupului în întreaga localitate.

Numărul de locuințe dintr-o amplasare poate depăși dimensiunea țintă pentru un sector. De exemplu, dacă dimensiunea țintă a sectorului este 100, atunci o amplasare cu 150 de locuințe depășește dimensiunea țintă de 100. Utilizăm un proces în mai multe etape pentru a stratifica datele în două grupuri: amplasările cu un număr de locuințe sub și peste un interval țintă. În cazul în care este vorba de o depășire, atunci alocarea temporară a dimensiunii locuinței este stabilită la un număr situat chiar sub dimensiunea țintă din localitate, ceea ce este o cerință necesară a modelului. După finalizarea algoritmului de grupare în clustere, rezultatele înregistrează dimensiunea reală a locuinței/clădirii. Această limitare poate explica rezultatele cu dimensiunile de locuințe/clădiri care

³ De exemplu, amplasările cu locuințe care se află în afara UAT sunt excluse (și anume, 156 amplasări în 100353).

se abat de la dimensiunea țintă a localității, în special atunci când diferența dintre dimensiunea țintă și dimensiunea reală a locuinței/clădirii este mare (de exemplu, amplasări cu clădiri cu 250 de locuințe, dimensiunea țintă fiind de 100 de locuințe).

2.2. Informații necesare

Modelul necesită numărul estimat de sectoare, o toleranță pentru numărul de locuințe dintr-un sector din rezultatele modelului, numărul de locuințe dintr-o amplasare, adiacența spațială între amplasările dintr-o localitate și numărul de simulări. Numărul estimat de sectoare depinde de numărul total de locuințe cu date dintr-o localitate și de un număr țintă de locuințe per sector. Pe lângă numărul de locuințe dintr-o localitate, amplasarea locuințelor este esențială pentru a identifica vecinii din fiecare amplasare cu locuințe și toate amplasările cu locuințe sunt conectate în aceeași localitate. Datele georeferențiate despre drumuri pot documenta în continuare identificarea vecinilor pentru a ușura navigarea în sector în concordanță cu realitatea la fața locului privind infrastructura de transport. Cazul ideal include date georeferențiate despre străzi, cu acoperire geografică actuală și completă, precum și segmente de străzi conectate digital (și anume, corecte din punct de vedere topologic). Cu toate acestea, metodologia este flexibilă, pentru a utiliza cele mai bune date disponibile despre străzi și pentru a completa golurile de date, după cum este necesar în vederea stabilirii că toate amplasările sunt conectate prin includerea de sub-proceduri pentru a adapta această abatere în mod rezonabil. Această abordare profită de două structuri principale de date spațiale: raster și vector. Datele raster sunt spațial contigue prin definiție și îndeplinesc nevoia unei acoperiri geografice complete ca o cerință a sectorizării, în timp ce datele vector pot să nu fie corecte din punct de vedere topologic și pot fi disjunctive cu o localitate.

3. Pregătirea datelor

3.1. Surse de date de intrare

Intrările furnizate includ:

- Un set de limite administrative georeferențiate, inclusiv UAT și 13.736 de localități din mediile urban sau rural. Limita localității este oficială și nu oferă o acoperire geografică completă a țării (Figura 1). În plus, limitele localității nu sunt neapărat conectate geografic (Figura 2).

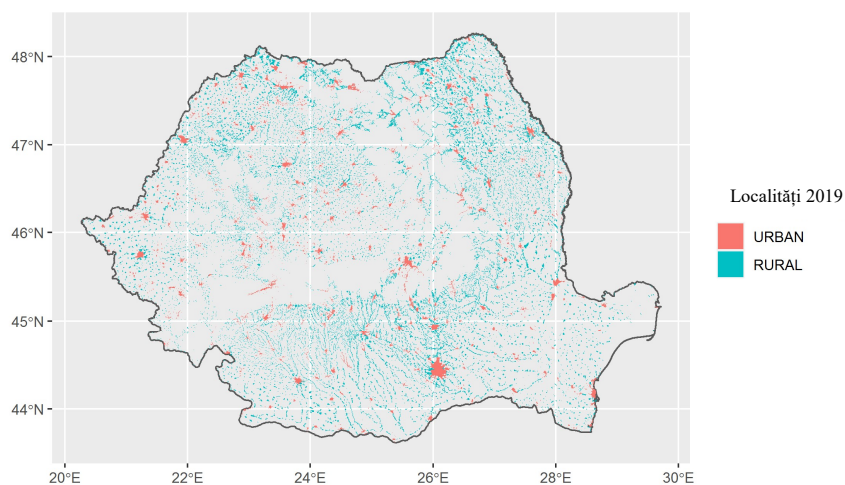


Figura 1: Harta localităților cu zone rurale și delimitări (2019)

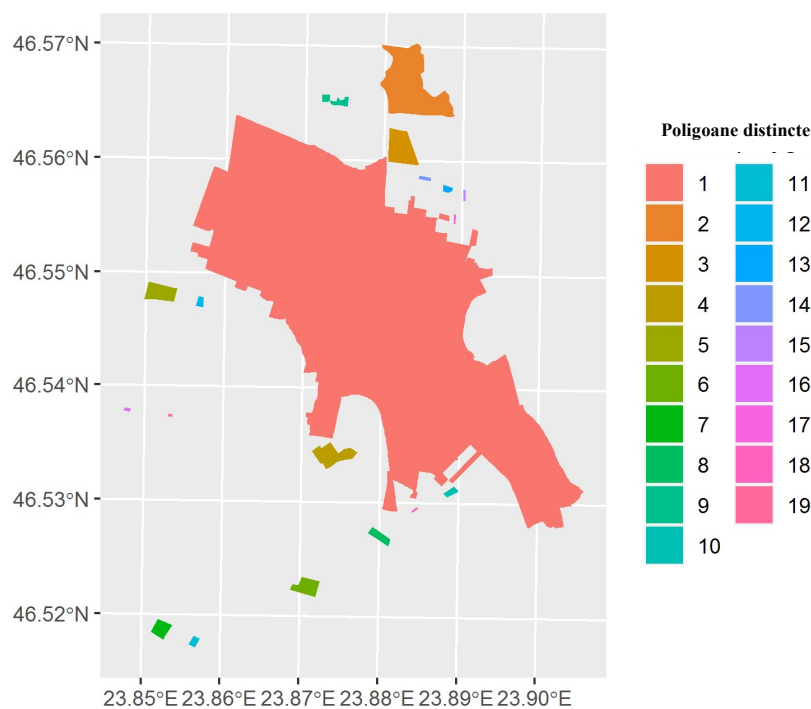


Figura 2: Harta unei localități cu zone nealăturate

- Baza de date cu locuințe georeferențiate (WEBGEO) are un identificator unic pentru fiecare amplasare și număr de locuințe.
- Baza de date cu drumuri georeferențiate are linii cu drumuri și nu este corectă din punct de vedere topologic. Majoritatea drumurilor au o denumire; drumurile din localitățile rurale pot să nu prezinte variație în ceea ce privește denumirile. Mai multe informații referitoare la tratarea datelor din baza de date cu străzi sunt prezentate în ***Rezultatul 7a – Raport privind serviciile de asistență tehnică oferite Beneficiarului privind Metodologia de evaluare și promovare continuă și în timp real a calității și acoperirii datelor colectate pentru RPL2021 și protecția/securitatea datelor***, parte a acestui acord RAS.

3.2. Pregătirea datelor de intrare

Pregătirea datelor de intrare este crucială pentru proces. Metodologia necesită ca toate amplasările să îndeplinească aceste cerințe: un număr pozitiv de locuințe, coordonate distincte (fără coordonate dublate), o singură înregistrare pentru fiecare amplasare sub formă de punct (fără geometrii de mai multe tipuri), amplasări cu locuințe care sunt situate complet într-o zonă administrativă identificabilă și codurile administrative derivate din definițiile spațiale, astfel încât intersecția geometrică a zonelor administrative să fie în concordanță cu codurile din baza de date cu locuințe. Astfel, dacă nu îndeplinesc aceste cerințe și nu conțin suficiente informații pentru tratarea datelor, înregistrările necesită o procedură manuală.

Metodologia include un tabel cu înregistrările excluse pentru a modifica înregistrările în vederea îndeplinirii cerințelor, dacă este cazul, înainte de a genera sectoarele pentru localitatea în cauză. Dacă sunt identificate înregistrări care nu îndeplinesc cerințele, acestea sunt corectate și apoi este folosit din nou algoritmul pe toate înregistrările inițiale din localitate.

3.3. Tratarea datelor

Tratarea datelor este secvențiată în următoarea ordine: amplasări cu locuințe, drumuri și limita intravilană a localității. Secțiunea 7 oferă informații despre scripturile documentate pe larg scrise în software-ul R ([R software](#)) cu privire la tratarea datelor, modelarea și gestionarea datelor.

În primul rând, identificăm toate locuințele din limitele administrative imbricate ale UAT și ale localității. Tratarea datelor privind datele despre locuințe include corectarea automată a următoarelor deficiențe în datele de intrare:

- În cazul în care o localitate are locuințe cu mai mult de un cod de județ, codul de județ este atribuit codului din numărul maxim de observații dintr-o localitate.
- În cazul în care o observație nu include coordonate unice și o geometrie validă, observația a fost exclusă.
- În cazul în care numărul de locuințe dintr-o amplasare nu este în format pozitiv întreg (valoare negativă sau 0), observația a fost exclusă.
- În cazul în care observația are un identificator neunic în datele de intrare, s-a creat un nou identificator unic în scopul exclusiv al acestui algoritm.

- În cazul în care amplasările cu locuințe erau aflate la 5 metri de o altă locuință, au fost modificate în punctul lor central pentru a reflecta suma locuințelor în ambele amplasări. (Numărul de metri este relevant pentru scala de procesare raster).

În al doilea rând, stabilim legătura spațială dintre locuințele din localitate folosind datele despre drumuri. Mai întâi, sub-clasificăm datele despre drumuri alocate localității. În continuare, algoritmul efectuează următorii pași pentru a stabili o relație spațială:

- Algoritmul construiește noi segmente de linie de la fiecare locuință până la cel mai apropiat drum.
- Rezultatul noilor segmente de linie poate să nu fie complet, astfel încât sunt necesare modificări suplimentare pentru a conecta toate segmentele de linie de drum, iterând distanța minimă. Acest lucru a fost, din păcate, complicat de lipsa sau caracterul incomplet al datelor coerente privind topologia străzilor.
- Ultimul pas transformă linia într-un poligon și apoi atribuie zona din poligon până la cea mai apropiată locuință pentru a construi matricea de adiacență spațială.

În al treilea rând, limita oficială a localității din 2019 necesită o ajustare, deoarece localitatea nu oferă o acoperire geografică completă pentru o alocare a amplasării cu locuințe într-o singură localitate. Localitatea face parte doar din zona intravilană, care include majoritatea amplasărilor cu locuințe. Chiar dacă intravilanul cuprinde toate locuințele, metoda necesită o acoperire geografică completă per localitate. Așadar, programul modifică limita localității intravilane pentru a se asigura că toate locuințele se află într-un singur poligon, ceea ce este o cerință tehnică a modelului.

3.4. Eșantionul de localități

Oferim două seturi de rezultate: un set aleatoriu de localități și un set predefinit de două localități rurale și două localități urbane, furnizate de INS. În primul rând, un eșantion aleatoriu de 100 de localități demonstrează utilitatea metodologiei în cadrul unei distribuții. Folosim comanda eșantion în R fără înlocuire. Suntem interesați în mod special să arătăm momentele în care metodologia de sectorizare necesită un număr estimat de sectoare mai mare sau egal cu doi. Așadar, folosim un subeșantion din cele 100 de localități pentru a examina algoritmul, care include 61 de localități cu între 150 și 6.000 de amplasări cu locuințe.

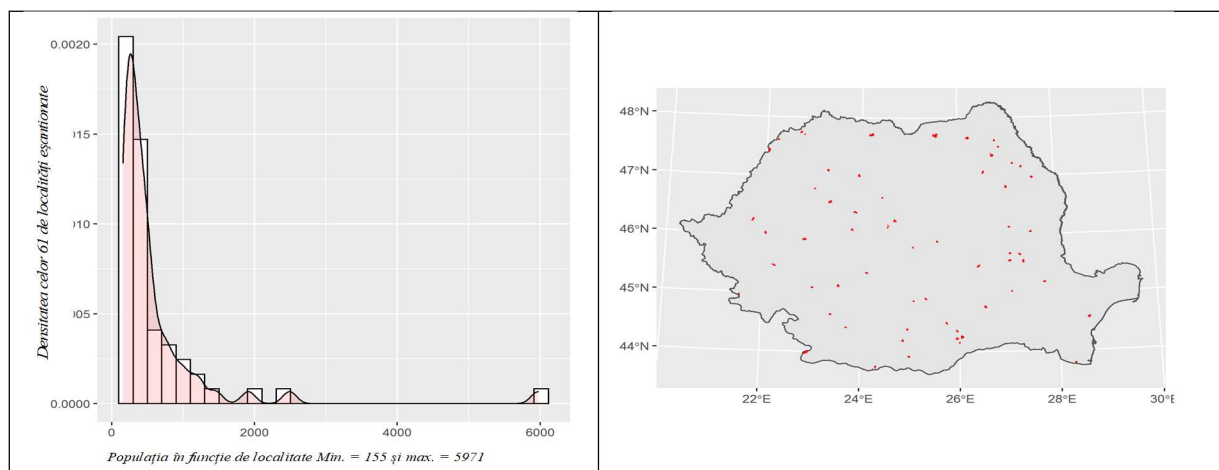


Figura 3: Harta subeșantionului prezintă populația în funcție de localitate (stânga) și amplasările (dreapta)

Distribuția locuințelor totale într-o localitate este inegală în rândul populațiilor mai mici (Figura 3a). Acest lucru se reflectă în numărul total de localități rurale și urbane. De asemenea, amplasarea geografică prezintă observații în toată țara (Figura 3b).

3.5. Exemple de localități propuse de INS

INS a pus la dispoziție două localități rurale și două localități urbane pentru a demonstra metodologia. Mai jos, hărțile panou ale fiecărei localități ilustrează principalele etape de tratare a datelor privind amplasările cu locuințe, drumuri și limita localității. Harta panou A prezintă amplasările cu locuințe și drumurile după tratarea de bază al amplasărilor cu locuințe și dimensiunii locuințelor, împreună cu subșetul de drumuri alocate localității (în intravilan sau extravilan). Harta panou B ilustrează legătura dintre segmentele de drum și toate locuințele. Harta panou C ilustrează conectarea iterativă a tuturor segmentelor de drum și a legăturilor cu toate locuințele într-o singură linie conectată cu limita intravilanului, ajustată pentru a include toate locuințele și pentru a păstra numai zonele intravilane cu amplasări cu locuințe. Harta panou D ilustrează transformarea liniei conectate într-o zonă tampon.

3.6. Localități rurale

Cele două localități rurale sunt Perieți și Sâncrăieni.

- Sâncrăieni este situată în Harghita, cu 862 de amplasări cu 1.063 de locuințe. Localitatea Sâncrăieni cuprinde mai multe zone intravilane, care pot avea sau nu locuințe. Etapele de tratare a datelor modifică limita localității pentru a îngloba toate amplasările cu locuințe, cu includerea zonelor intravilane și a drumurilor relevante (Figura 4).
- Perieți este situată în Ialomița, cu 560 de amplasări cu 570 de locuințe. Localitatea Perieți conține doar câteva zone intravilane distincte și aproape toate amplasările cu locuințe au o singură locuință. Etapele de tratare a datelor modifică limita localității pentru a îngloba toate amplasările cu locuințe, cu includerea zonelor intravilane și a drumurilor relevante (Figura 5).

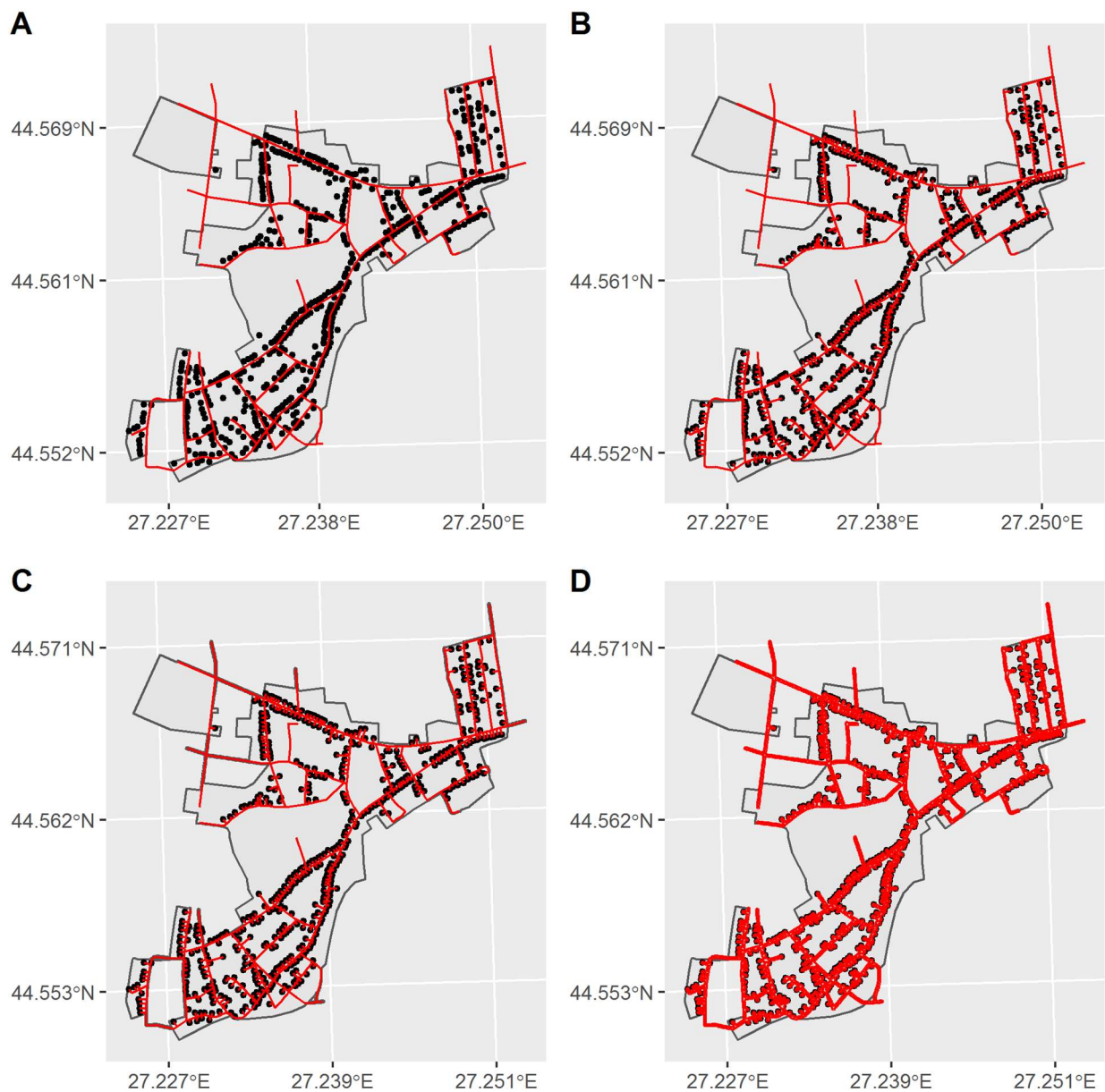


Figura 4: Hărțile (panou A-D) ilustrează principalele etape de tratare a datelor pentru modelul din Săncrăieni.

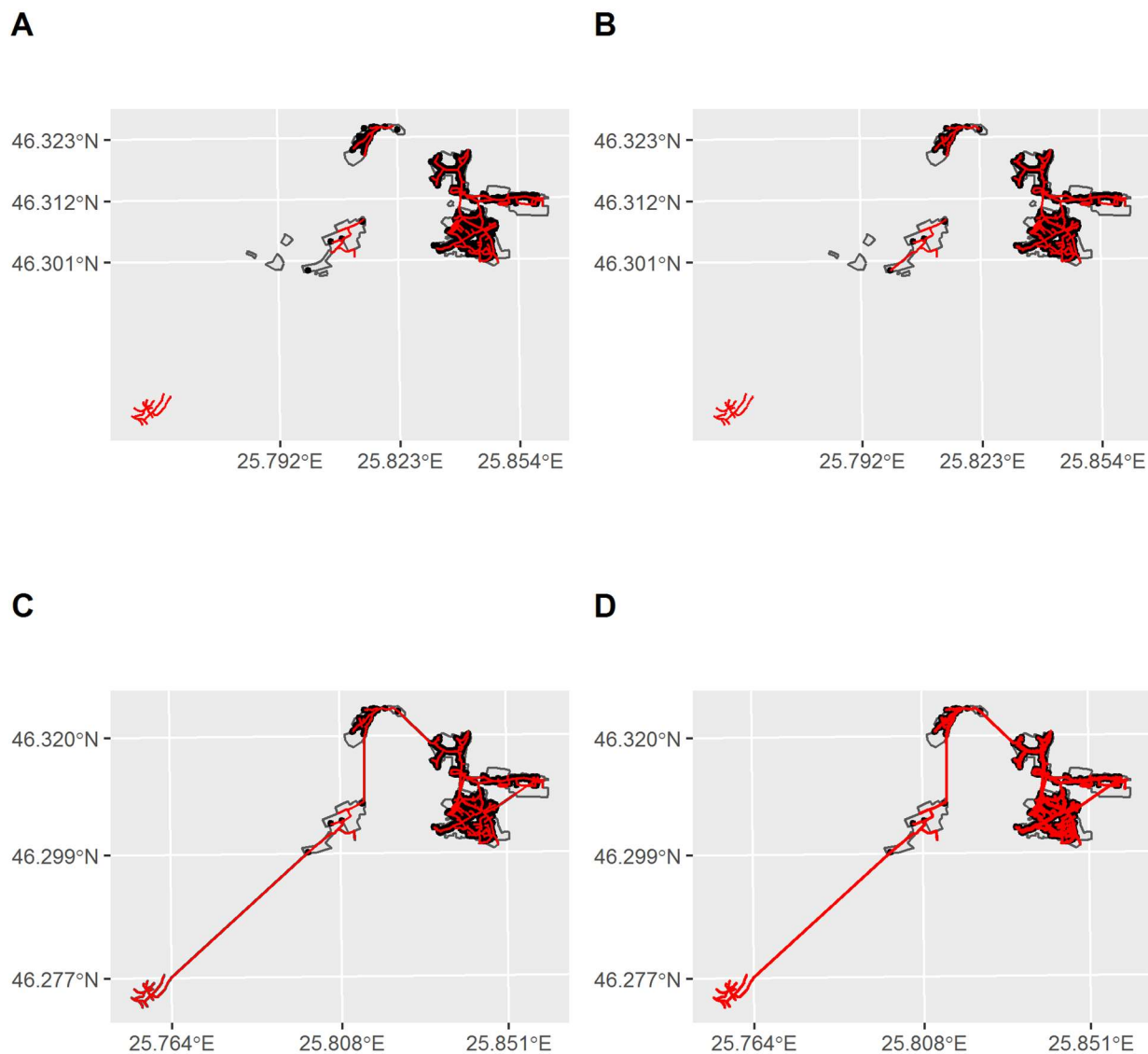


Figura 5: Hărțile (panou A-D) ilustrează principalele etape de tratare a datelor pentru modelul din Perieți.

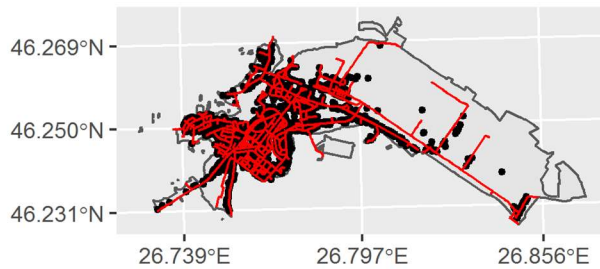
3.7. Localități urbane

Cele două localități urbane sunt Onești și Câmpia Turzii.

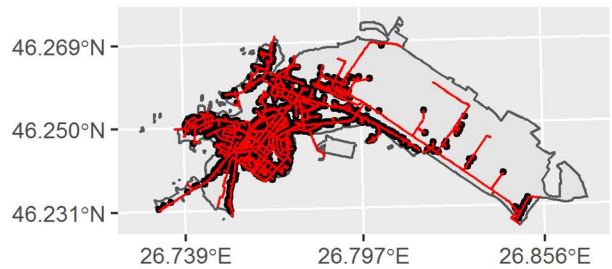
- Onești este situată în Bacău, cu 2.650 de amplasări cu 21.745 de locuințe. Localitatea Onești cuprinde mai multe zone intravilane, care pot avea sau nu locuințe. Etapele de tratare a datelor modifică limita localității pentru a îngloba toate amplasările cu locuințe, cu includerea zonelor intravilane și a drumurilor relevante.

- Câmpia Turzii este situată în Cluj, cu 4.071 de amplasări cu 10.317 locuințe. Localitatea Câmpia Turzii cuprinde mai multe zone intravilane, care pot avea sau nu locuințe. Etapele de tratare a datelor modifică limita localității pentru a îngloba toate amplasările cu locuințe, cu includerea zonelor intravilane și a drumurilor relevante.

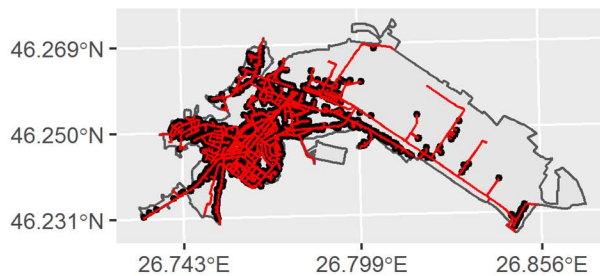
A



B



C



D

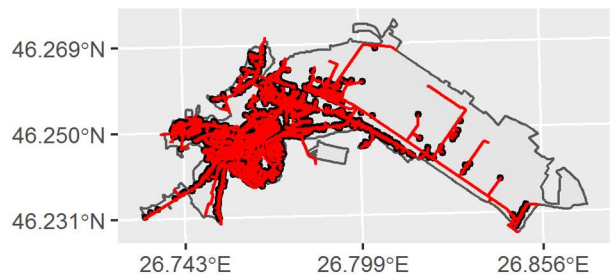


Figura 6: Hărțile (panou A-D) ilustrează principalele etape de tratare a datelor pentru modelul din Onești.

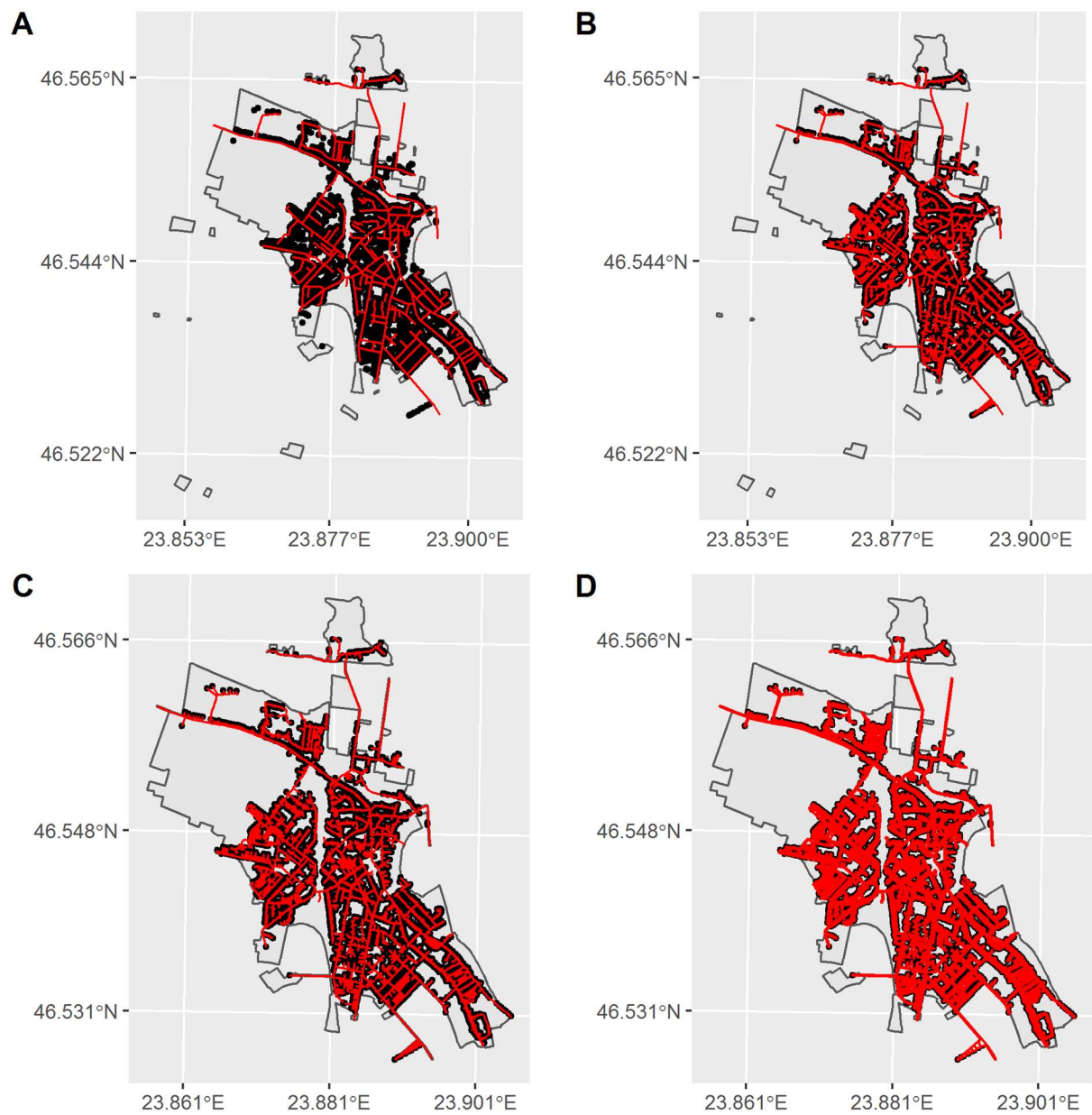
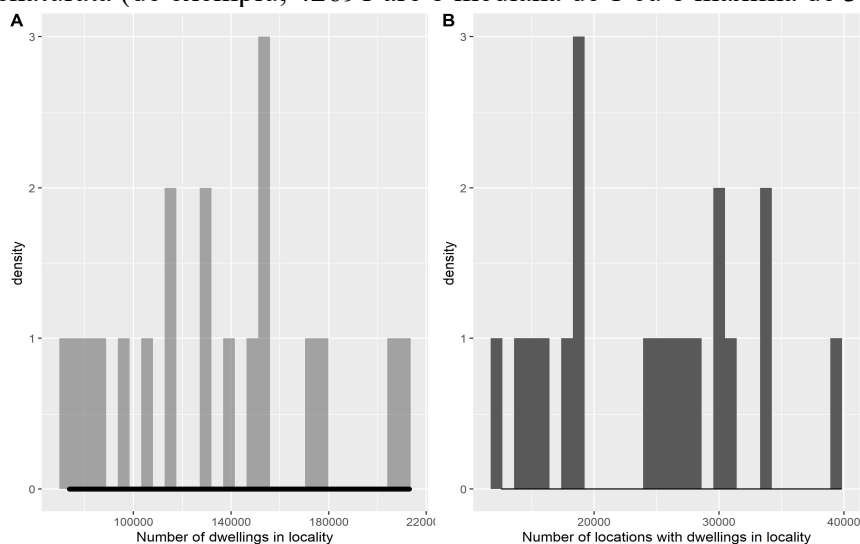


Figura 7: Hărțile (panou A-D) ilustrează principalele etape de tratare a datelor pentru modelul din Câmpia Turzii

3.8. Zona rurală – localități de mari dimensiuni

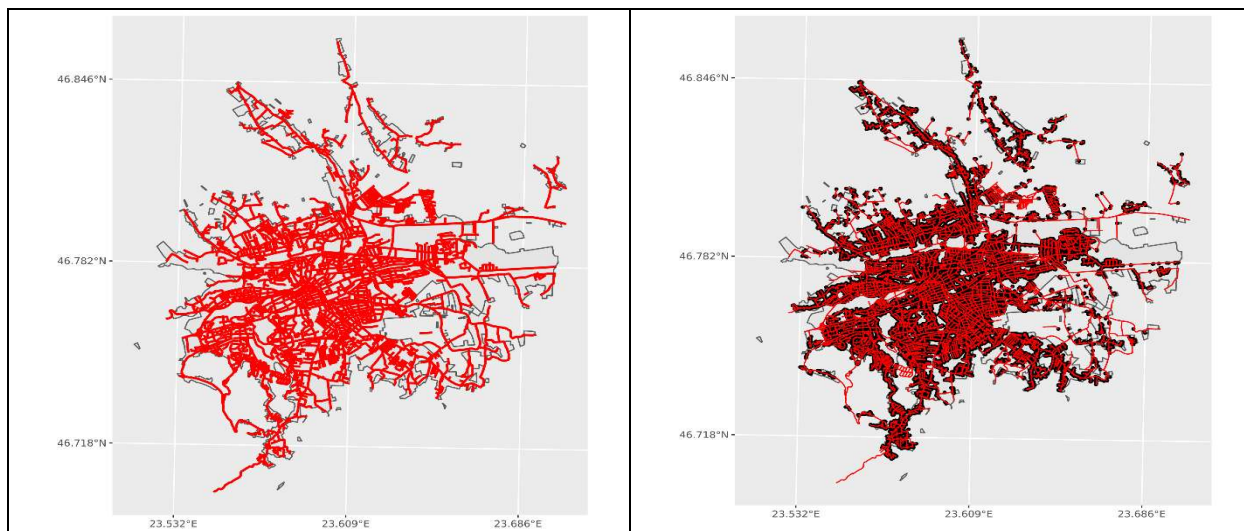
INS a furnizat o listă de 20 de localități⁴ cu număr mare de locuințe⁵ (a se vedea Figura Numărul de locuințe din localitate Numărul de amplasări cu locuințe din localitate

Figura 8 de mai jos). Șase localități (sectoare) se află în București. Distribuția numărului de locuințe este denaturată (de exemplu, 42691 are o mediană de 1 cu o maximă de 396).



Numărul de locuințe din localitate Numărul de amplasări cu locuințe din localitate
Figura 8: Histograma numărului de locuințe dintr-o localitate pentru eșantionul urban de localități de mari dimensiuni

Raportul va scoate în evidență localitatea Cluj-Napoca din această listă de 20. Cluj-Napoca are 28969 de amplasări cu 168.705 de locuințe în baza de date. Tratarea datelor include conectarea a 430 de străzi deconectate (Figura 9).



⁴ Localitățile includ: Arad (9271), Bacău (20304), Brăila (42691), Brașov (40205), București Sectorul 1 (179141), București Sectorul 2 (179150), București Sectorul 3 (179169), București Sectorul 4 (179178), București Sectorul 5 (179187), București Sectorul 6 (179196), Cluj-Napoca (54984), Constanța (60428), Craiova (69919), Galați (75105), Iași (95079), Oradea (26573), Ploiești (130543), Sibiu (143469) și Timișoara (155252)

⁵ Localitatea cu codul 20304 nu are un fișier de delimitare corespunzător, deci nu a fost procesată.

Figura 9: Străzile din Cluj-Napoca (stânga) sunt tratate pentru a conecta toate străzile (partea dreaptă)

4. Rezultate

În general, setul de localități din eșantionul aleatoriu funcționează corespunzător. Sectoarele automatizate sunt conectate spațial și sunt influențate de date cu privire la drumuri, astfel cum se evidențiază de forma fiecărui sector. Populația totală din fiecare sector al unei localități este echilibrată în cazul în care cea mai mare parte a eșantionului are o pondere a populației cu o abatere medie de 0,2, cu o țintă de 100 de locuințe. În timp ce numărul de localități cu mai mult de un sector la o țintă de 600 de locuințe este redus în eșantion, rezultatele selectate sunt mai aproape de paritate (Figura 10).

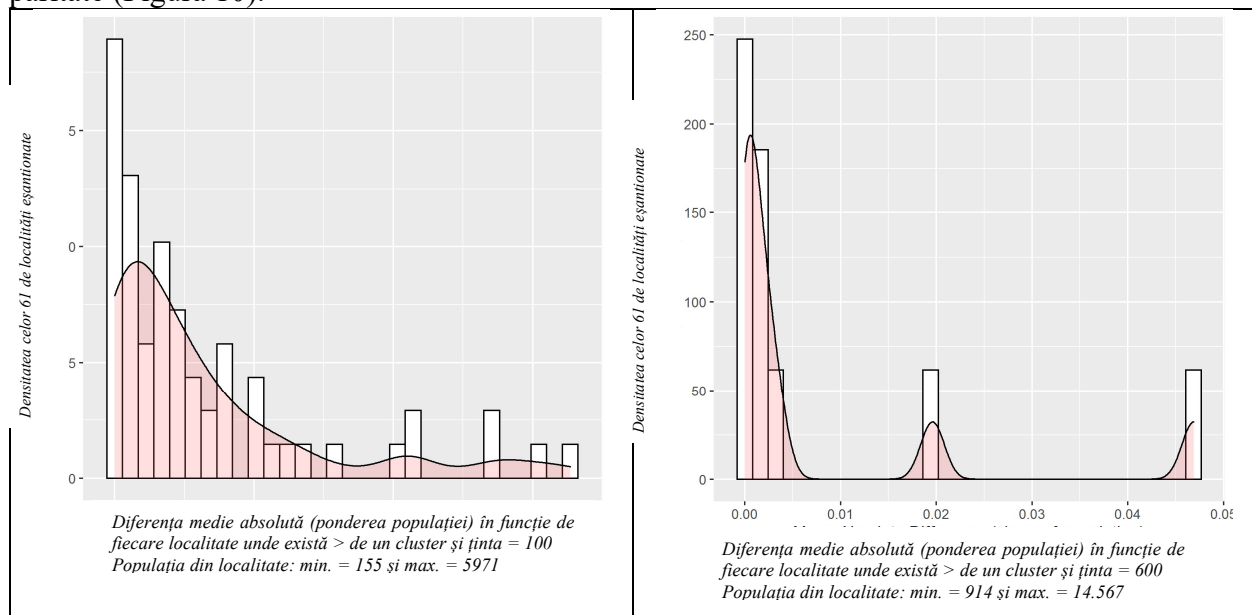


Figura 10: Aceste histograme ilustrează diferența medie absolută (ponderea numărului de locuințe) cu o țintă de 100 (stânga) și 600 (dreapta).

Mai jos sunt prezentate rezultatele eșantionului rural (4.1) și ale eșantionului urban (4.2).

4.1. Eșantionul rural

Hărțile panou de mai jos ilustrează alocarea sectorului în urma rezultatului algoritmului de grupare în clustere și reprezentarea geografică a sectorului. Algoritmul se aliniază cu soluțiile cu o toleranță medie a populației mai mică de 0,10 pentru o localitate.

- Sâncrăieni are 6 sectoare cu un număr de 92-97 de locuințe per sector din localitate (Figura 11: Hărțile (panou A-B) ilustrează rezultatul modelului (A) și sectorului (B) din Sâncrăieni și Figura 13).

- Perieți are 11 sectoare cu un număr de 89-104 de locuințe per sector din localitate (Figura 12 și Figura 13).

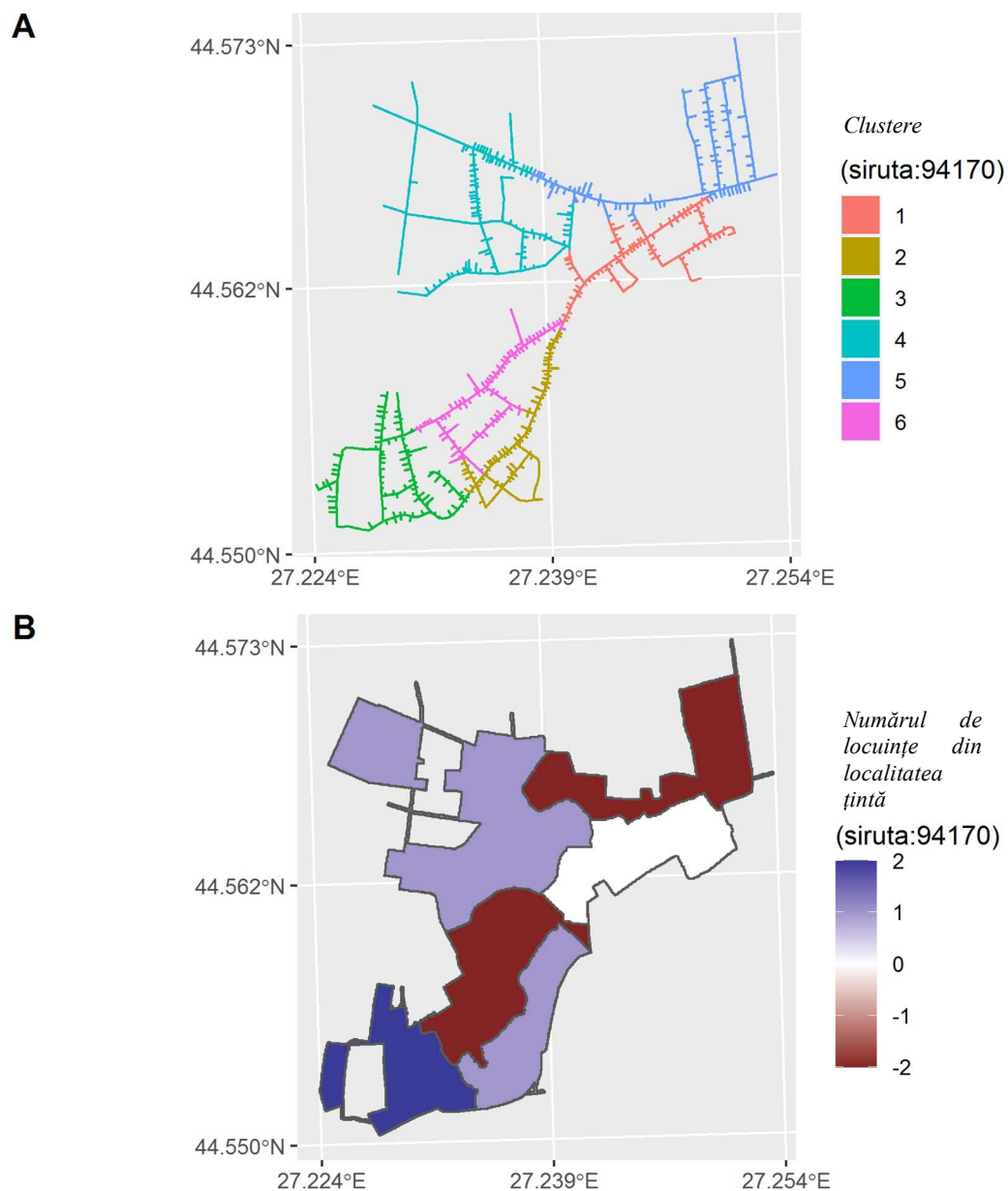


Figura 11: Hărțile (panou A-B) ilustrează rezultatul modelului (A) și sectorului (B) din Sâncrăieni

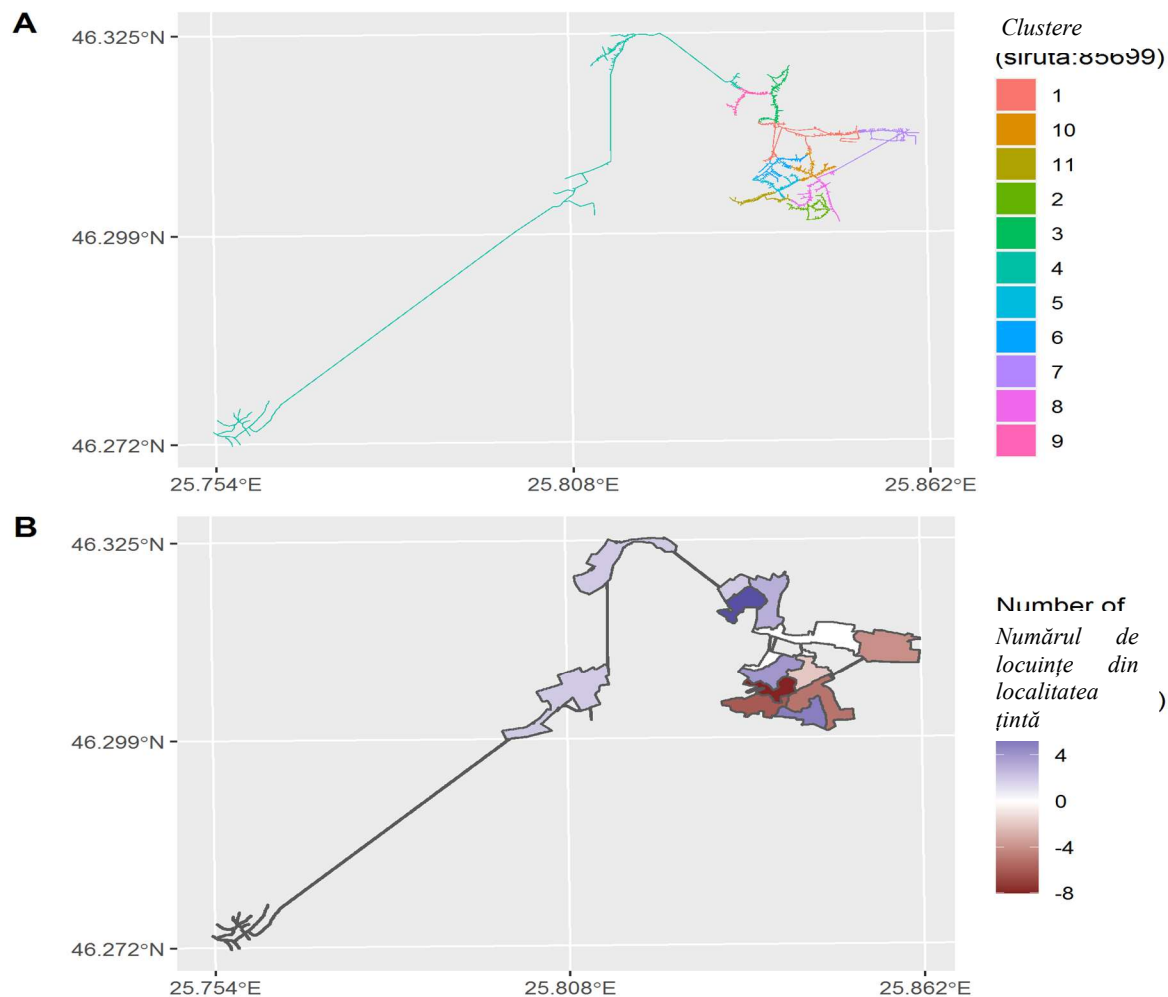
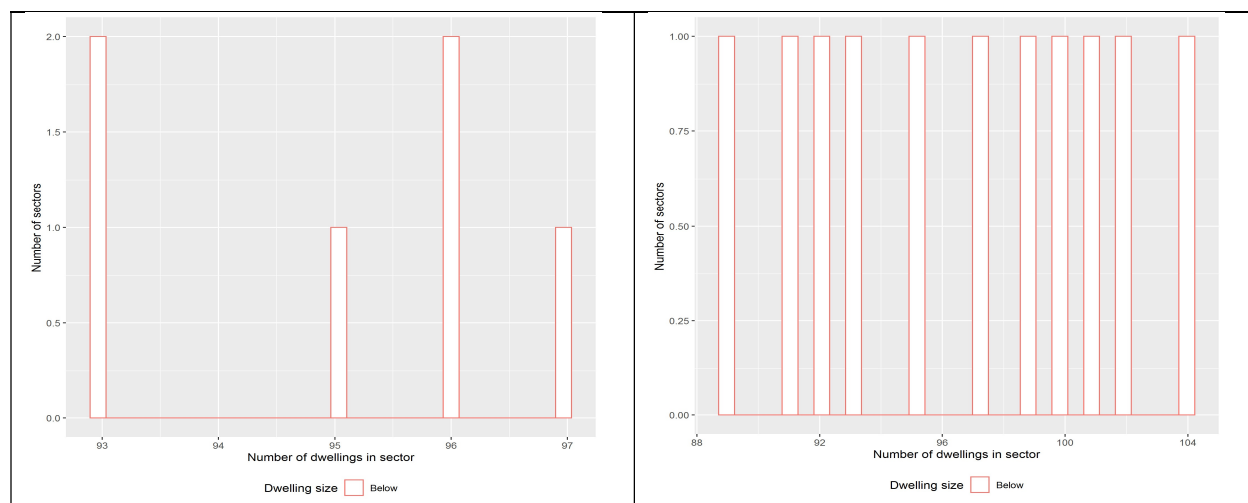


Figura 12: Hărțile (panou A-B) ilustrează rezultatul modelului (A) și sectorului (B) din Perieți



Legendă:
 Numărul de sectoare
 Numărul de locuințe din sector
 Dimensiune locuință
 Sub

Numărul de sectoare
 Numărul de locuințe din sector
 Dimensiune locuință
 Sub

Figura 13: Histograma afișează numărul de locuințe dintr-un sector situat peste și sub dimensiunea țintă în Sâncrăieni (stânga) și Perieți (dreapta).

4.2. Eșantionul urban

Hărțile panou de mai jos ilustrează alocarea sectorului în urma rezultatului algoritmului de grupare în clustere și reprezentarea geografică a sectorului. Mai jos sunt două seturi de rezultate atât pentru o dimensiune țintă de 100, cât și pentru o dimensiune țintă de 600. Dimensiunea țintă de 100 este mai mică decât numărul total de locuințe dintr-o amplasare; cu toate acestea, dimensiunea țintă de 600 este mai mare decât numărul maxim de locuințe dintr-o amplasare. Aceste rezultate permit compararea între o dimensiune țintă mai mică și o dimensiune țintă mai mare în ceea ce privește diferența dintre numărul de locuințe pentru fiecare sector față de dimensiunea țintă a localității.

Pentru dimensiunea țintă de 100, intervalele numărului de locuințe dintr-un sector sunt mai mari decât în cazul celor două exemple rurale din cauza abaterii însemnate a numărului de locuințe, incluzând amplasările care depășesc dimensiunea țintă de 100.

- Onești are 103 sectoare, cu un interval de 63-136 de locuințe per sector din localitate (Figura 14 și Figura 16). Onești include un număr de sectoare cu amplasări cu locuințe care depășesc ținta, astfel cum este menționat în histogramă.
- Câmpia Turzii are 202 sectoare, cu un interval de 52-312 de locuințe per sector din localitate (Figura 15 și Figura 16).

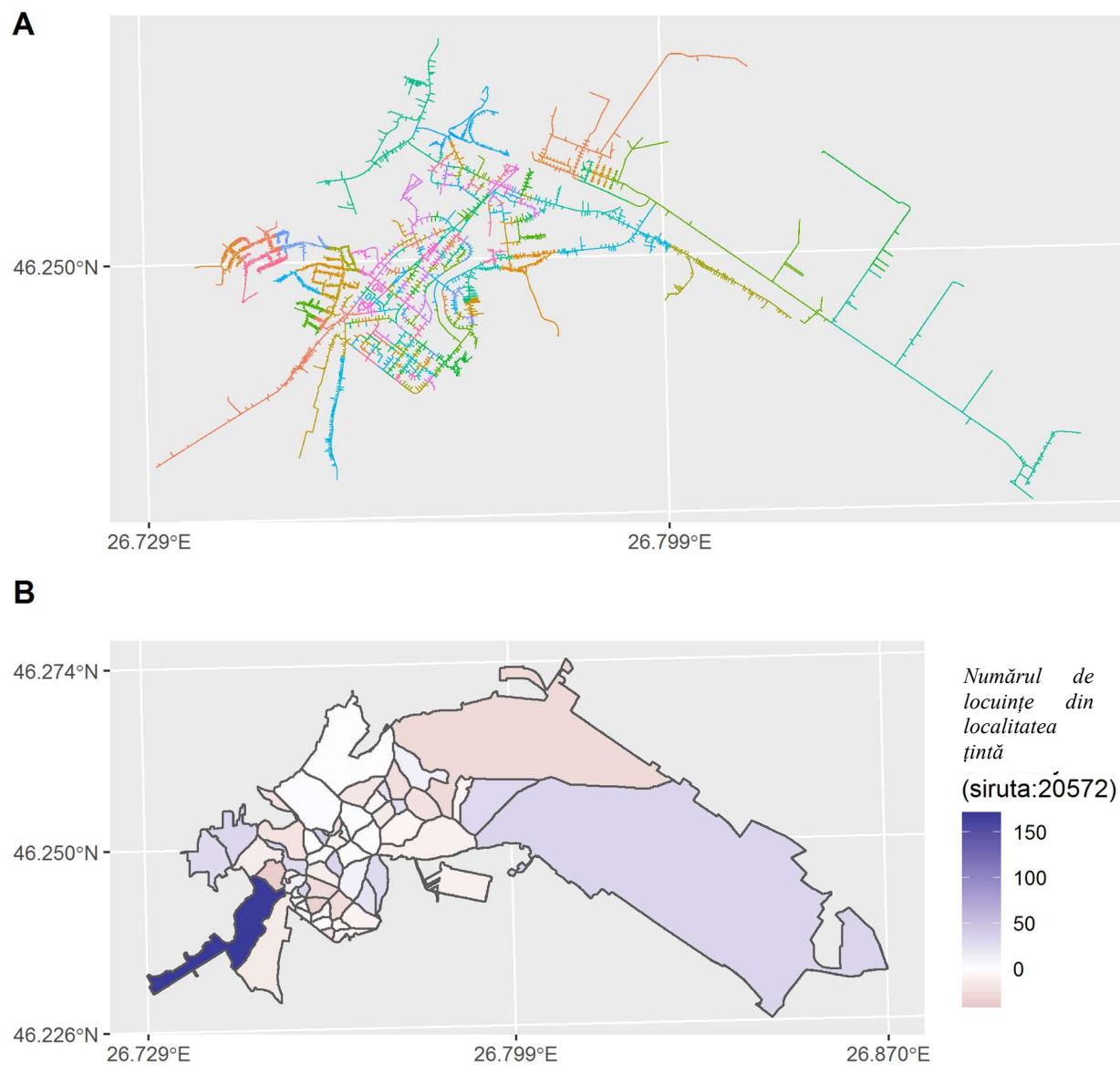


Figura 14: Hărțile (panou A-B) ilustrează rezultatul modelului (A) și sectorului (B) din Onești.

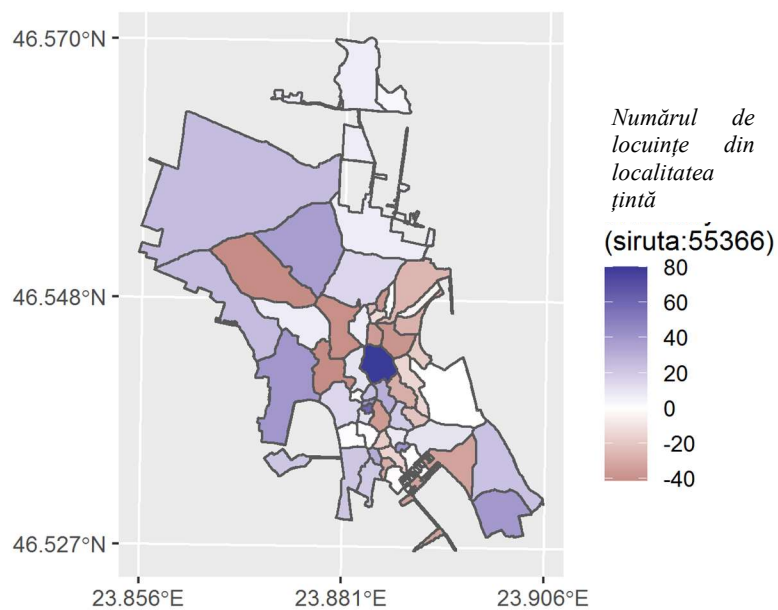
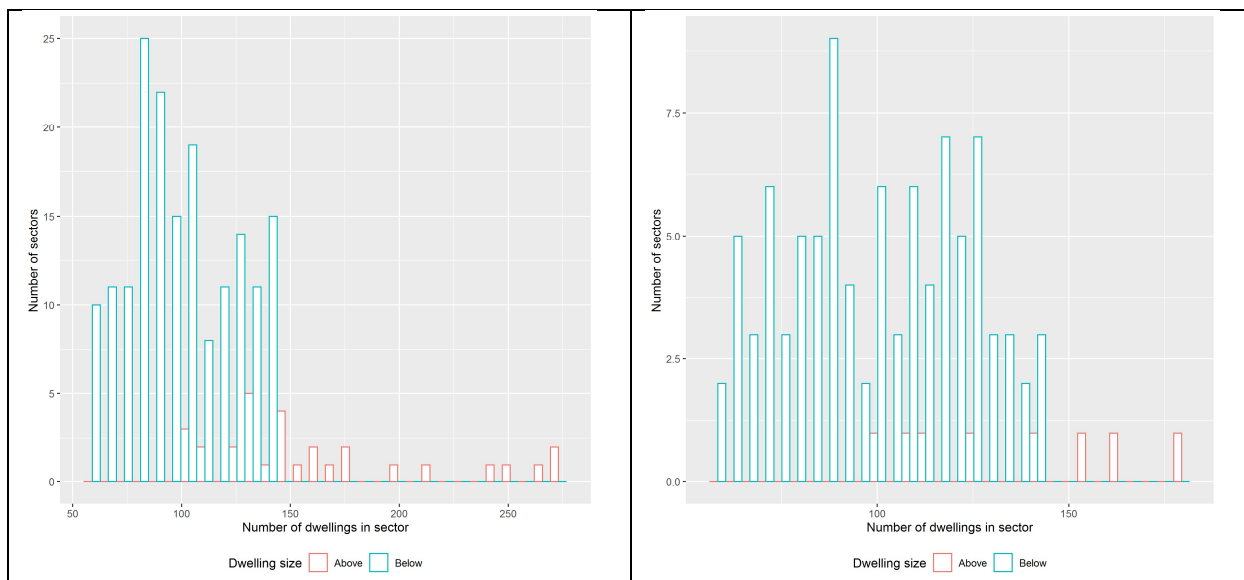
A**B**

Figura 15: Hărțile (panou A-B) ilustrează rezultatul modelului (A) și sectorului (B) din Câmpia Turzii



Legendă:

Numărul de sectoare
Numărul de locuințe din sector
Dimensiune locuință
Peste
Sub

Numărul de sectoare
Numărul de locuințe din sector
Dimensiune locuință
Peste
Sub

Figura 16: Histograma afișează numărul de locuințe dintr-un sector care include o amplasare peste și sub dimensiunea țintă în Onești (stânga) și Câmpia Turzii (dreapta).

Pentru dimensiunea țintă de 600, nu există amplasări cu locuințe care să depășească dimensiunea țintă. După cum era de așteptat, diferența dintre numărul de locuințe față de țintă este mai bine aliniată cu rezultatele rurale cu o dimensiune țintă de 100. Acest aspect oferă dovezi că numărul maxim de locuințe dintr-o amplasare în comparație cu dimensiunea țintă are o importanță semnificativă. Algoritmul se aliniază cu soluțiile cu o toleranță a populației de 0,10.

- Onești are 36 de sectoare, cu un interval de 560-653 de locuințe per sector din localitate (Figura 17 și Figura 19).
- Câmpia Turzii are 17 sectoare, cu un interval de 573-627 de locuințe per sector din localitate (Figura 18 și Figura 19).

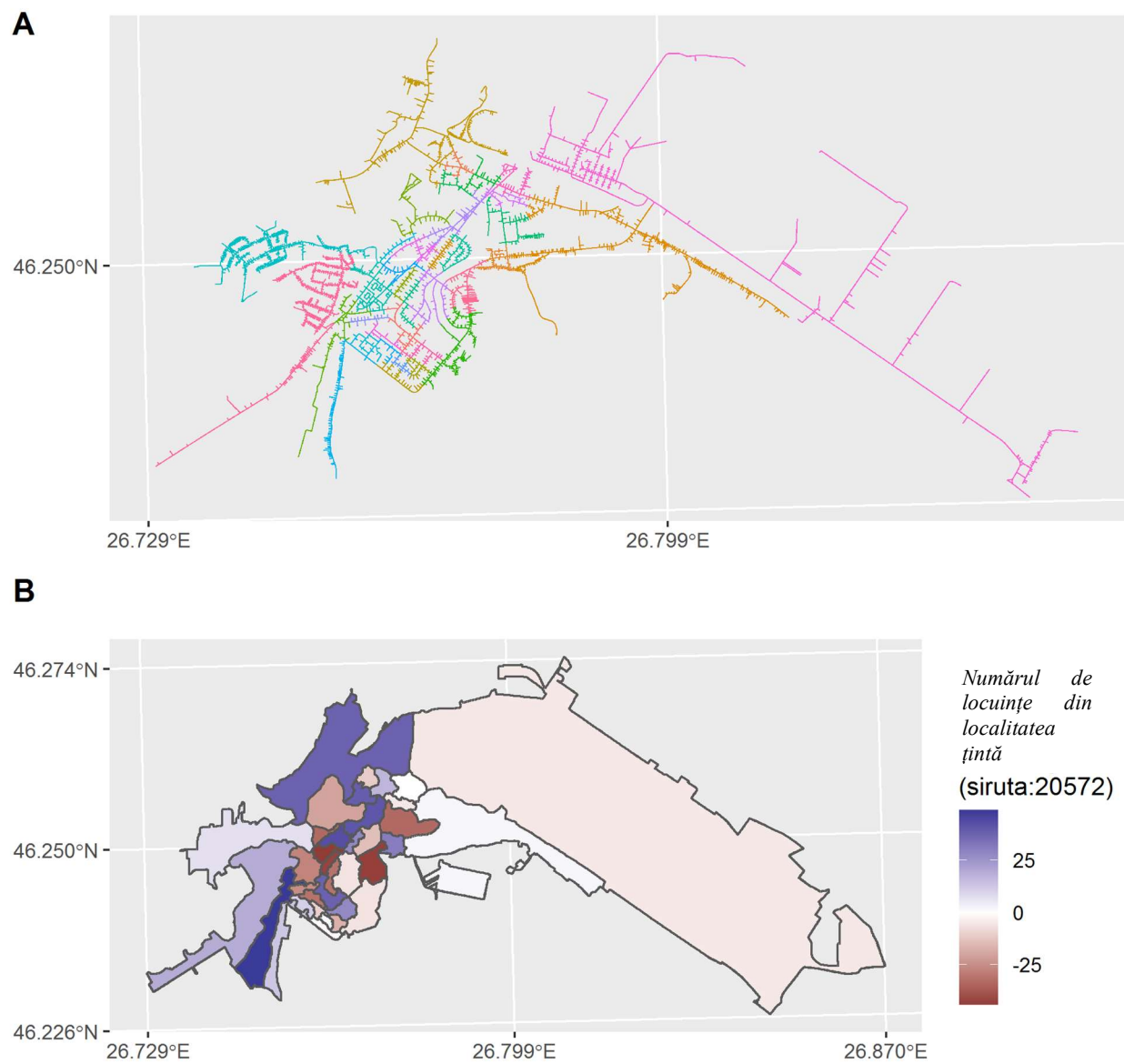
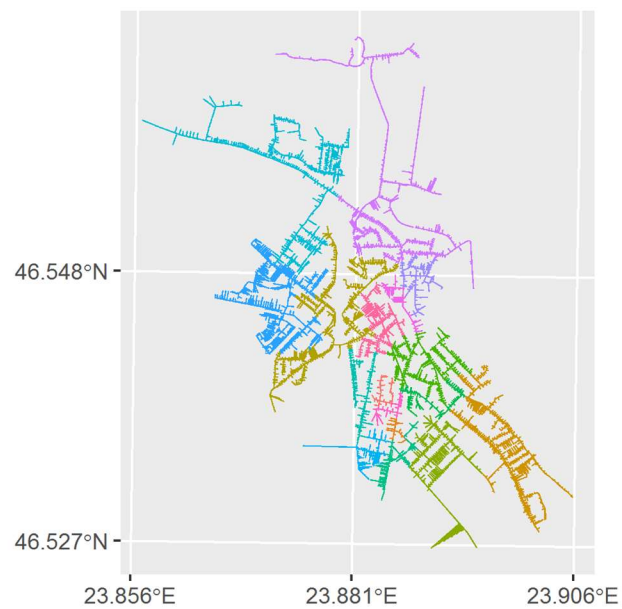
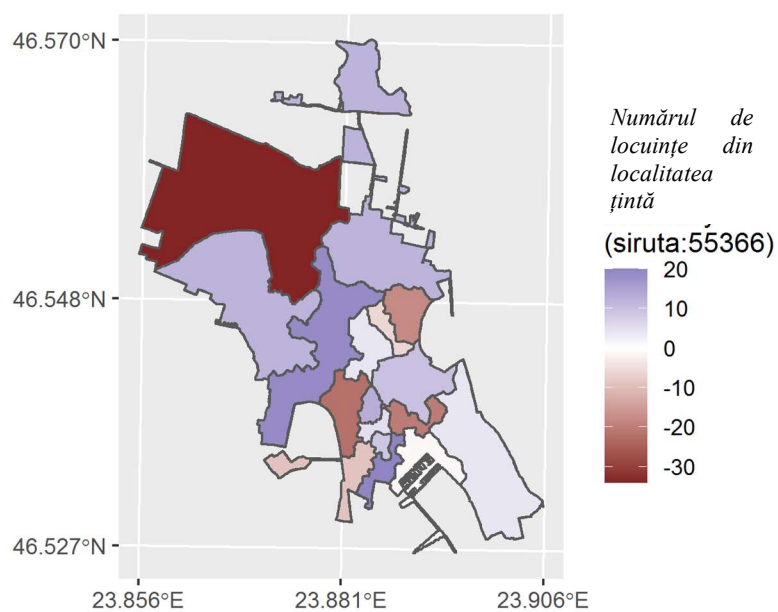


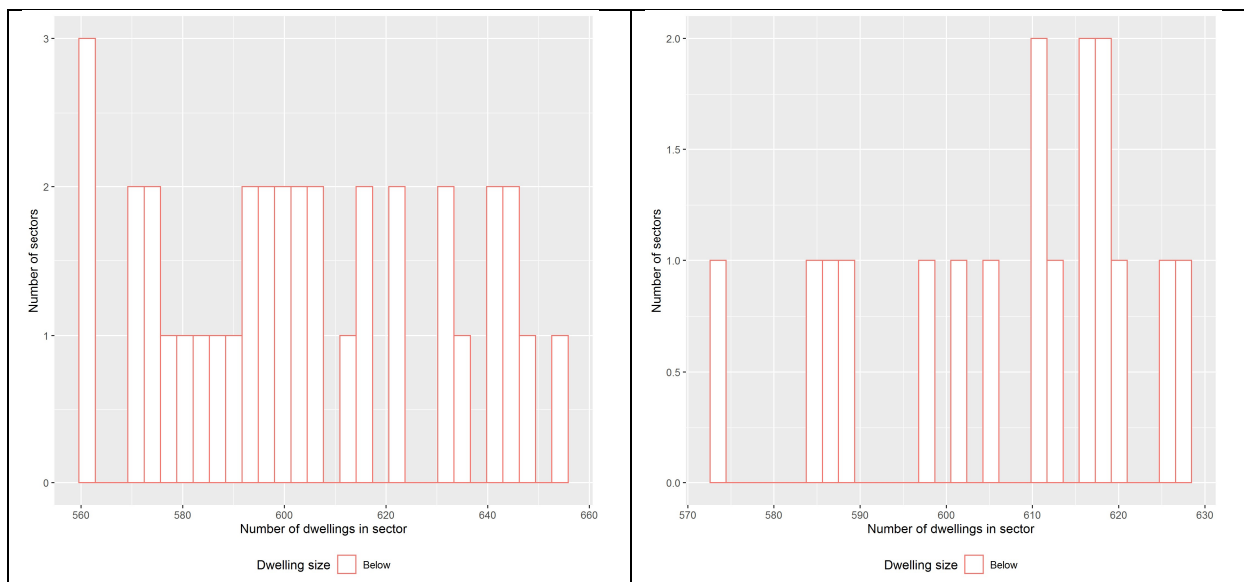
Figura 17: Hărțile (panou A-B) ilustrează modelul cu dimensiunea țintă de 600 (A) și sectoarele (B) din Onești

A**B**

Legendă:

Numărul de locuințe din localitatea țintă (SIRUTA:55366)

Figura 18: Hărțile (panou A-B) ilustrează modelul cu dimensiunea țintă de 600 (A) și sectoarele (B) din Câmpia Turzii



Legendă:

Numărul de sectoare

Numărul de locuințe din sector

Dimensiune locuință

Sub

Numărul de sectoare

Numărul de locuințe din sector

Dimensiune locuință

Sub

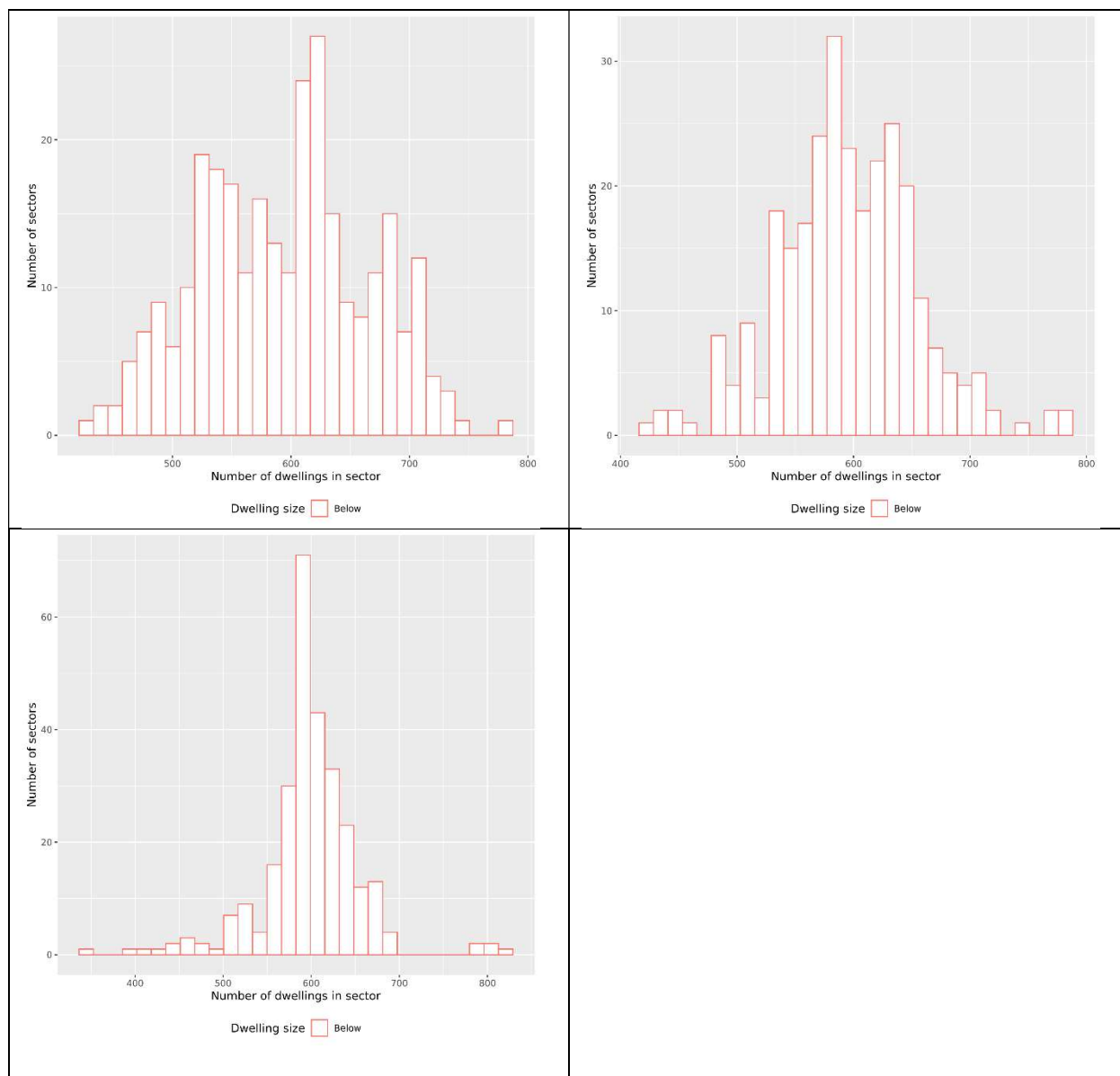
Figura 19: Histograma afișează numărul de locuințe dintr-un sector situat peste și sub dimensiunea țintă de 600 în Onești (stânga) și Câmpia Turzii (dreapta).

4.3. Eșantion urban – număr mare de localități cu locuințe

Rezultatele pentru 10 localități din numărul mare de localități cu locuințe sunt furnizate direct pe mediul INS serverul **'Data processing'** calea *arhiva2/ins_sectorization*.

Rezultatele de mai jos evidențiază un exemplu (respectiv Cluj-Napoca) din eșantionul cu un număr mare de localități cu locuințe. Rezultatele ilustrează o masă mare în cadrul unui număr +/-100 de locuințe din dimensiunea țintă (adică 600). Algoritmul care permite abaterea parametrilor de toleranță are o masă mai mare în jurul dimensiunii țintă decât rezultatele care asigură o singură toleranță în întreaga localitate (a se vedea Figura 20). Parametrii de toleranță includ: *pop_tolerance*, care este toleranța minimă, și *pop_tolerance_iter*, care este pasul de toleranță pentru fiecare iterație. Combinația acestor doi parametri produce o serie de valori care converg către o soluție fezabilă din punct de vedere tehnic. Figura 20 ilustrează importanța parametrilor de toleranță. Histograma din stânga sus arată distribuția cu o singură toleranță în întreaga localitate. Histogramele din dreapta sus arată distribuția bazată pe pasul de toleranță 0,05 (*pop_tolerance_iter* în script) evaluat la nivel de subgrup. Histogramele din stânga jos arată distribuția bazată pe pasul de toleranță de 0,02 (*pop_tolerance_iter* în script) evaluat la nivel de subgrup. O consecință a alegerii parametrilor de toleranță este creșterea timpului de procesare cu fiecare evaluare a toleranței. Evaluarea timpului de expirare este de aproximativ 10-15 minute pe iterație. Timpul de procesare pentru Figura 20a este un set de toleranțe pentru întreaga localitate până când converge o soluție. Timpul de procesare pentru Figura 20b crește din cauza necesității de a procesa mai mulți pași de timp pentru un subset de divizări. (O divizare se poate rezolva și în prima iterație și, prin

urmare, nu are nevoie de procesare ulterioară). Timpul de procesare din Figura 20c crește chiar mai mult decât în Figura 20b, cu o creștere mai mică a toleranței la 0,02 per pas de timp.



Legendă:

Numărul de locuințe din sector

Dimensiune locuință Sub

Numărul de sectoare

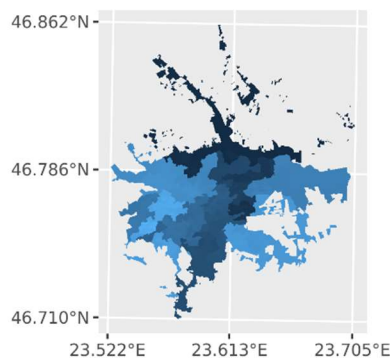
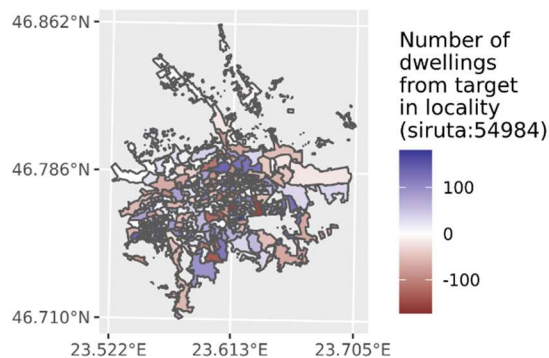
Numărul de locuințe din sector

Dimensiune locuință Sub

Numărul de locuințe din sector

Dimensiune locuință Sub

Figura 20: Numărul de locuințe din sector pentru Cluj-Napoca

A**B**

Legendă: Numărul de locuințe din țintă în localitate (siruta: 54984)

4.4. Observații

Dezvoltarea unei metode automate a necesitat pași suplimentari și specifici pentru tratare. Deși pașii manuali pot adapta abaterile datelor și metodelor, metodele automate necesită pași consecvenți (și anume, liniile directe ale ONU) pentru a crea un sector. Am efectuat pași suplimentari de tratare pentru a ajusta datele și metodele, pași care sunt rezumați mai jos.

- 1.) Nevoia de acoperire geografică actuală și completă a localităților
 - Pași pentru conectarea locuințelor în afara limitelor localității
 - Pași pentru a conecta toate zonele intravilane separate ale unei localități
- 2.) Nevoia de amplasare unică cu clădiri
 - Pași pentru combinarea amplasărilor cu clădiri pe o rază de 5 metri
 - Pași pentru a crea un identificator unic
- 3.) Nevoia de atribuire a vecinilor de la fiecare locuință
 - Pași pentru conectarea tuturor locuințelor la o stradă
 - Pași pentru conectarea tuturor străzilor dintr-o localitate

Metodologia de sectorizare se bazează fundamental pe baza de date cu locuințe (WebGIS). Pentru dezvoltarea acestei metodologii, este necesar ca înregistrările să includă un număr **pozitiv** de locuințe, coordonate distincte (fără coordonate **dublate**), o **singură înregistrare** per amplasare (fără geometrii de mai multe tipuri), amplasarea cu locuințe **în cadrul** limitei unei localități și codurile administrative derivate din definițiile spațiale, astfel încât intersecția geometrică a zonelor

administrative să fie **în concordanță** cu codurile din baza de date cu locuințe. Metoda necesită, de asemenea, ca poligonul localității să se suprapună cu amplasarea cu locuințe.

A doua observație este faptul că totalul și intervalul numărului de locuințe per amplasare prezintă o provocare pentru soluția dezirabilă sau lipsa de convergență pentru mulți algoritmi de clustere constrânși spațial, deoarece optimizarea trebuie să ia în considerare atât relațiile spațiale, cât și un echilibru al numărului de locuințe dintr-un sector. Algoritmul de îmbinare și divizare funcționează corespunzător cu aceste constrângeri în comparație cu alți algoritmi de grupare în clustere constrânși spațial, cum ar fi S'KATER (Assunção și colab. 2006), RECAP (Guo 2008) și regiunile MAX-P (Duque și colab. 2012), deoarece ceilalți algoritmi nu constrâng o valoare maximă, ceea ce poate duce la soluții cu sectoare în afara locuințelor din intervalul țintă, atât din cauza sectorizării, cât și a datelor de intrare. De asemenea, performanța modelului se îmbunătățește deoarece numărul maxim de locuințe este mai mic decât dimensiunea țintă a unui sector dintr-o localitate. Rezultatele urbane sunt mai aproape de dimensiunea țintă de 600, comparativ cu dimensiunea țintă de 100.

În al treilea rând, localitățile cu un număr mare de locuințe ar putea beneficia de delimitările cunoscute ale cartierelor cu limite georeferențiate în interiorul unei localități. Acest lucru poate facilita procesarea cu o întindere geografică mai restrânsă și mai puține amplasări și poate consolida limitele semnificative dintr-o localitate. Timpul de procesare, cerințele de memorie și convergența pot reprezenta o provocare în localitățile care au un număr mare de amplasări cu locuințe. Metodologia împarte localitatea în grupuri de amplasări cu locuințe pentru a putea efectua procesarea sub un prag prevăzut pentru sectoare (și anume, 50 de sectoare sau mai puțin). Algoritmul de schimb este utilizat cu o pondere puternică a parității, sectoarele fiind procesate separat în funcție de grup, astfel încât sectoarele din apropierea unei margini cu un grup nu includ locuințele învecinate pentru a simplifica cerințele de procesare și memorie.

În al patrulea rând, metoda necesită informații cu privire la adiacența spațială a locuințelor aflate în imediata apropiere geografică; în consecință, este necesară coerența între datele georeferențiate, astfel încât amplasările cu locuințe, drumurile georeferențiate și zonele administrative să se alinieze. Distanța minimă dintre locuințe este un parametru important, deoarece rezoluția spațială respectivă afectează viteza procesării raster, chiar dacă algoritmul profită de procesarea paralelă și îmbinarea cu fișiere temporare locale. Fiecare celulă trebuie să conțină o singură amplasare cu locuințe, astfel încât amplasările cu locuințe care se află la o distanță mică să fie agregate. O consecință a procesării raster este că rezultatul are mai multe margini în urma conversiei de la formatul raster în formatul vector. Scriptul are un parametru, *sector smoothing*, care poate executa un algoritm de omogenizare pe marginile zimțate atunci când este setat la *TRUE*; totuși, rezultatele sunt responsabilitatea utilizatorului și necesită o inspecție pentru a asigura alinierea spațială adecvată între amplasarea cu locuințe și limita atribuită sectorului.

În al cincilea rând, dimensiunea țintă ar trebui să ia în considerare intervalul numărului de locuințe. Amplasările cu locuințe pot avea sute de locuințe, iar dimensiunea țintă poate fi depășită.

În al șaselea rând, poate fi fezabil ca unele exemple să compare rezultatele acestor metode cu sectorizarea din abordările anterioare pentru a oferi o perspectivă asupra parametrizării.

În al șaptelea rând, localitățile cu amplasări cu un număr mare de locuințe pot să nu aibă soluții având în vedere parametrii de toleranță. Un exemplu scurt din 42691 este ilustrat mai jos (a se vedea Figura 21):

1. 3 locuințe cu 396 pe aceeași stradă
2. Toleranță pentru o soluție la fiecare 0,30 eșecuri (420-780)
 - a. 396 este o valoare prea mică
 - b. $396 + 396 = 792$ este o valoare prea mică
 - c. Este nevoie de o toleranță de $= 0,35$

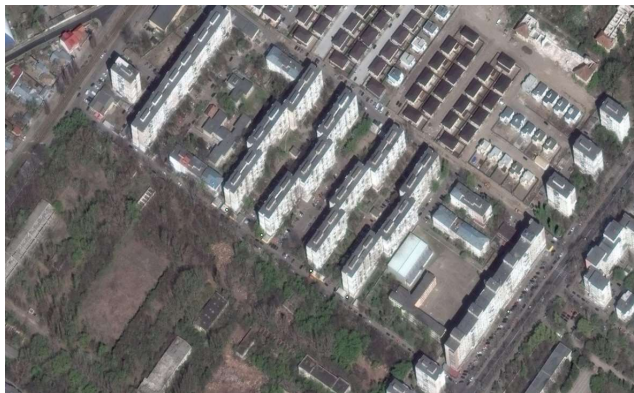


Figura 21: Imagini din satelit care arată amplasarea clădirilor

5. Recomandări

5.1. Date de intrare

Limitele depind de calitatea și disponibilitatea datelor. În delimitarea sectoarelor, algoritmul profită de amplasările cu locuințe și de rețeaua de drumuri.

- Baza de date WebGIS este esențială în aceste procese și trebuie să îndeplinească cerințele discutate în raport. Identificatorii trebuie să fie consecvenți ca nume și tip de date în diferitele seturi de date. În ceea ce privește intervalele observate în numărul de locuințe dintr-o amplasare (în special la nivel urban), este important ca alocarea amplasării să fie cât mai detaliată posibil (georeferențiată la nivelul clădirii) pentru a facilita convergența algoritmilor de grupare în clustere. În mod ideal, numărul de locuințe dintr-o amplasare este cu mult sub dimensiunea țintă a sectorului.
- Baza de date cu drumuri este, în mod ideal, pe cât de precisă posibil în acoperirea geografică, pentru a facilita alocarea blocurilor ca rezultat al unui poligon închis derivat din liniile de drumuri. De asemenea, acoperirea geografică a drumurilor ar trebui să fie cât mai completă posibil, astfel încât liniile de drumuri să se extindă până la limita localității, pentru a se asigura că este fezabil un poligon închis din punct de vedere tehnic atunci când drumurile intersectează limita. În cele din urmă, datele despre drumuri ar trebui să fie corecte din punct de vedere topologic, astfel încât liniile să se conecteze corespunzător în algoritm. Identificatorul de obiect al datelor despre străzi din baza de date cu locuințe nu furnizează amplasarea pe stradă, care este o cerință a modelului. Procedura actuală utilizează atât abordări vectoriale, cât și raster, pentru a aborda orice discontinuitate a liniilor; structura de date raster oferă acoperire continuă într-o celulă. Procesarea raster este necesară pentru a atribui vecinii. Deficiențele datelor de intrare nu permit analiza rețelei, care a fost testată. Dimensiunea celulei este mică. Deci, limita sectorului are mai multe margini ca rezultat al conversiei datelor din formatul raster în formatul vector.
- S-ar putea lua în considerare date georeferențiate suplimentare, cum ar fi căile navigabile și căile ferate, care pot facilita navigarea la fața locului, cu caracteristici naturale și artificiale ușor de identificat, și pot, de asemenea, oferi informații pentru delimitarea segmentelor prin cuantificarea asemănărilor amplasărilor cu locuințe în metodele de segmentare. Pe măsură ce aceste date sunt disponibile, se pot ajusta parametrii algoritmilor pentru a optimiza conexiunea, compactitatea geografică și paritatea numărului de locuințe.
- Este important de revizuit orice tabel cu înregistrări care nu îndeplinesc cerințele pentru a îmbunătăți caracterul complet. Dacă la etapa de sectorizare se identifică înregistrări care nu respectă cerințele și nu au fost corectate în etapa de actualizare a bazei de date, se generează un raport cu detaliile acestora de către algoritmul de sectorizare. Pe baza raportului, se vor efectua corectări ale înregistrărilor eronate, apoi se va reutiliza algoritmul de generare a sectoarelor.
- În mod similar, sectoarele pot avea limite vizavi de o stradă, totuși acest lucru poate fi armonizat și rezolvat logistic pe teren în timpul colectării datelor.

- Ca parte a angajării mai ample a gestionării datelor geospațiale, Organizația Națiunile Unite oferă un [Cadru de informații geospațiale integrate](#) pentru a asigura un ghid pentru dezvoltarea, integrarea, consolidarea și maximizarea managementului informațiilor geospațiale.⁶

5.2. Metode

Algoritmii necesită relații spațiale, iar liniile directoare prevăd că sectoarele ar trebui să fie exhaustive (acoperire geografică **completă** a întregii zone. În prezent, se utilizează datele zonei intravilane pentru a reprezenta zona din afara localității, care poate fi disjunsă cu una sau mai multe zone intravilane dintr-o localitate și datele privind amplasările cu locuințe includ deja localitatea. Atunci când acoperirea geografică nu este exhaustivă, este necesară o procedură externă pentru alocarea unei localități cu locuințe care nu se află în intravilan, disponibilă la momentul colectării datelor. Procedura trebuie să fie cuprinzătoare din punct de vedere geografic și exclusivă reciproc (să nu se suprapună), astfel încât orice amplasare din țară să aibă o singură alocare a unei localități conform liniilor directoare ale ONU (a).

- Dimensiunea țintă este un aspect important al modelului. Amplasările cu locuințe care au un număr mai mare de locuințe decât dimensiunea țintă contribuie la abaterea de la paritatea locuințelor dintr-o localitate. Aceste amplasări pot fi supuse sectorizării individuale; cu toate acestea, este posibil ca alte sectoare dintr-o soluție să nu adere la principiul contiguității spațiale. Alternativ, amplasările sunt luate în considerare în model și oferă abateri mai mari de la soluții.
- Simulările multiple oferă o distribuție de soluții pentru a selecta sectoarele cele mai dorite per localitate. Numărul de simulări și toleranța parității depind de distribuția numărului de locuințe. Parametrul de toleranță necesită o analiză atentă: toleranța prea scăzută poate să nu conducă la o soluție fezabilă; cu toate acestea, o toleranță prea mare poate duce la mai multe soluții care nu sunt de dorit. De asemenea, modelul a optimizat paritatea numărului de locuințe; cu toate acestea, compactitatea geografică este o constrângere suplimentară de luat în considerare în model, deoarece rezultatele pot oferi hărți cu navigare îmbunătățită.

Îmbinarea și divizarea au funcționat corespunzător în comparație cu alți algoritmi; cu toate acestea, pot fi adaptați și alți algoritmi de clustere constrânși spațial la fluxul de lucru actual, după cum se consideră necesar.

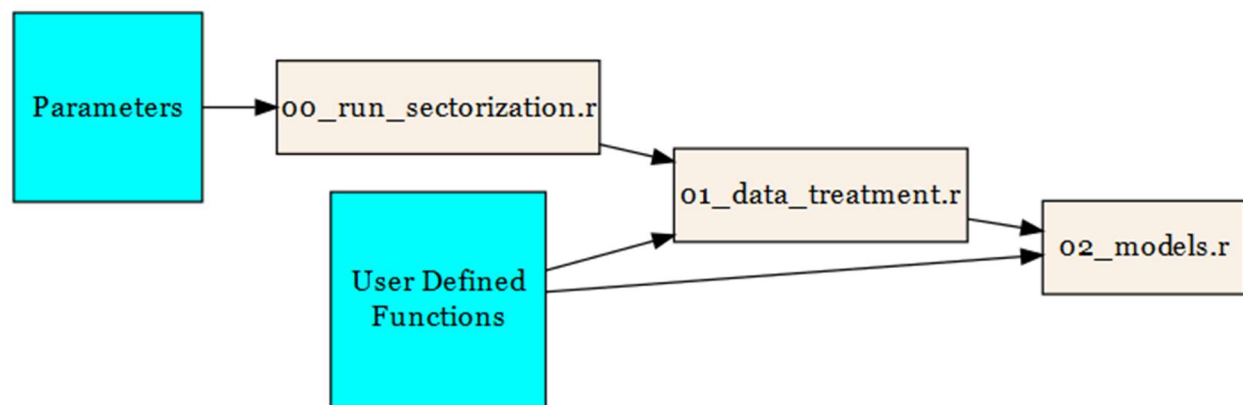
⁶ A se vedea mai multe detalii în anexă.

6. Implementare

Implementarea algoritmului necesită cunoștințe fundamentale în formatele de date geospațiale pentru a pregăti și evalua datele de intrare și pentru a aplica programarea în mediul R. Pregătirea datelor de intrare include o evaluare a acoperirii geografice, a amplasării, a valorilor și a topologiei. Scripturile R facilitează evaluarea datelor de intrare cu funcții în tratarea datelor pentru problemele întâlnite în modelarea datelor.

Pentru realizarea implementării se recomandă crearea unui login separat pentru echipa și să se configureze mediul de lucru. În acest sens, se recomandă, de asemenea, ca una dintre persoanele familiarizate cu R să însoțească echipa GIS, astfel încât orice probleme apărute, după caz, în rularea scriptului să poată fi rezolvate imediat.

Pe directorul „root”, trei scripturi R facilitează implementarea metodei descrise mai sus folosind Rstudio pe serverul INS. Un script de pornire ușurează implementarea prin încărcarea pachetelor R relevante, configurarea directoarelor, a parametrilor, a sarcinilor care acceptă automat funcțiile definite de utilizator în scripturi și execută secvențial metoda pe tratamentul datelor și model, care reprezintă al doilea, respectiv al treilea script R (a se vedea diagrama fluxului). Funcțiile definite de utilizator sunt scrise cu comentarii în linie ale scriptului care pot fi executate în Rmarkdown folosind pachetul „spin”.



Legendă:

Parametri

Funcții definite de utilizator

6.1. Pornire

Fișierul de pornire include toate informațiile pentru a începe procesul. Scriptul profită de o bază prolifică de utilizatori și încarcă automat peste 25 de pachete relevante. Scriptul de pornire identifică directoare, inclusiv datele de intrare, tratarea și rezultatele modelului. Acest lucru permite unui utilizator să specifice directorul webgis.gdb care urmează aceeași structură de denumiri de straturi și variabile. Scriptul încarcă, de asemenea, funcțiile de sprijin definite de utilizator, urmate de executarea secvențială a tratării și modelului datelor.

Fișierul de pornire include parametri importanți. Dimensiunea țintă a numărului de locuințe dintr-o localitate poate varia, totuși este setată la 600 pe baza datelor [metodologiei de sectorizare](#) publicate (*dimensiune țintă* <- 600). Localitatea sau localitățile de procesat poate/pot fi configurate cu codul SIRUTA.

6.2. Etape suplimentare și specifice pentru tratarea datelor și model

Tratarea datelor și modelul sunt executate secvențial. Fiecare script creează un folder cu localități cu rezultatele fiecărui pas, împreună cu un fișier de metadate și fișiere jurnal. Fișierul jurnal înregistrează codul localității și timpul de procesare. Fișierul de eroare înregistrează codul localității și câteva cuvinte pe funcția cu eroarea (de exemplu, localitatea nu are un fișier de delimitare corespunzător).

Acoperirea geografică actuală și completă a localităților este esențială în dezvoltarea unei soluții cuprinzătoare și oportune. Pașii necesari pentru integrarea datelor sunt proporționali cu complexitatea datelor. Lipsa datelor cuprinzătoare din punct de vedere geografic a necesitat pași pentru conectarea tuturor locuințelor din afara limitelor localității care există în baza de date. Fiecare amplasare nouă cu locuințe va avea nevoie de o atribuire a unei localități, iar limita localității trebuie să se suprapună cu amplasarea. În mod similar, pașii tratează zonele intravilane pentru a asigura o conexiune și astfel limita localității îndeplinește cerințele tuturor amplasărilor cu un număr pozitiv de locuințe care se află în limita unei localități. O extensie a acestui principiu este de a oferi o limită la nivel de localitate care are o acoperire geografică completă pe întreg teritoriul.

Baza de date cu locuințe este o bază de date bogată care oferă coordonate cu numărul de locuințe dintr-o amplasare. Pașii de tratare au inclus identificatori unici, excluderea valorilor zero sau negative și agregarea amplasărilor în unele cazuri pentru a facilita procesarea. Modelul folosește o structură de date raster tip grid care îndeplinește cerințele de acoperire geografică completă în localitate. Mărimea celulei influențează direct timpul de procesare în raport cu numărul și amplasarea celulelor corespunzătoare. O funcție examinează distanța cea mai apropiată dintre amplasările cu locuințe și agregă numărul de locuințe în instrumentele de măsurare specificate.

Atribuirea vecinilor de la fiecare locuință la toate celelalte locuințe necesită integrarea corespunzătoare a bazei de date cu locuințe, a bazei de date cu străzi și a limitei localității. Conexiunea locuinței cu amplasarea străzii este automatizată, deoarece datele privind numărul străzii nu sunt complete și oferă o amplasare exactă pe stradă. O hartă a conexiunii stradale oferă utilizatorului informații despre distanța necesară pentru a se conecta la datele existente despre străzi. Discrepanțele mari identifică zone pentru a examina actualitatea și caracterul complet al datelor despre străzi. Pașii necesari pentru conectarea tuturor străzilor dintr-o localitate abordează provocările din datele despre străzi care nu sunt conectate digital (corecte din punct de vedere topologic).

Tratarea datelor include mai mulți pași și se produc fișiere intermediare în cazul în care datele nu îndeplinesc cerințele necesare (a se vedea graficele de mai jos din Figura 22, Figura 23 și Figura 24).

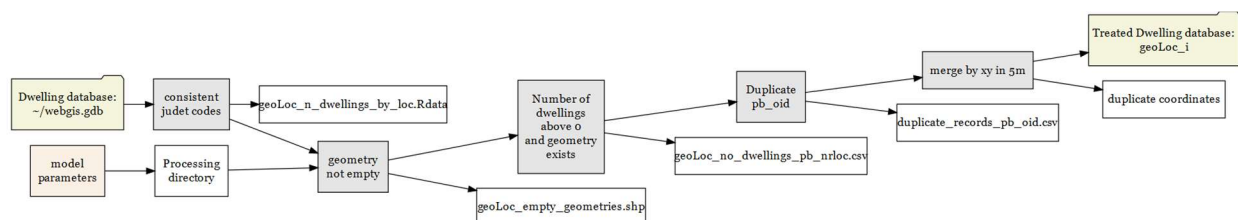


Figura 22: Diagrama tratării locuințelor

Legendă:

Baza de date cu locuințe

Coduri consecvente județ

Numărul de locuințe peste 0 și geometria există

Duplicat

Fuziune prin xy în 5m

Baza de date cu locuințe tratate

Parametri model

Director de procesare

Geometrie care nu este ocupată

Coordonate duplicat

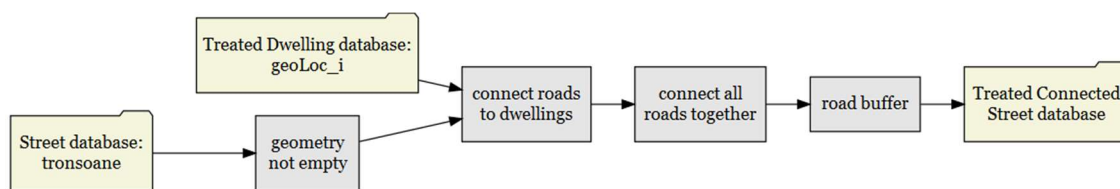


Figura 23: Diagrama tratării străzii cu locuințe

Legendă:

Bază de date străzi: tronsoane

Baza de date cu locuințe tratate:

Conectează drumurile cu locuințele

Conectează toate drumurile împreună

Nod rutier

Bază de date străzi conectate tratate

Geometrie care nu este ocupată

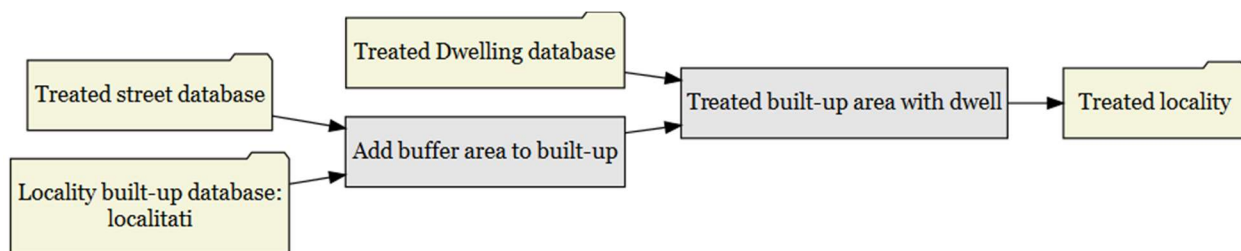


Figura 24: Diagrama tratării zonei cu localități

Legendă:

Bază de date străzi tratate

Bază de date locuințe tratate

Zonă intravilană cu locuințe tratată

Localitate tratată

Bază de date localități intravilane: localități

Adaugă nod în zona intravilană

Secțiunea model include mulți pași (a se vedea diagrama din Figura 25). Rezultatele modelului includ un shapefile al limitei de sector, o histogramă și o parcelă de hartă.

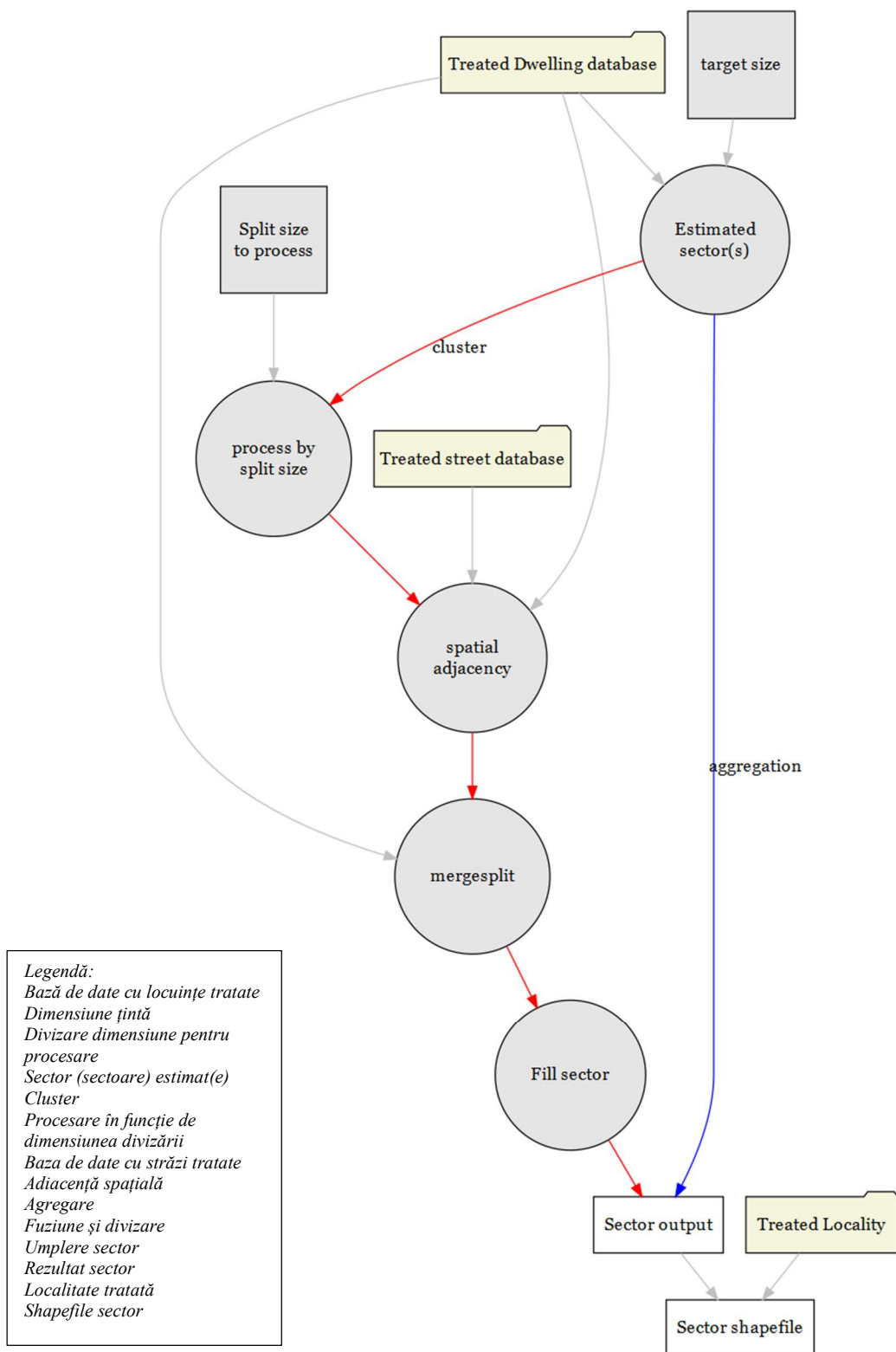


Figura 25: Diagrama pașilor de model

7. Managementul datelor geospațiale în contextul RPL2021

„Managementul datelor este procesul de încorporare, stocare, organizare și întreținere a datelor create și colectate de o organizație” (Stedman & Vaughan, 2019).

În RPL2021, există trei tipuri principale de date din resurse geospațiale care vor fi utilizate în proces, precum și o sursă de date opțională:

- i. Denumirea (și numerele) străzilor, satelor, localităților (comune și orașe), municipiilor și județelor
- ii. Amplasările clădirilor, inclusiv numărul de locuințe din interiorul clădirilor
- iii. În funcție de sectorizare: limitele sectoriale
- iv. (opțional: baza de date de fotografii aeriene)

Cu excepția sursei de date iv., toate sursele de date vor fi stocate într-o bază de date(PostGIS⁷) PostgreSQL⁸, astfel cum este furnizată de INS. Această bază de date urmează, de asemenea, structura de codare idiosincronică a INS, care, prin urmare, nu va fi comentată în acest raport. Presupunem că aceste intrări sunt întreținute corespunzător pe partea INS.

În plus, datele de intrare necesită, de asemenea, o calitate suficientă pentru a fi utilizate în procesul de segmentare, astfel cum s-a discutat mai sus, precum și în procesul de (re)identificare a respondenților, utilizat ulterior în verificările de acoperire. Se presupune că baza de date furnizată de INS îndeplinește aceste cerințe.

Soluțiile furnizate aici sunt scripturi scrise în R, care pot fi executate pe sistemul INS și care pot utiliza

- a) Pachetul API Survey Solutions (<https://github.com/michael-cw/SurveySolutionsAPI>)
- b) Pachetul Rpostgres (<https://cran.r-project.org/web/packages/RPostgres/>)
- c) Pachetul DBI (<https://cran.r-project.org/web/packages/DBI/>).

Acestea sunt cele 3 pachete principale utilizate în transportul și stocarea datelor. Deoarece vor fi realizate și unele activități transformative, scripturile furnizate folosesc și pachete generale R, precum și două pachete geospațiale foarte importante, și anume

- a) Pachetul SF (<https://cran.r-project.org/web/packages/sf/>); și
- b) Pachetul stars (<https://cran.r-project.org/web/packages/stars/>)

⁷ <https://postgis.net/>

⁸ <https://www.postgresql.org/>

7.1. Utilizarea datelor geospațiale în RPL2021

Un recensământ constă din trei etape și, prin urmare, un recensământ georeferențiat ar trebui să utilizeze aceste date pe tot parcursul procesului. În RPL2021, Figura 26 prezintă etapele și diferitele aplicații în care datele geospațiale vor fi relevante.

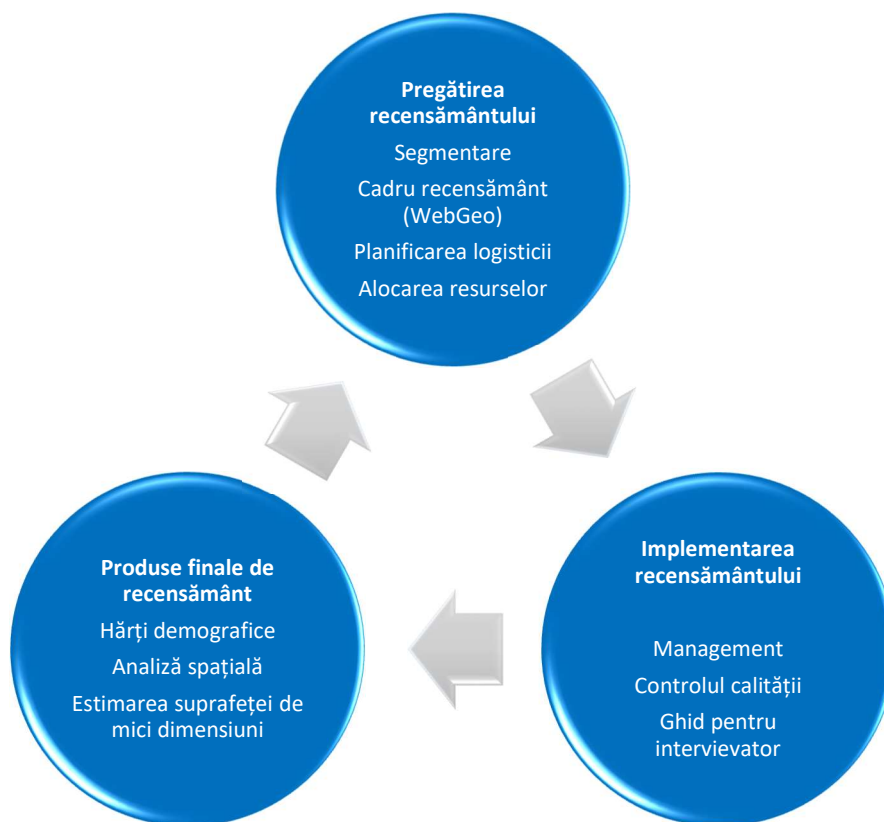


Figura 26: Cele trei etape principale ale unui recensământ

Important de remarcat aici este că „ciclul de viață” de mai sus este un ciclu literal, deoarece fiecare element susține îmbunătățirea succesorilor săi.

Recensământul în sine va face o utilizare inovatoare a datelor georeferențiate, în special în mai multe privințe:

- Automatizarea segmentării (a se vedea secțiunea anterioară).
- Utilizarea datelor de teledetecție (și a altor date disponibile georeferențiate) pentru testarea acoperirii populației
- Navigarea recenzorilor CAPI
- (Limitele segmentelor pentru tabloul de bord al hărții CAPI)
- Vizualizarea rezultatelor recensământului.

Prin urmare, formatul de bază de date selectat trebuie să poată face față acestor scenarii diferite. Cu alegerea PostgreSQL ca format de bază de date, acest lucru este totuși destul de simplu de realizat, deoarece ambele au implementări API în R (setul de funcții I/O pentru R se află în Anexa 0) și Python. Pentru a putea exporta și importa date geospațiale din și în Survey Solutions, baza de

date trebuie să fie totuși accesibilă pentru celelalte aplicații, care sunt fie la sediul STS, fie la cel al INS. Detaliile trebuie să fie conforme cu configurarea hardware finală. Dacă totuși serverele se află într-o amplasare geografică diferită, atunci VPN-ul în rețea dedicată ar fi optim, iar soluțiile sub-optime trebuie evaluate pentru a vedea dacă sunt capabile din punctul de vedere al vitezei de transfer de date și al volumului datelor.

7.2. Armonizarea nomenclurii amplasărilor și străzilor

O componentă importantă în procesul recensământului este georeferențierea tuturor unităților de recensământ (și anume, clădirile, locuințele, gospodăriile și persoanele) și aplicarea unei nomenclaturi spațiale consistente folosind identificatori imbricați.

Procesul principal de înregistrare a recensământului se bazează în mare măsură pe legătura dintre o bază de date (armonizată) de denumiri de străzi legată de coordonatele sale geografice. Dacă este posibil, nomenclatorul corespunzător al denumirilor de străzi reflectă structura imbricăată prezentată în Figura 27. Atunci când utilizați o bază de date PostgreSQL, această structură imbricăată poate reflecta și structura bazei de date. Funcțiile necesare din anexă permit implementarea acestei structuri a bazei de date, dacă se dorește.

Figura 27 arată structura imbricăată a unităților de recensământ⁹. Deși este teoretic posibil ca o unitate de recensământ să fie membră a două unități în același timp, practic poate fi rezultatul fie al unei clasificări eșuate, fie al unei alocări eșuate în ceea ce privește conceptul de rezidență. Din perspectiva bazei de date, totuși, nu contează, deoarece pentru toate motivele enunțate mai sus, este necesară o structură unică.

⁹ Deși adresele nu sunt unități de recensământ în sine, ele ar trebui totuși tratate ca atare, având în vedere structura idiosincronică (adică Scara 1,2,3, 40, a, b, c) și rolul important pe care îl joacă în identificarea și georeferențierea tuturor celorlalte unități de recensământ de mai jos.

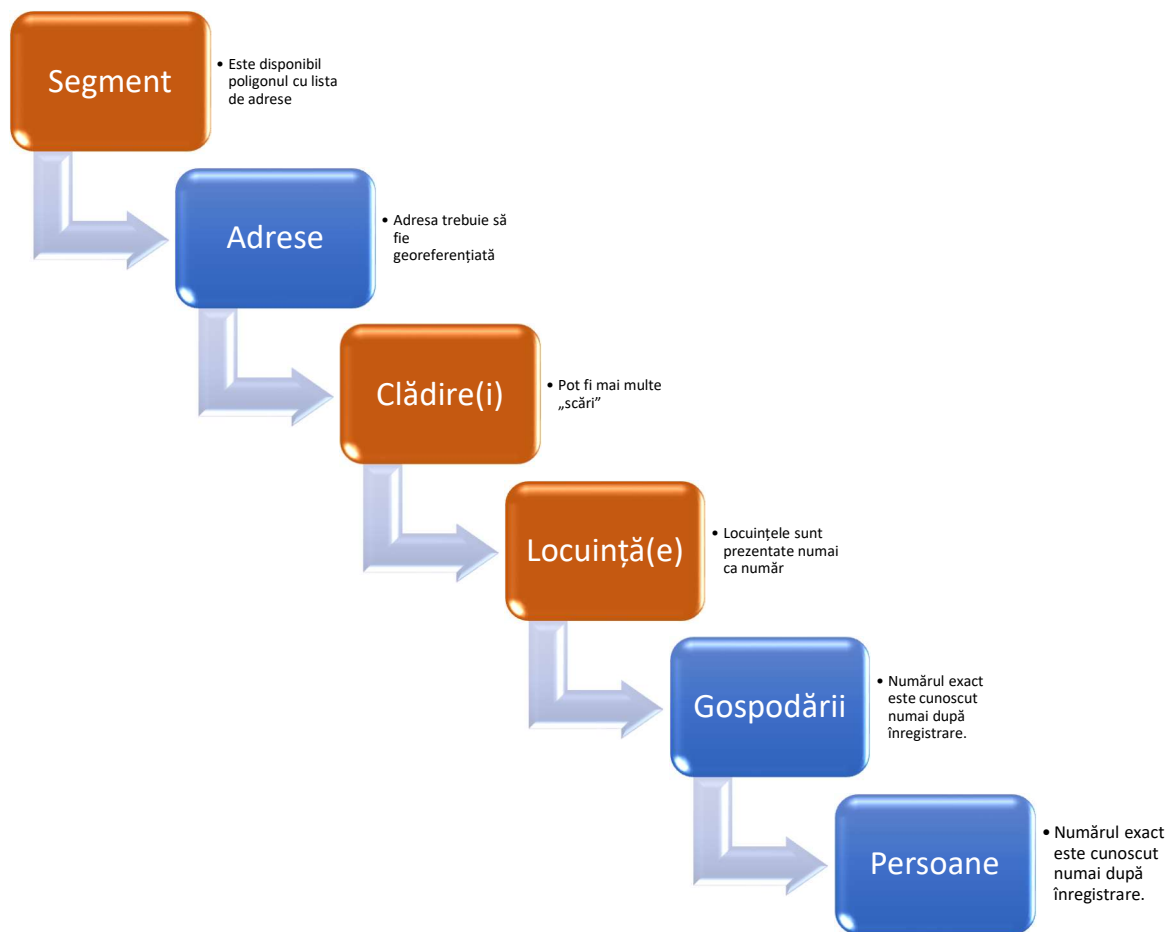


Figura 27: Structura imbricată a unităților de recensământ

În RPL2021, recensarea persoanelor a avut loc prin portalul de autoînregistrare. Pentru persoanele care nu s-au auto-recenzat, datele au fost colectate prin interviuri față în față prin intermediul recenzorilor.

Anexe

7.3. Cod de segmentare

Codul de segmentare are un scripturi de pornire și două scripturi principale pentru a funcționa în ordine:

- (1) 00_run_sectorization.r
- (2) 01_data_treatment.r ;
- (3) 02_model.r

Codul sursă al aplicației este furnizat în format electronic, arhivat împreună cu acest raport „Rezultatul 7c – Anexa 7.1 Sectorization_code.zip”

7.4. Anomalii WebGIS

Baza de date WebGIS implică eforturi uriașe ale INS pentru a reprezenta întrebuințarea și acoperirea locuințelor și străzilor, împreună cu zonele intravilane la nivel de 2019, însă, în câteva cazuri, nu îndeplinește cerințele furnizate de noi. Există oportunități specifice pentru a îmbunătăți calitatea și acoperirea datelor de intrare, care pot fi aplicate și în alte fluxuri de lucru.

7.4.1. Date complete

Modelul necesită ca localitățile să aibă același (cod) SIRUTA în cele trei seturi de date: locuințe, străzi și zonă intravilană. Din eșantionul mare de localități urbane, două localități nu au limite de zonă intravilană corespunzătoare: Brașov (40205) și Bacău (20304). Deși rezultatele sunt procesate și furnizate la nivel de localitate, un set de date auxiliare cu zone geografice definite mai mici ale localităților mari ar facilita procesarea în zone semnificative de cartier mai mici decât sectorizarea localității în două etape: mai întâi întreaga localitate în subgrupuri și apoi subgrupul în sectoare.

7.4.2. Actualitate și acoperire (a datelor)

O funcție definită de utilizator estimează conexiunea străzilor atunci când se omite posibilitatea de a executa modelul, însă acesta nu este cazul ideal. Conexiunile omise pot apărea din lipsa datelor topologice, dar pot reprezenta și lipsa unor date actuale și a acoperirii complete în raport cu baza de date cu locuințe. Localitățile urbane mari au necesitat multe străzi (de exemplu, 430 în Cluj-Napoca, 99 legături pentru 179178, 196 pentru 95079 etc.) Deși ponderea străzilor poate fi mică (sute din zeci de mii de linii stradale), această anomalie a datelor mărește timpul de procesare a datelor (executarea funcției de tratare a datelor), precum și limitează modelul în mod artificial pentru a oferi rezultate optime.

La o inspecție mai atentă a unui eșantion de cazuri, actualitatea și acoperirea datelor despre străzi pot duce la sectoare suboptime. De exemplu, un sector cu un număr redus de locuințe nu are străzile identificate în datele din imagini. Această lipsă de conexiuni ale străzilor impune artificial

Legendă.
Locuințe
Sectoare
Străzi

În 2019, Națiunile Unite și Banca Mondială au lansat Cadrul de informații geospațiale integrate ([IGIF](#)), care oferă un ghid pentru națiuni pentru a-și dezvolta infrastructura și serviciile geospațiale. IGIF cuprinde trei părți ca documente separate, dar conectate: Partea 1 este un cadru strategic general; Partea 2 este un Ghid de implementare; iar Partea 3 este un Plan de acțiune la nivel de țară. Cele trei părți includ un cadru cuprinzător de informații geospațiale integrate, care deservește nevoilor unei țări în abordarea factorilor economici, sociali și de mediu, care depind de informațiile despre amplasare într-o lume în continuă schimbare. Ghidul de implementare comunică utilizatorului ceea ce este necesar pentru stabilirea, implementarea, consolidarea, îmbunătățirea și/sau menținerea unui sistem și capacitățile naționale de management al informațiilor geospațiale.

IGIF se concentrează pe informații despre amplasare care sunt integrate cu orice alte date semnificative pentru a rezolva problemele societale și de mediu, acționează ca un catalizator pentru creșterea economică și dezvoltarea oportunităților și pentru a înțelege și a beneficia de prioritățile de dezvoltare ale unei națiuni și de Obiectivele de dezvoltare durabilă.

7.7. Referințe

- Assunção, R. M., Neves, M. C., Câmara, G., & da Costa Freitas, C. (2006). Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees. *International Journal of Geographical Information Science*, 20(7), 797-811.
- Carter, D., Hunter, Z., Herschlag, G., & Mattingly, J. (2019). A merge-split proposal for reversible Monte Carlo Markov chain sampling of redistricting plans.
- Chen, J., & Rodden, J. (2013). Unintentional gerrymandering: Political geography and electoral bias in legislatures. *Quarterly Journal of Political Science*, 8(3), 239-269.
- DeFord, D., Duchin, M., & Solomon, J. (2019). Recombination: A family of Markov chains for redistricting. *arXiv preprint arXiv:1911.05725*.
- Duque, J. C., Anselin, L., & Rey, S. J. (2012). The max-p-regions problem. *Journal of Regional Science*, 52(3), 397-419.
- Fifield, B., Imai, K., Kawahara, J., & Kenny, C. T. (2020a). The essential role of empirical validation in legislative redistricting simulation. *Statistics and Public Policy*, 7(1), 52-68.
- Fifield, B., C. T. Kenny, C. McCartan, A. Tarr, and K. Imai (2020b). *redist: Simulation Methods for Legislative Redistricting*. Available at The Comprehensive R Archive Network (CRAN). URL <https://CRAN.R-project.org/package=redist>.
- Guo, D. (2008). Regionalization with dynamically constrained agglomerative clustering and partitioning (REDCAP). *International Journal of Geographical Information Science*, 22(7), 801-823.
- Qader, S., Lefebvre, V., Tatem, A., Pape, U., Himelein, K., Ninneman, A., ... & Bird, T. Semi-automatic mapping of pre-census enumeration areas and population sampling frames. *Humanities and Social Sciences Communications*, 8(1), 1-14.
- Rousseeuw, P. J., & Kaufman, L. (1990). Finding groups in data. *Hoboken: Wiley Online Library*, 1.
- Stedman, C., & Vaughan, J. (2019). What is data management and why is it important?, Whitepaper. URL <https://searchdatamanagement.techtarget.com/definition/data-management>
- United Nations. (2017). Principles and Recommendations for Population and Housing Censuses, Revision 3 United Nations. Available at : <https://www.un-ilibrary.org/content/books/9789210573948>



Competența face diferența!

Proiect selectat în cadrul Programului Operațional Capacitate Administrativă, cofinanțat de Uniunea Europeană din
Fondul Social European