

Svojstva podataka o raznolikosti vrsta prikupljenih ljudskim opažanjem iz baze Global Biodiversity Information Facility (GBIF)

Šapina, Ivan

Master's thesis / Diplomski rad

2022

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:217:025890>

Rights / Prava: [In copyright](#)/[Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-11-19**



Repository / Repozitorij:

[Repository of the Faculty of Science - University of Zagreb](#)



Sveučilište u Zagrebu
Prirodoslovno-matematički fakultet
Biološki odsjek

Ivan Šapina

**Svojstva podataka o raznolikosti vrsta
prikupljenih ljudskim opažanjem iz baze
Global Biodiversity Information Facility
(GBIF)**

Diplomski rad

Zagreb, 2022

Ovaj rad je izrađen u Laboratoriju Evolucije na Zoologijskom zavodu Prirodoslovno-matematičkog fakulteta u Zagrebu, pod mentorstvom prof. dr. sc. Damjana Franjevića, te komentorstvom dr. sc. Josipa Skeje. Rad je predan na ocjenu Biološkom odsjeku Prirodoslovno-matematičkog fakulteta Sveučilišta u Zagrebu radi stjecanja zvanja Magistar struke znanosti o okolišu.

Zahvale i posveta

Zahvaljujem se mentoru prof. dr. sc. Damjanu Franjeviću na bezuvjetnoj podršci koju mi je pružio tokom cijelog postupka izrade, prijave i obrane diplomskog rada. Bez njegovih čeličnih živaca i riječi punih potpore ovoj agoniji ne bi bilo kraja. Zahvaljujem se komentoru dr. sc. Josipu Skeji na pomoći sa prijavom teme, na „*big words*“ i što je sa mnom proživio stresne trenutke. Zahvaljujem se dr. sc. Franu Rebrini na provjeri prijave teme. Zahvaljujem se prof. dr. sc. Ivani Maguire na literaturi koja mi je otvorila oči kako u ekologiji postoji toliko zanimljivih ideja i misli za koje nisam niti znao kako postoje i na savjetima kada nisam znao kuda dalje. Zahvaljujem se prijateljici i neslužbenoj mentorici iz programiranja mag. biol. mol. Dori Zvjerković bez čijih poduka ovaj rad ne bih mogao izvesti u ovakvom obliku. Zahvaljujem se prijateljima Ivi Džakuli, mag. oecol. Janu Regulu i Đurđici Majetić na lektoriranju i podršci tokom izrade diplomskog rada. Zahvaljujem se svojim prijateljima: Andrijani i Katarini Štulić, Ivi Čupić, Katarini Radman, Luciji Perković, Kim Vučinić, Đini Noli, Mariti Cvitanović, Marti Ivanović, Filipu i Valeriji Plazonić i Andreji Lucić i svima drugima koji su mi ovo razdoblje života učinili lijepim iskustvom.

Želim se posebno zahvaliti svojoj djevojci Anđeli koja je sve ovo vrijeme bila uz mene i pružila mi potporu i ljubav.

Neizmjerne sam zahvalan svojoj obitelji, mami, tati, babi, nonotu, sestri i njenom suprugu što su mi pružili sigurnost tokom studiranja. Njima posvećujem ovaj rad.

TEMELJNA DOKUMENTACIJSKA KARTICA

Sveučilište u Zagrebu
Prirodoslovno-matematički fakultet
Biološki odsjek

Diplomski rad

Svojstva podataka o raznolikosti vrsta prikupljenih ljudskim opažanjem iz
baze Global Biodiversity Information Facility (GBIF)

Ivan Šapina

Rooseveltov trg 6, 10000 Zagreb, Hrvatska

Ljudsko opažanje stvara velike količine podataka o raznolikosti vrsta. Tako prikupljeni podatci sve češće se koriste u ekološkim istraživanjima. Poznate su mnoge pristranosti ovih podataka i načini ispravljanja istih, ali do sada ljudsko opažanje nije istraženo kao metoda prikupljanja podataka o raznolikosti svih dostupnih svojti. Cilj ovog rada je zaključiti svojstva ljudskog opažanja kao metode uzorkovanja raznolikosti vrsta analiziranjem podataka koje stvara, pod pretpostavkom kako pristranosti podataka bolje opisuju kolektivno ponašanje ljudi prilikom opažanja nego samu raznolikost vrsta. Podatke sam preuzeo s javno dostupne mrežne baze Global Biodiversity Information Facility za odabrane države u periodu od 22 godine. Podatke sam analizirao pomoću mjere bogatstva vrsta, heterogenosti i prostorne raspodjele. Za sve metode postoje teoretske vrijednosti i svojstva koje postižu u hipotetskim situacijama i time sam usporedio dobivene rezultate. Rezultati su pokazali da podatci imaju logaritamski normalnu raspodjelu vrsta i korisnika, broj vrsta ovisi o broju opažanja i kako ljudsko opažanje vrsta nije nasumično. Također su podatci pokazali značajnu prostornu grupiranost. Zaključio sam kako podatci koje ljudsko opažanje prikuplja najbolje opisuje interakciju ljudi s organizmima za ona područja gdje čovjek obitava. Takvi podatci svoj najveći potencijal ostvaruju pri opisivanju ljudskog ponašanja unutar okoliša.

Ključne riječi: građanska znanost, raznolikost, ponašanje, raspodjela, akumulacija, rarefakcija (63 stranice, 27 slika, 3 tablice, 82 literaturnih navoda, jezik izvornika: hrvatski)

Rad je pohranjen u Središnjoj biološkoj knjižnici

Mentor: prof. dr. sc. Damjan Franjević

Komentor: dr. sc. Josip Skejo

Ocjenitelji:

Prof. dr. sc. Sven Jelaska

Prof. dr. sc. Ivana Maguire

Doc. dr. sc. Zorica Petrincec

Doc. dr. sc. Luka Valožić

Zamjena: Prof. dr. sc. Darko Tibljaš

Rad prihvaćen: 08. 09. 2022.

BASIC DOCUMENTATION CARD

University of Zagreb
Faculty of Science
Department of Biology

Master thesis

Properties of Species Diversity Data Collected by Human Observation from the Global Biodiversity Information Facility (GBIF) Database

Ivan Šapina

Rooseveltova trg 6, 10000 Zagreb, Croatia

Human observation is creating a large amount of species diversity data. Data collected in this manner is being used more frequently in ecological research. Many biases are known for this type of data and methods for correcting them, but still there is no research which investigates human observation as a method for collecting species diversity data especially for all available taxa. Goal of this thesis is to deduce features of human observation as a method for sampling species diversity, by analyzing data which it generates; under the assumption that data bias better describes collective human behavior while observing than diversity itself. Data was downloaded from the open access internet database Global Biodiversity Information Facility for selected states for a 22-year period. I have analyzed the data using species richness, heterogeneity and spatial distribution. For all methods there are theoretical values and features which can be achieved in hypothetical situations and using them I have compared given results. Results show that data has a logarithmic normal distribution of species, that the number of species depends on the number of observation and that human observation of species is not random. Data also shows significant spatial grouping. I have deduced that human observation data best describes the interaction between humans and organisms for areas that humans visit. This kind of data achieves its fullest potential while describing human behavior in environment.

Keywords: citizen science, diversity, behavior, distribution, accumulation, rarefaction
(63 pages, 27 figures, 3 tables, 82 references, original in: croatian)

Thesis is deposited in Central Biological Library.

Mentor: prof. dr. sc. Damjan Franjević

Co-mentor: dr. sc. Josip Skejo

Reviewers:

Prof. Sven Jelaska, PhD

Prof. Ivana Maguire, PhD

Asst. Prof. Zorica Petrinec, PhD

Asst. Prof. Luka Valožić, PhD

Substitute: Prof. Darko Tibljaš, PhD

Thesis accepted: 08. 09. 2022.

SADRŽAJ

1. UVOD	1
1.1. RAZVOJ BIOLOŠKE INFORMATIKE OD 1999. GODINE.....	2
1.1.1. Globalna podatkovna infrastruktura bioraznolikosti (GBIF).....	4
1.1.2. Darwin jezgra.....	5
1.1.2.1. Vokabular Darwin jezgre	5
1.2. GRAĐANSKA ZNANOST I LJUDSKO OPAŽANJE RAZNOLIKOSTI VRSTA.....	9
1.2.1. Pristranosti građanske znanosti u kontekstu opažanja raznolikosti vrsta	11
1.2.1.1. Taksonomska pristranost.....	11
1.2.1.2. Prostorna pristranost	13
1.2.1.3. Vremenska pristranost.....	14
1.2.2. Ponašanje građana znanstvenika	14
1.3. MJERE RAZNOLIKOSTI VRSTA	14
1.3.1. Bogatstvo vrsta.....	14
1.3.2. Mjere heterogenosti	15
1.3.3. Uvjet za metodu uzorkovanja raznolikosti.....	15
1.3.4. Akumulacijska krivulja vrsta ili krivulja otkrivanja vrsta	16
1.4. PROSTORNA RASPODJELA PODATAKA (PROSTORNI UZORAK TOČAKA)	21
1.4.1. Ripleyjeva K-funkcija	21
2. CILJEVI.....	23
3. MATERIJALI I METODE	24
3.1. MATERIJALI	24
3.1.1. Preuzimanje podataka	24
3.1.2. Struktura metapodataka i odabir atributa	26
3.1.3. Struktura podataka	26
3.2. METODE	27
3.2.1. Definiranje uzorka i njegove veličine, te njihova podjela	27
3.2.2. Promatrani parametri	27
3.2.3. Računalne aplikacije	27
3.2.4. Problemi nalaza.....	28
3.2.5. Mjere raznolikosti	28
3.2.5.1. Rang-abundancijski modeli.....	28

3.2.5.2. Logaritamski normalna raspodjela	29
3.2.5.3. Akumulacijska krivulja	29
3.2.5.4. Rarefakcija	30
3.2.6. Ripleyjeva K-funkcija	31
4. REZULTATI	33
4.1. LOGARITAMSKE SERIJE	33
4.1.1. Rang-abundancijski ili dominacija / raznolikost modeli.....	33
4.1.2. Prestonova logaritamski normalna raspodjela	38
4.2. AKUMULACIJSKE KRIVULJE	43
4.3. RAREFAKCIJA	46
4.4. RIPLEYJEVA K-FUNKCIJA	49
5. RASPRAVA	52
6. ZAKLJUČAK.....	56
7. LITERATURA	57
9. PRILOG	63
8. ŽIVOTOPIS	

KRATICE:

.csv – Datoteka s vrijednostima odvojenim zarezom (engl. *comma-separated values*)

AIC – Akaikeov informacijski kriterij

BAP – Podatci bez atributa „problem“

BIC – Bayesov informacijski kriterij

Dev. – Devijacija

DwC – Darwin jezgra (engl. *Darwin core*)

GBIF – Globalna podatkovna infrastruktura bioraznolikosti (engl. *Global biodiversity Information Facility*)

LNS – Logaritamski normalna serija (raspodjela)

nK – Broj korisnika

nO – Broj opažanja

nP – Broj porodica

nR – Broj rodova

nRA – broj razreda

nRE – Broj redova

nV – Broj vrsta

OECD – Organizacija za gospodarsku suradnju i razvoj (engl. *Organisation for Economic Co-operation and Development*)

PLNR – Prestonova logaritamski normalna raspodjela

RK-funkcija – Ripleyjeva K-funkcija

SD – Standardna devijacija

1. UVOD

Globalna podatkovna infrastruktura bioraznolikosti (engl. *Global Biodiversity Information Facility*; dalje u tekstu: GBIF) je međunarodna javno dostupna baza podataka o raznolikosti vrsta koja objedinjuje:

1. zapise nastale ljudskim opažanjem (engl. *human observation*), tj. građanska opažanja svojiti koja najčešće uključuju fotografiju i detaljne podatke (npr. datum i lokacija)
2. neobjavljene podatke iz muzejskih zbirki
3. objavljene znanstvene podatke iz radova (Telenius 2011, What is GBIF? 2021).

Podatci iz ove i sličnih baza podataka sve se češće koriste i kao podatci u znanstvenim istraživanjima (Luo i sur. 2021, Waller i sur. 2021). Do sada nije istražena pristranost ovakvih baza podataka i mogućnosti koje one pružaju kao izvor podataka za cjelokupnu raznolikost vrsta, već je to napravljeno samo za određene taksonomske skupine (Strien i sur. 2013, Beck i sur. 2014, Geldmann i sur. 2016). Iako se podatci iz baza poput GBIF-a koriste u istraživanjima, do sad postoji malo sistematičnih istraživanja o tome o kakvim podacima se zapravo radi (Beck i sur. 2014, Rocha-Ortega i sur. 2021). Od ukupno 1,9 milijardi zapisa u GBIF-u, 84% pripada ljudskim opažanjima (GBIF 2021). Iako je upitna valjanost ovakvih podataka, isto se može reći za muzejske podatke koji su također rezultat ljudskih opažanja (Specht i Lewandowski 2018, Nekola i sur. 2019). Valjanost podataka prikupljenih ljudskim opažanjem za ekološka istraživanja je upitna, tj. neprovjerljiva, jer se radi o podacima koji su drugačiji od podataka koji se koriste za takva istraživanja (Specht i Lewandowski 2018). Posljednjih godina fokus se promijenio ka proučavanju ponašanja samih građana znanstvenika, iako i dalje u svrhu ispravljanja podataka građana znanstvenika kako bi odgovarali analizama u ekološkim istraživanjima (Boakes i sur. 2016, August i sur. 2020, Di Cecco i sur. 2021). Dvije struje mišljenja su proizašle iz tog pristupa:

1. grupiranje korisnika po ponašanju
2. promatranje korisnika po intenzitetu i frekvenciji korištenja platformi za bilježenje vrsta (Boakes i sur. 2016, August i sur. 2020, Di Cecco i sur. 2021).

Tek radom Di Cecco i sur. (2021) ponašanje građana znanstvenika analizirano je za kompletnu raznolikost vrsta, a ne samo pojedinih taksonomskih skupina. Navedeni pristupi analiziranja ponašanja građana znanstvenika daju uvid u njihovu selektivnost (koje vrste ili skupine opažaju), ali ne u nasumičnost koja je uvjet za podatke koji se koriste pri mjerama raznolikosti (Preston 1948). Stoga u ovom diplomskom radu podatke prikupljenim ljudskim opažanjem, tj. kolektivom ljudi, promatram kao metodu (uređaj) za uzorkovanje organizama. O svojstvima

zaključujem promatrajući rezultate mjera raznolikosti i usporedbom s istim, ali nasumičnim podacima, a pomoću tih svojstava i dosadašnjih spoznaja zaključujem o svrsi tih podataka.

1.1. Razvoj biološke informatike od 1999. godine

Kroz povijest prirodnih znanosti, glavni način komunikacije i pohrane podataka dobivenih znanstvenom metodom smatra se literaturni rad (Bowker 2000). On sadrži hipotezu i podatke dobivene u svrhu njenog prihvaćanja ili odbacivanja, te kao takav, predstavlja konačni cilj znanstvenog rada (Bowker 2000). Takva pohrana podataka o bioraznolikosti ima nedostatke međusobne kompatibilnosti, oni nisu dobiveni istim metodama, nisu javno i lako dostupni i u takvom obliku ne doprinose rješavanju problema smanjenja bioraznolikosti na svjetskoj razini (Wieczorek i sur. 2012). Zadnjih 40 godina konačni cilj znanstvene metode sve više postaje izrada baza podataka koje su teorijski neutralne, u službi napretka različitih disciplina (Bowker 2000). Tome je pridonijelo stvaranje svjetske mreže koja je omogućila brzu i efikasnu virtualnu povezanost sa zainteresiranim strankama iz cijelog svijeta (Bonney i sur. 2014). Izradu takvih baza podataka uvjetuju prioriteta koji se tiču ljudske vrste kao što je slučaj s medicinskim znanostima i izradom mnoštva javnih baza podataka koje su ujedinjene pod mrežnom stranicom Nacionalni centar za biotehnoške informacije (engl. *National Center for Biotechnology Information*) (NCBI 2021). Shvaćanje važnosti gubitka bioraznolikosti kao prioriteta za dobrobit ljudske vrste dovelo je do stvaranja baza podataka, koje su zbog nedostatka zajedničkog komunikacijskog protokola ostale međusobno izolirane. Spomenuti problem svladan je izradom i upotrebom Darwin jezgra (engl. *Darwin Core*; dalje u radu: DwC) standarda (slika 1.) (Wieczorek i sur. 2012).

gbifID	datasetKey	occurrenceID	kingdom	phylumclass	order	family	genus	species
3384033901	50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/95749135	Animalia	Arthropoda	Insecta	Lepidoptera	Pieridae	Colias
1785)	Colias croceus	PT	Azores	PRESENT	28eb1a3f-1c15-4a95-931a-4af90ecb574d	37.797945	-25.25544	2021-09-08T12:48:00
8	9	2021	5137612	5137612	HUMAN_OBSERVATION	iNaturalist	Observations	95749135
Luca Boscain	Luca Boscain	Luca Boscain	2021-09-22T12:27:53	CC_BY_NC_4_0	Luca Boscain	Luca Boscain	2022-05-27T11:31:07.034Z	StillImage
3337841669	50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/91677315	Plantae	Tracheophyta	Liliopsida	Zingiberales	Zingiberaceae	Hedychium
gardnerianum	Sheppard ex Ker Gawl.	Hedychium gardnerianum	PT	Azores	PRESENT	28eb1a3f-1c15-4a95-931a-4af90ecb574d	37.795522	-25.487727
61.0	2021-07-04T19:16:00	4	7	2021	2758700	2758700	HUMAN_OBSERVATION	iNaturalist
Observations	91677315	Milan Svoboda	Milan Svoboda	Milan Svoboda	2022-05-27T11:01:02.189Z	StillImage	COORDINATE_ROUNDED	

(a)

PMF

ZA HERBARIUM CROATICUM

Inv. No: Herbarium ID: 54717

Taxon: *Hyosecyamus albus* L.

Loc.: otok Palagruža

Hab.:

Leg.: Vedran Šegota, Anja Rimac
Leg. data: 9.5.2019.
Det.: Vedran Šegota, Anja Rimac
Det. data: 9.5.2019.

Nota: FGD

(b)

Slika 1. Dva primjera metapodataka za zapise organizama. (a) metapodatci u DwC standardu, csv. formata, kakav se može preuzeti sa GBIF.org. (b) metapodatci za pripadajući list herbarija, kakav se koristi u herbaru Botaničkog zavoda PMF-a u Zagrebu. Oba primjera metapodataka dijele osnovne informacije (atribute): *gbifID* = *Inv. No.*; *occurrenceID* = *Herbarium ID*, *VerbatimScientificNameAuthorship* = *Taxon*, *localitystateProvince* = *Loc.*, *recordedBy* = *Leg.*, *day*; *month*; *year* = *Leg. Data*, *dateIdentified* = *Det. Data*. No DwC standard ima još dodatnih atributa. Preuzeto i prilagođeno s: (a) www.gbif.org/occurrence, (b) www.hirc.botanic.hr/fcd/beta/herbar (pristupljeno 14. 08. 2022.)

1.1.1. Globalna podatkovna infrastruktura bioraznolikosti (GBIF)

Organizacija za gospodarsku suradnju i razvoj (engl. *Organisation for Economic Co-operation and Development*; dalje u radu: OECD) kojoj je cilj osmišljavanje međunarodnih standarda osnovanih na dokazanim metodama, održavanje foruma, sakupljanje i analiziranje podataka u svrhu stvaranja dobrih praksi koje donositelji političkih odluka mogu koristiti (About the OECD - OECD 2021), organizirala je 1999. godine OECD mega znanstveni forum (engl. *OECD Megascience Forum* (What is GBIF? 2021)). Unutar OECD mega znanstvenog foruma, Radna skupina za biološku informatiku (engl. *Working Group On Biological Informatics*) objavila je konačni izvještaj unutar kojeg predlažu osnivanje i podržavanje međusobno umreženih informatičkih alata i baza podataka koje bi zajedno tvorile GBIF (slika 2.) (OECD Megascience Forum 1999). Cilj ove podatkovne infrastrukture je omogućiti javni globalni pristup podacima o bioraznolikosti i ekosustavima, njegovo nadopunjavanje i napredak znanosti u korist svih zainteresiranih za potrebe širokog raspona opće dobrobiti (WGBI 1999). GBIF trenutno sadrži 1889008570 nalaza iz cijelog svijeta (Search 2022a) i podržava tri podatkovna standarda: DwC – Darwin jezgra (engl. *Darwin core*), EML – Ekološki meta podatkovni jezik (engl. *Ecological Metadata Language*) i ABCD – Pristup podacima bioloških zbirki (engl. *Access to Biological Collections Data*) (Data standards 2021). Od tri navedena standarda, u izradi ovog diplomskog rada koristio sam samo DwC.

Scientific name	Country or area	Coordinates	Month & year	Dataset
<i>Hydrilodes lentalls</i> Guenée, 1854	Chinese Taipei	25.0N, 121.5E	2021 January	Hu
<i>Tanaocenia halaria</i> Walker, 1861	Chinese Taipei	23.5N, 120.8E	2021 January	Hu
<i>Corvus brachyrhynchos</i> C.L.Brehm, 1822	United States of America	36.9N, 76.1W	2021 January	Hu
<i>Anas platyrhynchos</i> Linnaeus, 1758	United States of America	36.9N, 76.1W	2021 January	Hu
<i>Cardinalis cardinalis</i> (Linnaeus, 1758)	United States of America	36.9N, 76.1W	2021 January	Hu
<i>Toxostoma rufum</i> (Linnaeus, 1758)	United States of America	36.9N, 76.1W	2021 January	Hu
<i>Setophaga coronata</i> (Linnaeus, 1766)	United States of America	36.9N, 76.1W	2021 January	Hu
<i>Cyanocitta cristata</i> (Linnaeus, 1758)	United States of America	36.9N, 76.1W	2021 January	Hu

Slika 2. Sučelje mrežne baze podataka gbif.org za pretraživanje zapisa organizma. Na lijevoj strani sučelja nalazi se bočna traka sa filterima. Na desnoj strani prikazani su zapisi organizama s pripadajućim atributima u tabličnom obliku. Slika zaslona preuzeta i prilagođena sa www.gbif.org/occurrence (Pristupljeno: 19. 05. 2022.).

1.1.2. Darwin jezgra

Započet 1999. godine DwC standard je zamišljen kao skup termina i njihovih definicija kojima bi se podatci o bioraznolikosti mogli jednoznačno opisati i međusobno razmjenjivati. Time su omogućili njegov daljnji razvoj i nadopunjavanje bez gubitka u kompatibilnosti između njegovih različitih verzija. Svoje korijene vuče iz Dublin Core metapodatkovne inicijative (engl. *Dublin Core Metadata Initiative*) (Wieczorek i sur. 2012), organizacije čiji je cilj inovacija i podrška u korištenju metapodataka (DCMI: About DCMI 2021). Izradila ga je Radna skupina Darwin jezgre (engl. *Darwin Core Task Group*) (Wieczorek i sur. 2012, Darwin Core - TDWG 2021), podskupina Međunarodne radne skupine za taksonomske baze podataka (engl. *The International Working Group on Taxonomic Databases*) koja je znanstvena, neprofitabilna organizacija sa ciljevima osmišljavanja, razvijanja i promoviranja uputa i standarda korištenih u biološkoj informatici (Constitution - TDWG 2021). Termini su podijeljeni u devet kategorija uz pripadajuće podkategorije termina (Wieczorek i sur. 2012). Sedam kategorija se odnose na Jednostavni DwC (engl. *Simple DwC*) standard i one su: zapis nalaza, pojavnost, događaj, lokacija, identifikacija, svojta i geološki kontekst (za paleontološke nalaze), na način da svaki zapis nalaza ima jedan red, a svaki termin kategorije ima po jedan pripadajući stupac (Wieczorek i sur. 2012). Ostale dvije od devet kategorija, odnose se na Srodni DwC (engl. *Generic DwC*) standard koji se ne može jednostavno prikazati kao prethodni standard jer se kategorije odnose na mjerenja (koja zahtijevaju više redaka za isti nalaz) ili međusobne odnose (Wieczorek i sur. 2012). Popis i opis svih termina nalazi se na mrežnoj stranici (Core terms defined by Darwin Core 2021). Takav način korištenja i nadogradnje metapodataka omogućio je daljnji razvoj baza o bioraznolikosti koje nisu izolirane jedne od drugih te je olakšao njihovo korištenje (Wieczorek i sur. 2012). Takav primjer je i javno dostupna mrežna baza GBIF, koja pomoću tih standarda unosi podatke s drugih mrežnih stranica (slika 3.), baza i platformi kao što su: EOD (eBird), Artportalen, iNaturalist, Observation.org (slika 4.).

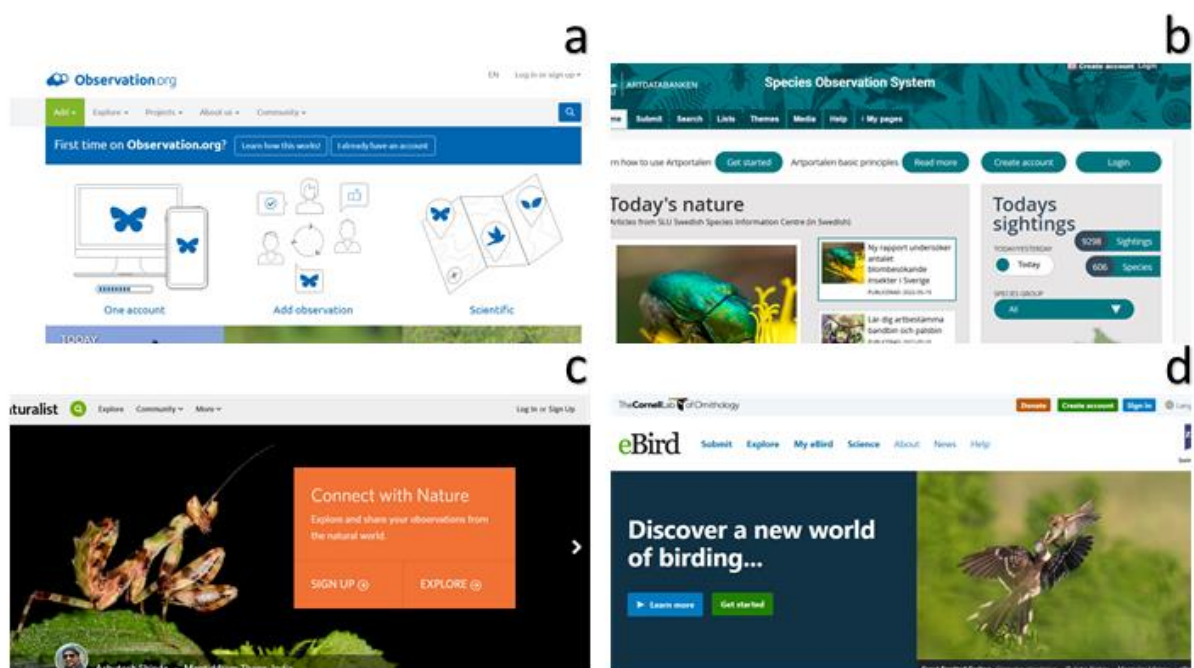
1.1.2.1. Vokabular Darwin jezgre

Pojmovi DwC vokabulara koji su važni za razumijevanje ovog diplomskog rada su:

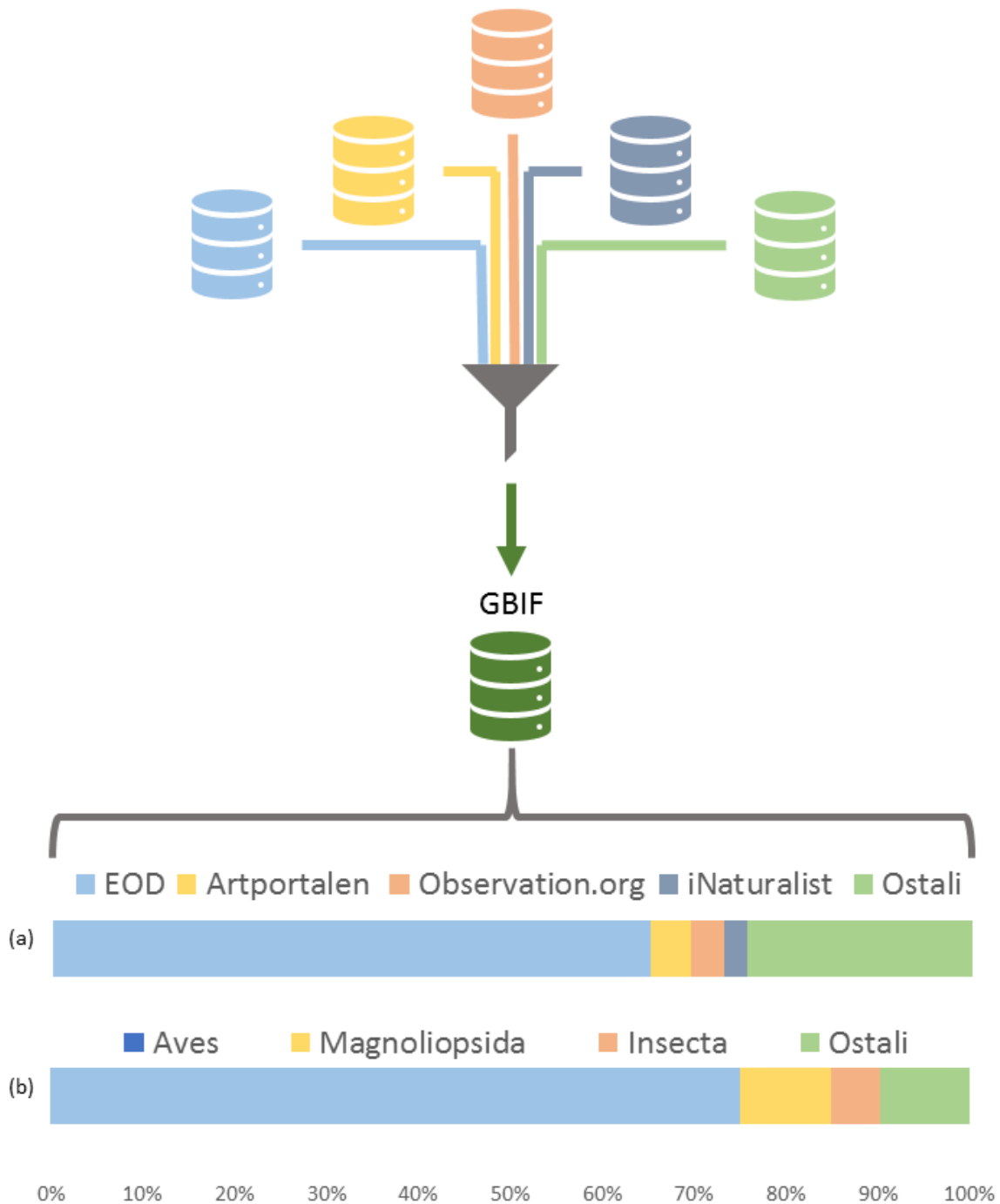
1. Ljudsko opažanje (skraćeno: opažanje; engl. *human observation*; DwC vokabular: *humanObservation*) je kategorija specifične prirode zapisa podatka (engl. *basis of record*; DwC vokabular: *basisOfRecord*) nastala kao rezultat ljudskog opažanja (slika 5.) (Core terms defined by Darwin Core 2021).

2. Nalaz (engl. *occurrence*; DwC vokabular: Occurrence) je postojanje organizma u određenom prostoru i vremenu (Core terms defined by Darwin Core 2021)
3. Korisnik (engl. *recorded by*; DwC vokabular: *recordedBy*; doslovni prijevod: zapisao) je osoba, grupa osoba ili institucija koja je odgovorna za nastanak zapisa (Core terms defined by Darwin Core 2021)

Dodatni pojam koji nije zaseban pojam u DwC vokabularu, ali potreban za razumijevanje, je *zapis*. Zapis su uneseni (zabilježeni) podatci o nalazu, u širem smislu zadržavanje informacije, a u informatičkom smislu pohranjivanje informacija na računalo (Record 2022) U kontekstu DwC vokabulara on se odnosi na zapis nalaza.



Slika 3. Početne stranice prvih četiriju mrežnih baza po broju zapisa preuzetih na gbif.org. (a) observation.org; (b) Artportalen; (c) iNaturalist; (d) EOD (eBird). Slike zaslona preuzete i prilagođene s: www.observation.org; <https://www.artportalen.se>; www.inaturalist.org; <https://ebird.org/home> (Pristupljeno: 19. 5. 2022.).



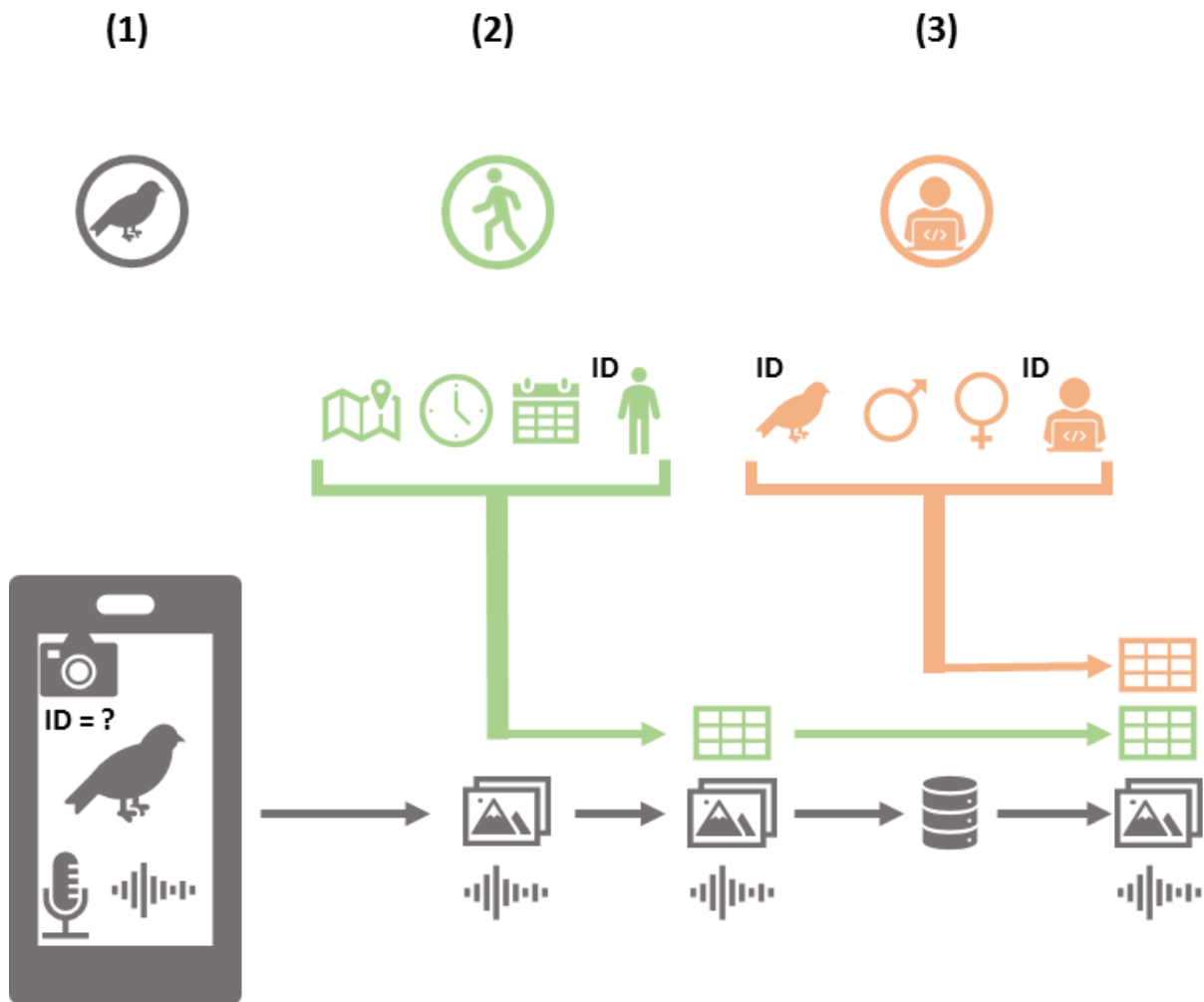
Slika 4. Proces unosa podataka mrežnih baza u GBIF i omjeri unesenih podataka prikupljenih ljudskim opažanjem iz ostalih baza i taksonomskih razreda. Baze filtriraju i šalju podatke zadovoljavajuće točnosti bazi GBIF. (a) najveći broj podataka (65%) pripada bazi EOD (eBird), zatim slijede Artportalen, Observation.org i iNaturalist (Search 2022b). Te četiri baze tvore 75% podataka baze GBIF, dok 25% podataka pripada ostalim bazama. (b) Razredu Aves pripada 75% zapisa u bazi GBIF, zatim slijede Magnoliopsida i Insecta. Tim trima razredima pripada 90% zapisa baze GBIF, dok svega 10% posto pripada ostalim razredima (Search 2022b).



Slika 5. Ljudsko opažanje je rezultat procesa opažanja ljudskim osjetilima. Čovjek vizualno opaža atraktivne organizme, npr. s upozoravajućim obojenjem; njuhom opazi organizme koji ispuštaju miris, npr. biljke koje cvatu; a sluhom može opaziti organizme koji se glasaju, npr. ptice, sisavce, cvrčke i zrikavce.

1.2. Građanska znanost i ljudsko opažanje raznolikosti vrsta

Građanska znanost oblik je masovne podrške za prikupljanje znanstveno valjanih podataka od strane građana (Vohland i sur. 2021). To je svjetska pojava u kojoj milijuni građana sudjeluju u sakupljanju i obradi mnoštva podataka u sklopu projekata, koji pokrivaju različita područja znanosti od mikrobioma do potresa, pa sve do galaksija (Bonney i sur. 2014, Bossu i sur. 2018, Quitariano i Wald 2020). Do ove pojave je došlo zbog svjetske mreže koja je omogućila građanima povezanost s projektima od osobnog interesa, a samim time edukaciju i izravno unošenje podataka u mrežne baze podataka (Bonney i sur. 2014). Građani znanstvenici nisu kvalificirani znanstvenici, ali su zbog svog mnoštva sposobni sakupiti višestruku količinu podataka od pojedinačnog znanstvenog tima (Bonney i sur. 2014). Podatci sakupljeni pomoću građanske znanosti, iako brojni, nisu nužno prihvaćeni kao valjani izvor podataka u znanstvenim krugovima i često su neiskorišteni u napretku znanosti (Bonney i sur. 2014). U kontekstu raznolikosti vrsta, to je prikupljanje nalaza putem opažanja, a prikupljanje može biti organizirano ili samostalno (Di Cecco i sur. 2021). Karakteristika takvog prikupljanja podataka o raznolikosti vrsta je oportunitizam; nalazi nisu nužno prikupljeni standardiziranim protokolima, ako i jesu ti protokoli nisu univerzalni za svaki projekt, a nalazi su zabilježeni na mjestima koja su pristupačna, te je pristranost prema određenim skupinama izražena (Strien i sur. 2013, Geldmann i sur. 2016, Di Cecco i sur. 2021). Ti nalazi mogu dalje ovisiti o drugim sudionicima masovne podrške koji identificiraju nalaze do moguće taksonomske razine (Kosmala i sur. 2016, Di Cecco i sur. 2021). Time konačni skup podataka ovisi o osobi koja je opazila, osobi ili osobama koje identificiraju opažanja i mrežnim bazama koje prosljeđuju prikupljene podatke GBIF-u (slika 6.).



Slika 6. Primjer procesa stvaranja zapisa ljudskog opažanja organizma. ID = identitet. (1) korisnik opaža nepoznati ili poznati organizam, te ukoliko može/želi uzima dokaz o opažanju u obliku fotografije ili zvučnog zapisa. (2) korisnik pridružuje zapisu metapodatke o lokaciji, vremenu, datumu i svom identitetu, te zapis sa metapodacima šalje bazi podataka koju koristi. (3) taksonom u bazi podataka pridružuje zapisu metapodatke o identitetu organizma (do njemu poznate taksonomske razine), spolu, starosti, itd. (kao što je moguće u bazi iNaturalist).

1.2.1. Pristranosti građanske znanosti u kontekstu opažanja raznolikosti vrsta

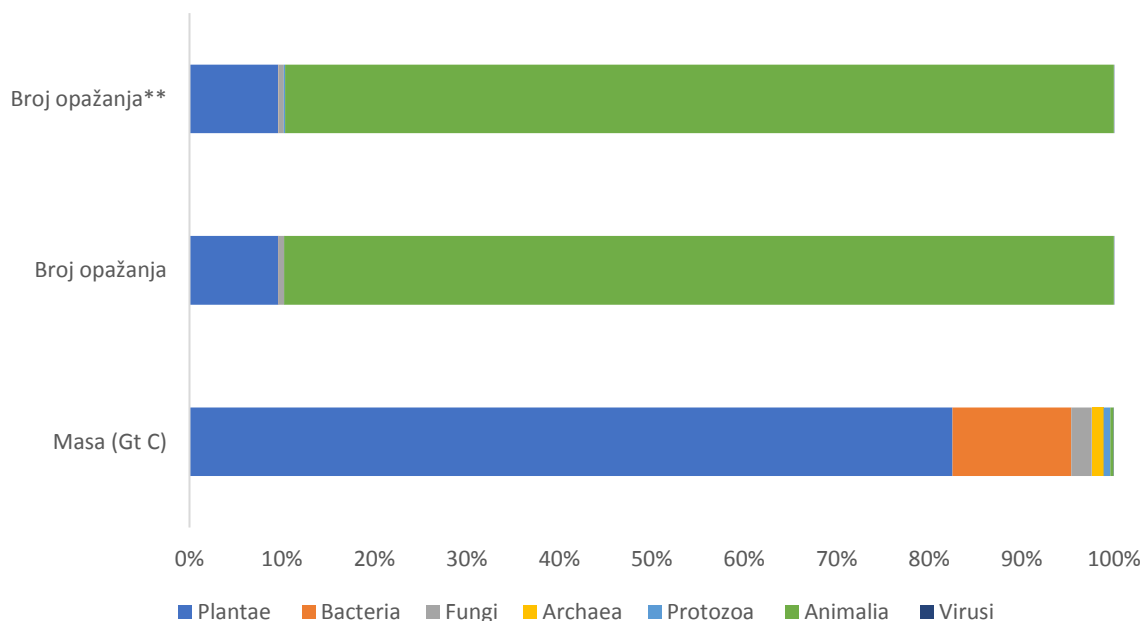
Postojeći skepticizam prema podacima građanske znanosti u ekološkim istraživanjima usmjeren je prema točnosti i preciznosti podataka u usporedbi s podacima sakupljenim od strane znanstvenika (Aceves-Bueno i sur. 2017). No taj skepticizam pokazao se neutemeljenim jer su usporedbe podataka građana znanstvenika i znanstvenika rađene bez referentnog seta podataka gdje bi se mogla usporediti točnost i preciznost podataka od znanstvenika (Specht i Lewandowski 2018). Rad Specht i Lewandowski (2018) smatra kako se podatci znanstvenika ne mogu smatrati točnijim od građana znanstvenika bez takve usporedbe, već drugačijima. U podacima o raznolikosti vrsta postoje pristranosti (Beck i sur. 2014, Bird i sur. 2014a, Geldmann i sur. 2016, Fanelli i sur. 2017, Rocha-Ortega i sur. 2021). One se dijele na tri kategorije:

1. taksonomska
2. prostorna
3. vremenska (Bird i sur. 2014b, Geldmann i sur. 2016, Di Cecco i sur. 2021).

1.2.1.1. Taksonomska pristranost

Taksonomska pristranost ili taksonomski šovinizam je pristranost prema karizmatičnim i bolje istraženim vrstama ili skupinama (Bonnet i sur. 2002, Titley i sur. 2017). Ona je prisutna u ekološkim i biološko konzervacijskim istraživanjima koji koriste podatke znanstvenika, odnosno nije isključivo prisutna u istraživanjima koja koriste podatke građana znanstvenika (Bonnet i sur. 2002, Clark 2002, Donaldson i sur. 2017, Troudet i sur. 2017, Di Cecco i sur. 2021). Prema podacima GBIF-a, u periodu od 1999. do 2021. godine za cijeli svijet, najviše ljudskih opažanja (brojnost) pripada carstvu Animalia (89,76%). nakon njega carstvo Plantae (9,60%) (slika 7.) (Search 2022b.). Na primjeru ta dva carstva najbolje se uočava taksonomska pristranost; carstvo Plantae, su autotrofni organizmi koji su primarni izvor ugljika u hranidbenom lancu i svojom biomasom (gigatonama ugljika; Gt C) u svijetu (82,54%) nadmašuju biomasu carstva Animalia (0,73%) (slika 7.) (Bar-On i sur. 2018). Iako su ta dva različita mjerila abundancije neusporediva, nesrazmjernost vrijednosti je indikativan za postojanje taksonomske pristranosti.

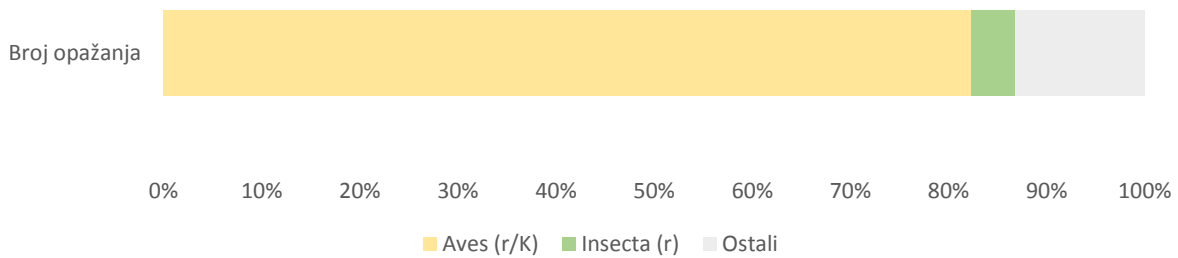
Omjeri biomase i ljudskih opažanja carstava u svijetu



Slika 7. Omjeri biomase carstava (Bar-On i sur. 2018) i omjeri broja ljudskih opažanja u GBIF-u (Search 2022b.). U podacima GBIF-a postoji carstvo Chromista, koje se ne nalazi u radu (Bar-On i sur. 2018), te su njihova opažanja pridružena carstvu Protozoa u stupcu Broj opažanja**. Tradicionalno su neke skupine carstva Chromista pripadale u carstvo Protozoa (Cavalier-Smith 1981).

Također, možemo zaključiti taksonomsku pristranost pomoću teorije r/K selekcije (MacArthur and Wilson 1967). Prema toj teoriji selekcija vrsta u ekosustavu ovisi o osobinama vezanim za kvalitetu brige za potomke i kvantiteti potomaka; te osobine određuju dvije strategije razmnožavanja: r -strategiju i K -strategiju (MacArthur i Wilson 1967). Što je veća kvaliteta brige za potomke, a manja kvantiteta potomaka to vrsta ima više osobina K -strategije; što je manja kvaliteta brige za potomke, a veća kvantiteta potomaka to vrsta ima više osobina r -strategije (MacArthur i Wilson 1967). Ako promatramo razred Insecta on generalno ima r -strategiju (Pianka 1970), dok razred Aves ima r/K -strategiju (Sæther i sur. 2016); time bi udio razreda Insecta bio veći od razreda Aves. Iz udjela opažanja u GBIF-u tih taksonomskih skupina vidljivo je kako ima više opažanja razreda Aves, nego razred Insecta (slika 8.) (Search 2022b).

Udio razreda po r/K strategijama



Slika 8. Udio taksonomskih razreda po pripadanju r/K strategijama selekcije u podacima ljudskog opažanja GBIF-a (Search 2022b)

Još jedan primjer koji ukazuje na taksonomsku pristranost ovih podataka je i usporedba broja opažanja reda Passeriformes s brojem opažanja reda Araneae. Red Araneae u prirodi postiže prosječnu gustoću 131 do 151 jedinki po jednom metru kvadratnom, a u idealnim uvjetima i do 1000 jedinki po metru kvadratnom, dok red Passeriformes 600 jedinki po kilometru kvadratnom (Emlen 1977, Nyffeler i Birkhofer 2017). Prema tome u nasumičnom uzorku očekivalo bi se više jedinki reda Araneae nego reda Passeriformes, no prema podacima ljudskog opažanja to nije slučaj; opažanja reda Passeriformes iznose 45 % svih ukupnih opažanja dok opažanja reda Araneae iznose 0,1 % (Search 2022b).

1.2.1.2. Prostorna pristranost

Generalno u znanstvenim istraživanjima prostorna pristranost nastaje prilikom uzorkovanja u prirodi, gdje zbog nepristupačnih područja znanstvenici češće biraju ona područja koja su generalno ljudima lako dostupna (kao što su putovi, ceste ili okolice naselja) umjesto područja koja su nasumično odabrana (Boria i sur. 2014). Rezultat prostorne pristranosti su podatci koji odaju dojam prostorne grupiranosti i dolazi do autokorelacije (Boria i sur. 2014). Specifično na prostornu pristranost podataka koje su prikupili građani znanstvenici utjecaj imaju infrastruktura (pogotovo izraženo kod needuciranih građana znanstvenika) i gustoća populacije (Geldmann i sur. 2016). Osim geografske prostorne pristranosti, građani znanstvenici su također prostorno pristrani prema staništima (Di Cecco i sur. 2021). Razvijena (urbana) staništa i šume (gdje ljudi provode svoje slobodno vrijeme) su zastupljenija u opažanjima nego što je očekivano, dok su grmolika, travnjačka, livadna i poljoprivredna staništa manje zastupljena nego što se može očekivati (Di Cecco i sur. 2021).

1.2.1.3. Vremenska pristranost

Kako biti građanin znanstvenik nije plaćeno zanimanje, sama građanska znanost ovisi o slobodnom vremenu pojedinca (Di Cecco i sur. 2021). Zbog toga više opažanja nastaje vikendima nego tijekom radnog dijela tjedna (Di Cecco i sur. 2021). Osim toga, vidljivo je i odstupanje opažanja od ostatka tjedana u godini (200 opažanja) između 15. i 45. tjedna u godini s maksimalnom vrijednosti (600 opažanja) oko 17. tjedna (Di Cecco i sur. 2021).

1.2.2. Ponašanje građana znanstvenika

Ponašanje građana znanstvenika do sada se pokušalo opisati na dva generalna načina:

1. grupiranjem u tipove ili grupe korisnika (Boakes i sur. 2016)
2. svrstavanjem u kontinuum (August i sur. 2020).

U radu Boakes i sur. (2016) korisnici su podijeljeni na:

1. povremene (engl. *dabbler*); najmanji broj opažanja, najmanja aktivnost s najvećom varijabilnošću
2. redovite (engl. *steady*); karakteristike između diletanata i entuzijasta
3. entuzijaste (engl. *enthusiast*) najveći broj opažanja, najveća aktivnost s najmanjom varijabilnošću.

Suprotno tome u radu August i sur. (2020) korisnici su podijeljeni na kontinuumu četiriju osi:

1. Intenzitet zapisivanja (engl. *recording intensity*)
2. Prostorni opseg (engl. *spatial extent*)
3. Potencijal zapisivanja (engl. *recording potential*)
4. Zapisivanje rijetkosti (engl. *rarity recording*)

Potrebno je napomenuti kako su u radu August i sur. (2020) korisnici s jednim opažanjem isključeni iz analiza kako bi podjela korisnika na kontinuumu imala smisla.

1.3. Mjere raznolikosti vrsta

1.3.1. Bogatstvo vrsta

Bogatstvo vrsta je u osnovi izraz za broj vrsta u zajednici (McIntosh 1967). Rarefakcijom standardiziramo broj vrsta po veličini uzorka (npr. po broju jedinki); to jest procjenjujemo koliko vrsta možemo očekivati u manjem nasumičnom uzorku jedinki uzetih iz veće kolekcije (Krebs 1989).

1.3.2. Mjere heterogenosti

Mjere heterogenosti istovremeno izražavaju bogatstvo i ujednačenost vrsta u zajednici (Krebs 1989). Jedan od načina kojim se to može prikazati je korištenje logaritamskih krivulja i logaritamski normalne raspodjele, koje koriste dva parametra:

1. broj vrsta u uzorku
2. broj jedinki svake vrste (Krebs 1989).

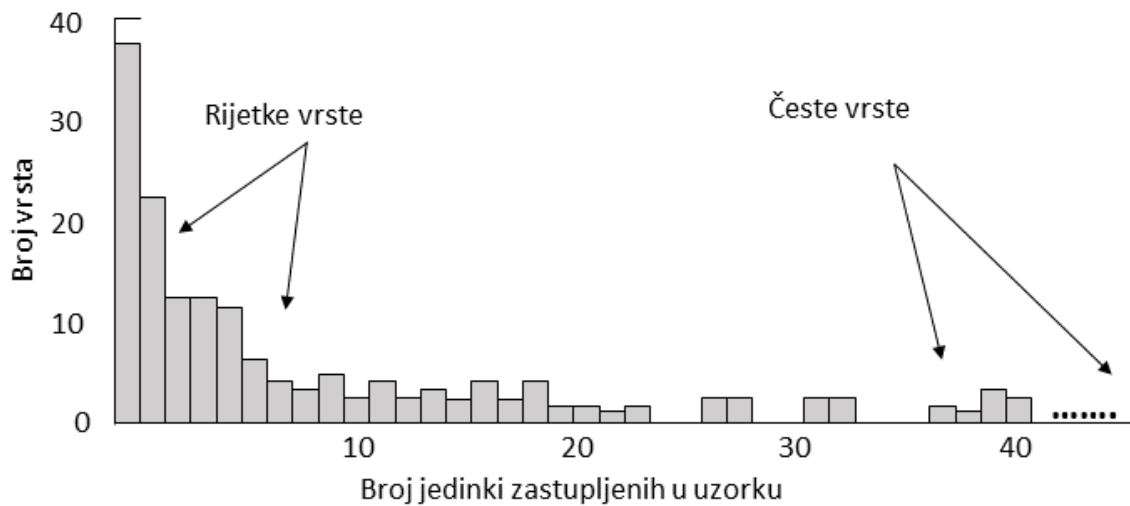
Jedna od karakteristika nekih zajednica je velik broj rijetkih i malen broj učestalih vrsta (Krebs 1989). Takvu raspodjelu najbolje opisuje logaritamska serija (slika 9.) (Fisher i sur. 1943). U suprotnosti tome je logaritamski normalna raspodjela koja pretpostavlja kako su u uzorku zajednica s najmanjim brojem zastupljene vrste s jednom (rijetke vrste) i najvećim brojem jedinki (učestale vrste) (slika 10.) (Preston 1948, Krebs 1989). Logaritamski normalna raspodjela ima nepotpunost koja se naziva skrivenim područjem logaritamski normalne raspodjele koji se teoretski otkriva sve većim brojem uzoraka, koji ne mijenjaju njen karakteristični oblik (Preston 1948, Krebs 1989). Kako bi se pojednostavio prikaz, brojnost jedinki je podijeljena u grupe logaritamskih oktava (Preston 1948). Prva oktava sadrži polovinu broja vrsta s jednom jedinkom, druga oktava sadrži polovinu broja vrsta s dvije jedinke i drugu polovicu vrsta s jednom jedinkom, zatim sa dvije, četiri, osam, šesnaest, itd. (Preston 1948). Takvo prelijevanje brojnosti vrsta, iz oktave u sljedeću, radi se zbog pretpostavke kako ne postoji dovoljno velik uzorak vrsta, te da rijetke vrste (sa samo jednom jedinkom) nisu zaista rijetke (Preston 1948).

1.3.3. Uvjet za metodu uzorkovanja raznolikosti

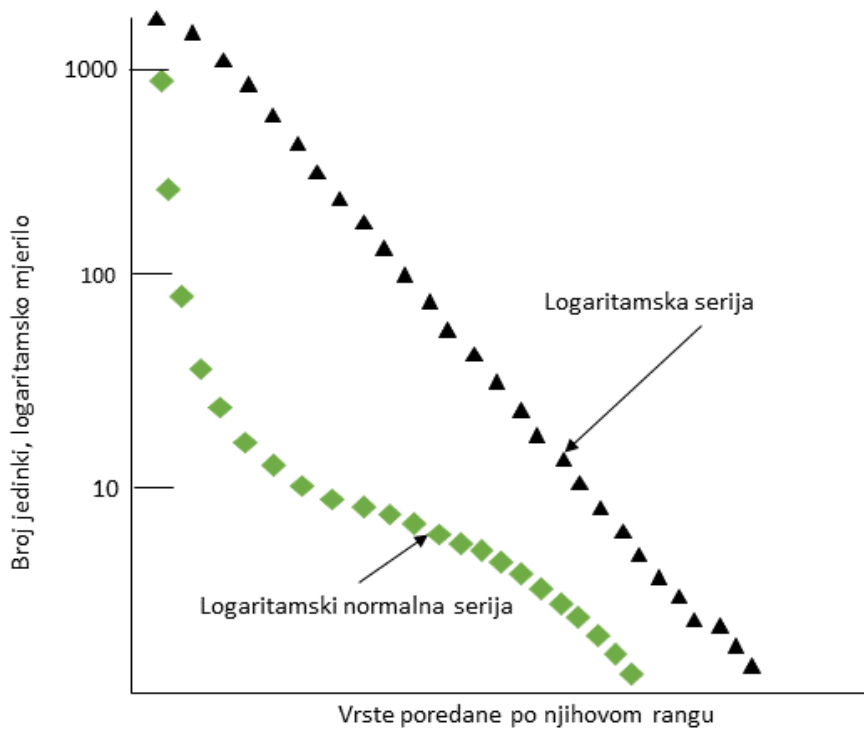
U ovom radu vodim se tvrdnjama iz rada Preston (1948) gdje je nasumičnost navedena kao uvjet koje metode moraju zadovoljiti kako bi prikupljeni podatci odgovarali analizama za mjerenje raznolikosti vrsta. Time brojnost vrsta u uzorku ovisi o brojnosti vrsta na istraživanom području (Preston 1948). Metode uzorkovanja su selektivne prema ciljnim skupinama istraživanja, npr. svjetlosna piramida je selektivna prema fototaksičnim leptirima (slika 11. prilog 1., 2.) (Preston 1948). Analogno tome i ljudsko opažanje je selektivno, ali iz pregleda dostupnih podataka vidljivo je kako ono nije ograničeno samo jednom taksonomskom skupinom, npr. rodnom, porodicom, pa niti jednim carstvom (slika 12.) (Search 2022b).

1.3.4. Akumulacijska krivulja vrsta ili krivulja otkrivanja vrsta

Akumulacijska krivulja ili krivulja otkrivanja vrsta opisuje rast broja novih vrsta prilikom postupka uzimanja novih uzorka ili jedinki, tj. otkrivanja novih vrsta dodatnim uzorkovanjem (Gotelli i Colwell 2001). Akumulacijska krivulja je srodna rarefakcijskoj krivulji zbog toga što rarefakcijska krivulja promatra koliki broj vrsta može biti u manjem uzorku čije su jedinke uzete iz ukupne kolekcije ponavljanjem uzorkovanja, što je statističko očekivanje akumulacijske krivulje (Gotelli i Colwell 2001).

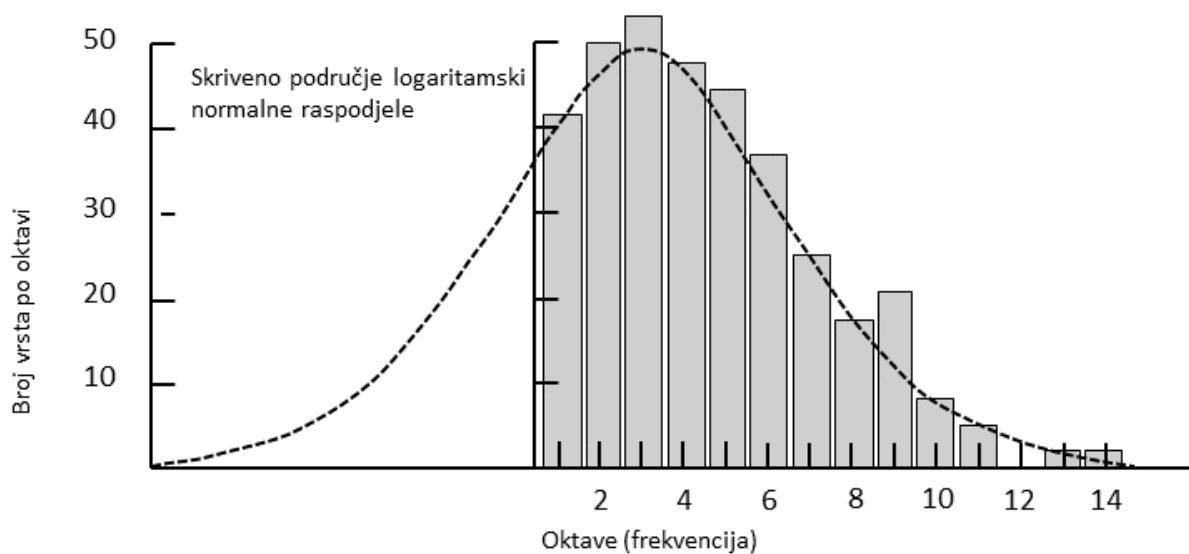


(a)

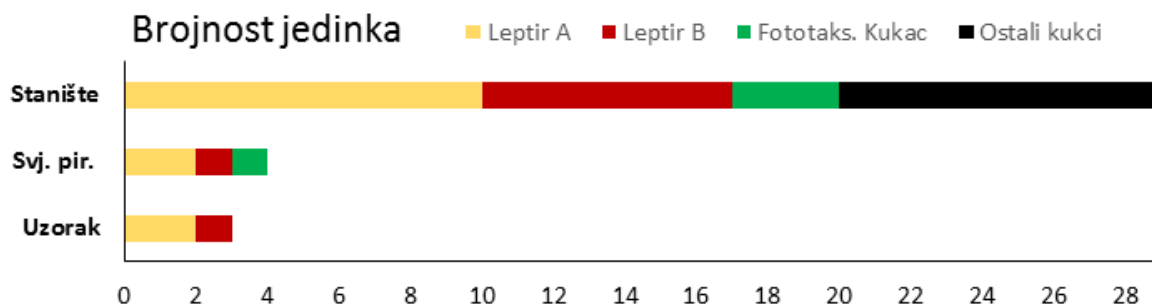
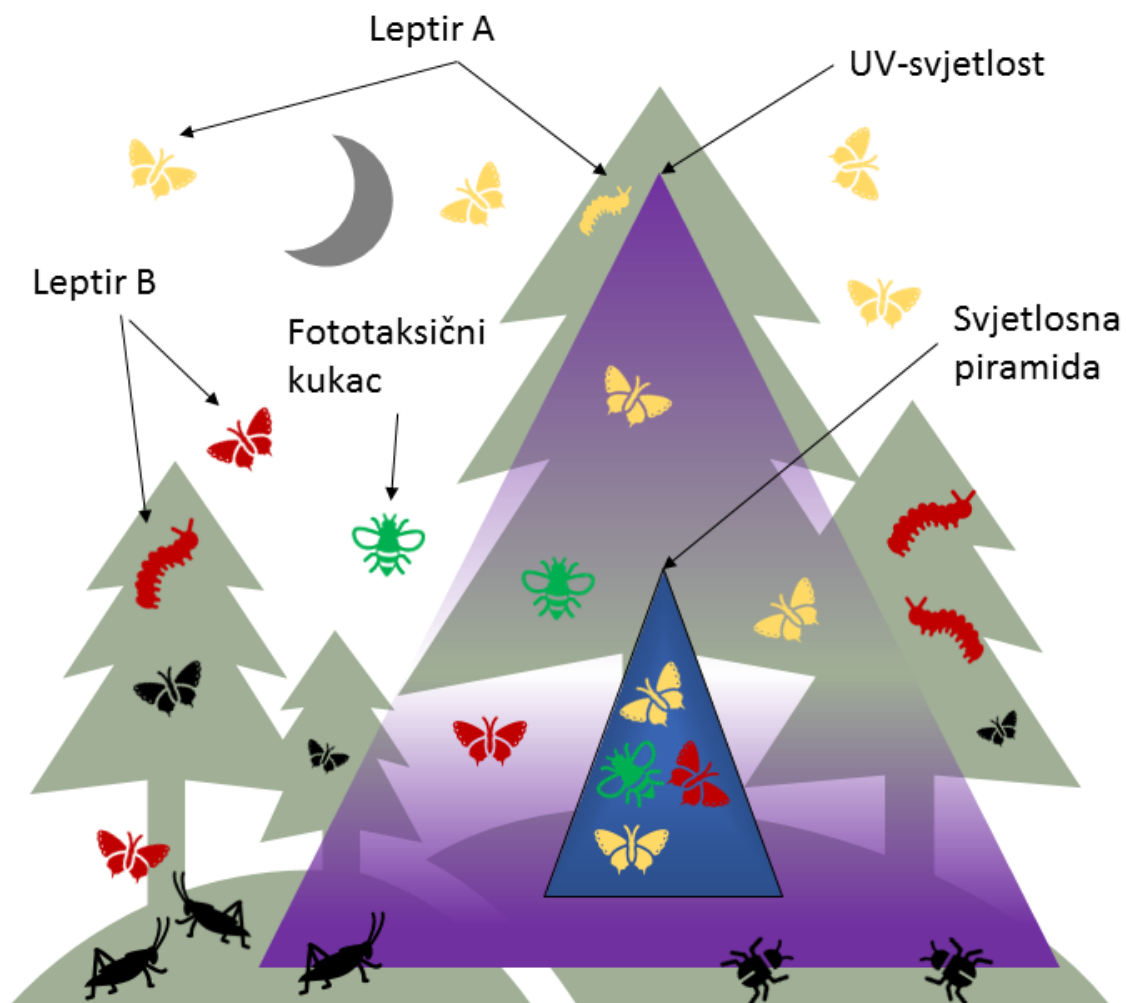


(b)

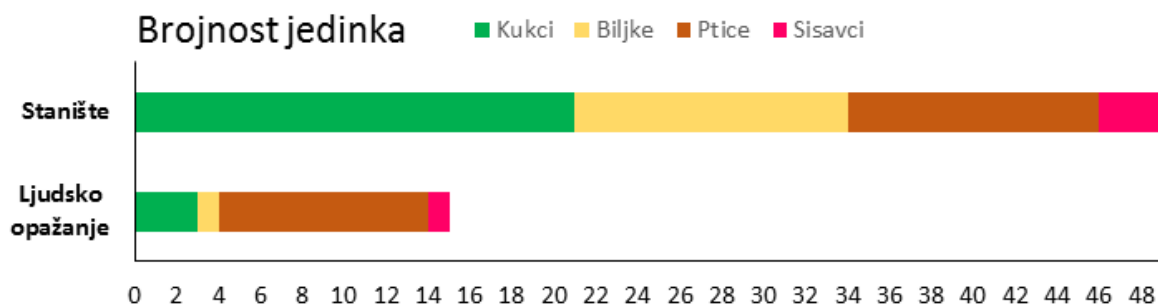
Slika 9. Logaritamska serija i rang-abundancijski prikaz heterogenosti. (a) Hipotetski uzorak vrsta područja prikazan brojem vrsta i brojem jedinki tih vrsta. Najveći broj vrsta je zastupljen najmanjim brojem jedinki, dok je najmanji broj vrsta zastupljen najvećim brojem jedinki (os x je odsječena). Takva je raspodjela vrsta i broja jedinki tih vrsta logaritamska. (b) U rang (broj vrsta)-abundancijskom prikazu logaritamska serija se ponaša kao pravac, dok se logaritamski normalna serija ponaša kao karakteristična krivulja. Preuzeto i prilagođeno iz Krebs (1989).



Slika 10. Prestonova logaritamski normalna raspodjela. Isprekidana linija – teoretska krivulja; sivi pravokutnici – opažena vrijednost broja jedinki svrstana u oktavu. Opažena raspodjela je prekinuta pri oktavi s jednom jedinkom; teoretski to nisu uistinu rijetke vrste već se one nalaze u skrivenom području logaritamski normalne krivulje. Preuzeto i prilagođeno iz Krebs (1989).



Slika 11. Nasumičnost i selektivnost metoda uzorkovanja raznolikosti. Primjer iz Preston (1948). Metode za uzorkovanje raznolikosti ovise o nasumičnosti svoje interakcije s organizmima kako bi postigli reprezentativan uzorak. Svjetlosna piramida je selektivna prema pozitivno fototaksičnim leptirima, a nasumična je jer omjer jedinki vrsta u uzorku ovisi o omjeru jedinki vrsta u staništu. Omjeri leptira A i leptira B u uzorku nisu identični kao u staništu jer svjetlosna piramida nije u interakciji s gusjenicama leptira.



Slika 12. Nasumičnost i selektivnost ljudskog opažanja. Čovjek je taksonomski pristran prema karizmatičnim skupinama kao što su ptice, biljke s cvjetovima ili leptiri. Ostale organizme koji se roje ili koji su sastavni dio izgleda okoliša se bilježi jednim primjerkom. Organizmi koji se nalaze dalje od puta se ni ne opažaju.

1.4. Prostorna raspodjela podataka (prostorni uzorak točaka)

Set podataka koji sadrži opažene prostorne lokacije događaja ili objekata. npr. stabala u šumi, epicentara potresa, opažanja organizama ili položaja galaksija, naziva se prostornim uzorkom točaka (Baddeley i sur. 2016). Takvom setu podataka mogu se pridružiti i kovarijanti tj. bilo koji podatci koji nisu dio prostornog uzorka točaka, nego služe njegovom objašnjavanju npr. visinska karta, karta nagiba, karta staništa, hidrološka karta, karta tla, karta naseljenih područja itd. (Baddeley i sur. 2016). Promatrajući međusobne ovisnosti točaka, u prostornom uzorku točaka možemo odrediti njihovu međusobnu korelaciju, tj. jesu li točke smještene u prostoru zbog međusobne ovisnosti ili su nezavisne jedne od drugih (Baddeley i sur. 2016). Klasična podjela međusobne ovisnosti točaka (slika 13.), koja se također koristi i u ekološkim istraživanjima, je na:

1. nasumičan uzorak točaka – međusobna korelacija ne postoji tj. točke su neovisne
2. uređen uzorak točaka – međusobna korelacija postoji tj. odnos točaka je izbjegavanje
3. grupirani uzorak točaka – međusobna korelacija postoji tj. odnos točaka je grupiranje (Krebs 1989, Haase 1995, Baddeley i sur. 2016).

Jedna od najčešće korištenih funkcija za određivanje korelacije za klasičnu podjelu međusobnih ovisnosti točaka je Ripleyjeva K-funkcija (Baddeley i sur. 2016).

1.4.1. Ripleyjeva K-funkcija

Ripleyjeva K-funkcija (dalje u radu: RK-funkcija) koristi se za analizu drugog reda nepomičnog procesa točaka X koji je definiran kao očekivani broj točaka u mjernoj jedinici prostora $\lambda K(r)$, koji je jednak očekivanom broju dodatnih nasumičnih točaka s udaljenosti r karakteristične nasumične točke u prostoru (Ripley 1977, Baddeley i sur. 2016). Kako bi RK-funkcija bila usporediva, koristi se i funkcija $K(r) = \pi r^2$ koja opisuje Poissonov (nasumični) proces točaka, te odstupanje $\hat{K}(r)$ iznad ili ispod $K(r)$ upućuje na grupiranost ili uređenost točaka (Baddeley i sur. 2016).

Jednadžba 1.

Jednadžba Ripleyjeve K-funkcije preuzeta iz Baddeley i sur. (2016):

$$\hat{K}(r) = \frac{a}{n(n-1)} \sum_i \sum_j I(d_{ij} \leq r) e_{ij}$$

r = Tipična udaljenost nasumične točke

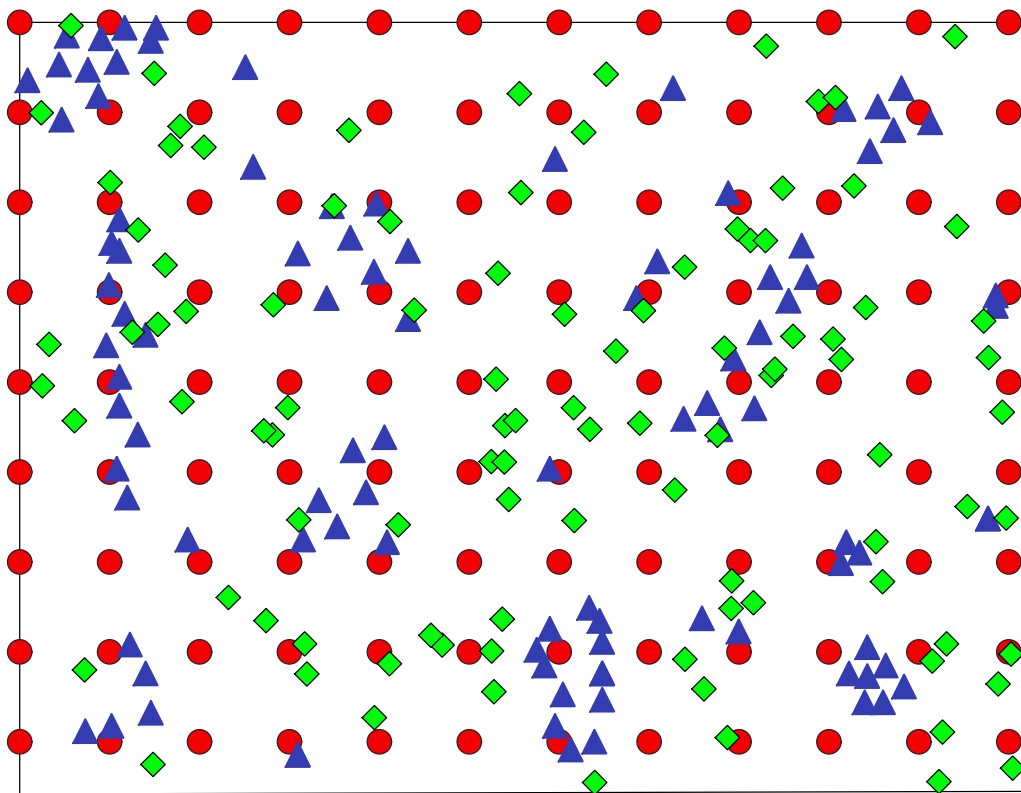
a = Površina okvira

n = Broj točaka

d_{ij} = Udaljenost između dviju točaka

$I(d_{ij} \leq r)$ = Indikator koji je jednak 1 ako je odaljenost manja ili jednaka r

e_{ij} = Težina rubnog ispravka



Slika 13. Prostorna raspodjela 99 nasumičnih, grupiranih i uređenih prostornih uzoraka točaka u dvodimenzionalnom prostoru. Crvene točke – uređene; plave piramide – grupirane; zeleni dijamanti – nasumične; crni okvir – dvodimenzionalni prostor. Točke u dvodimenzionalnom prostoru mogu biti grupirane, nasumične ili uređene.

2.CILJEVI

Osnovni cilj ovog rada je odrediti svojstva podataka raznolikosti vrsta iz baze GBIF prikupljenih ljudskim opažanjem. Njih ću odrediti analizama shodnima onima korištenim u istraživanjima raznolikosti vrsta. Pomoću tih svojstava zaključivat ću o kolektivnom ljudskom opažanju uspoređujući dobivene vrijednosti s nasumičnim i teoretskim vrijednostima i spoznajama. Specifični ciljevi su da se pomoću statističkih analiza za raznolikost vrsta i prostorne analize odredi:

1. heterogenost svojti (koliko je čestih, odnosno koliko je rijetkih vrsta), te provjeriti imaju li podatci s baze GBIF logaritamski normalnu raspodjelu za ljudsko opažanje ili neku drugačiju;
2. bogatstvo svojti:
 - a) akumulacija novih podataka (kako se s povećanjem uzorka povećava broj novih vrsta nekog područja) tj. provjerit ću dolazi li uopće i pri kolikom uzorku do prestanka bilježenja novih svojti;
 - b) rarefakcija podataka (kako se povećanjem uzoraka povećava broj vrsta koji je moguće opaziti na nekom području)
3. prostorna raspodjela opažanja, odnosno provjerit ću jesu li podatci sakupljeni nasumično ili se grupiraju po nekoj osnovi.

3.MATERIJALI I METODE

Postupak preuzimanja i analiziranja podataka prati slijed:

- (1) Filtriranje podataka
- (2) Preuzimanje podataka
- (3) Odabir atributa
- (4) Podjela podataka na uzorke
- (5) Rang-abundancijski ili dominacija / raznolikost modeli
- (6) Prestonova logaritamski normalna raspodjela
- (7) Akumulacija
- (8) Rarefakcijska
- (9) Ripleyjeva K-funkcija

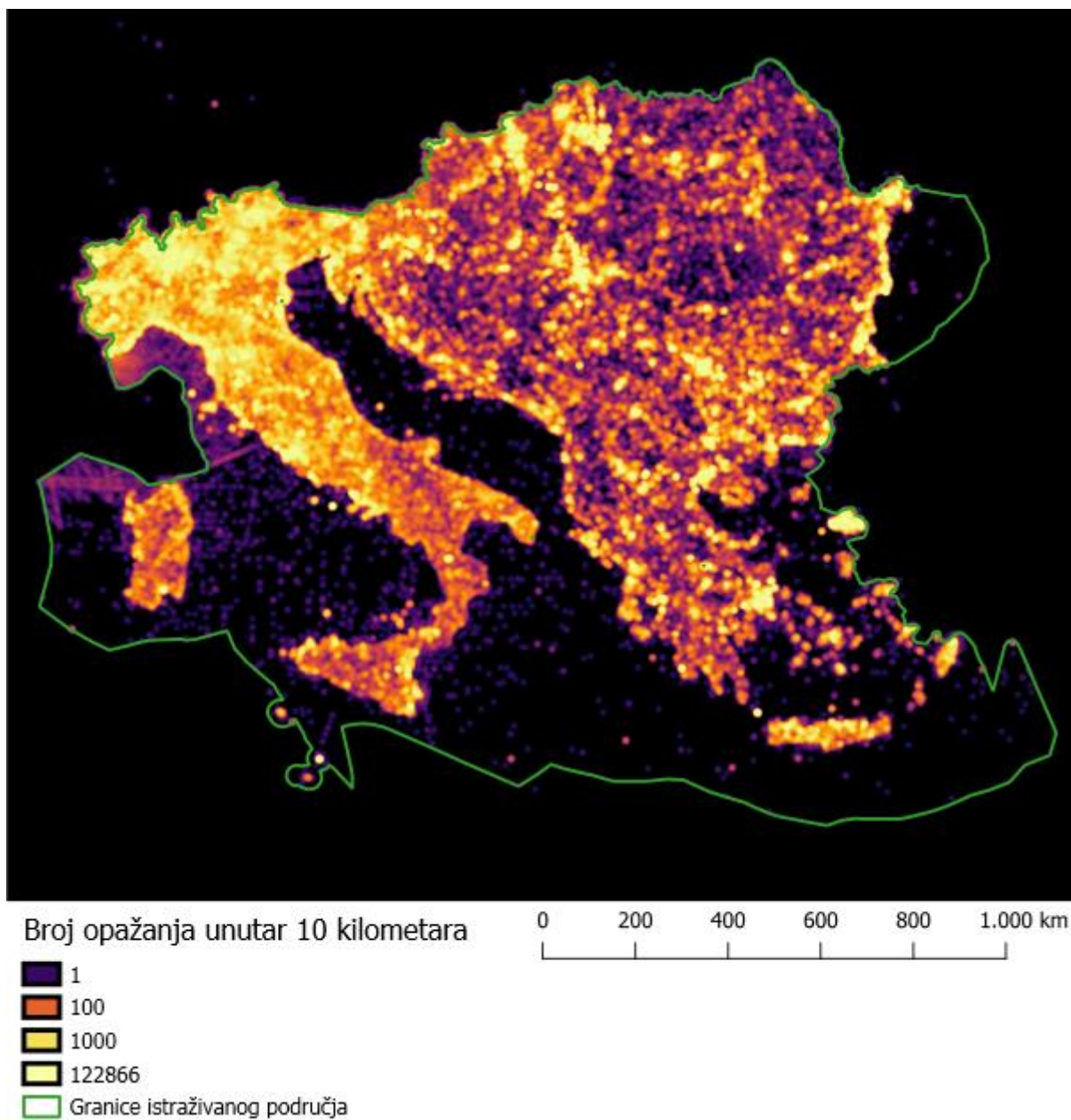
3.1. Materijali

3.1.1. Preuzimanje podataka

Iz baze podataka GBIF preuzeo sam tablicu u .csv formatu, zapisanu u Jednostavnom DwC standardu (Gbif.Org 2022). Podatke sam odabrao pomoću izbornika filtera na mrežnoj bazi GBIF koji su ponuđeni prilikom pretraživanja nalaza. Koristio sam sljedeće filtre i njihova stanja:

- (1) Status nalaza = Prisutan (engl. *Present*)
- (2) Specifična priroda zapisa = Ljudsko opažanje
- (3) Godina = 1999. do 2021.
- (4) Država ili područje = Albanija; Bosna i Hercegovina; Bugarska; Crna Gora; Grčka; Hrvatska; Italija; Kosovo; Mađarska; Makedonija; Rumunjska; Slovenija; Srbija (slika 14.).

Karta istraživanog područja i gustoće ljudskog opažanja



Slika 14. Karta istraživanog područja s gustoćom ljudskog opažanja. Na karti je vidljiva neujednačenost gustoće opažanja, s najvećom gustoćom na sjeverozapadnom dijelu Italije, a najmanjim gustoćama opažanja na moru u smjeru jugoistoka. Također su vidljivi podatci koji se nalaze u preuzetoj tablici, ali se ne nalaze u istraživanom području (vidi poglavlje 3.2.4.). Izrađeno u računalnoj aplikaciji Qgis 3.14.

3.1.2. Struktura metapodataka i odabir atributa

Preuzeta .csv tablica u Jednostavnom DwC standardu sadrži ukupno 53 atributa (stupca), koji zajedno tvore metapodatke za svaki nalaz. Od svih atributa odabrani su: (1) korisnik ; (2) godina (engl. *Year*); (3) problem (engl. *Issue*); (4) vrsta (engl. *Species*); (5) rod (engl. *Genus*); (6) porodica (engl. *Family*); (7) red (engl. *Order*); (8) razred (engl. *Class*).

3.1.3. Struktura podataka

Preuzeti podaci sadrže 6746339 zapisa, od toga je 6714238 zapisa određeno do taksonomske razine vrste, 24502 vrsta (tablica 1.) i 30101 korisnika. Podatci ne predstavljaju događaje prilikom kojih se vršilo uzorkovanje raznolikosti vrsta i samim time ne postoje ni definirani uzorci. Stoga su preuzeti podaci kolekcija koja predstavlja populaciju vrsta i ne može se smatrati manjim nasumičnim uzorkom veće populacije vrsta, osim ako ne postoji opravdani razlog (Pielou 1966). No kako opažanje povezuje vrstu i korisnika u vremenu i prostoru, korisnici vrše reprezentativno samo-uzorkovanje vlastitog ponašanja i interakcije sa vrstama. Kako u ljudskom opažanju sudjeluje kolektiv ljudi čija samo-uzorkovanja ne ovise jedna o drugima, to je kontinuirani proces bez prestanka, jedini način na koji se može promatrati jest kao kolekcija. Time postoji opravdani razlog korištenja kolekcije za predviđene analize pomoću kojih je moguće opisati ponašanje građana znanstvenika i njihovoj interakciji sa živim svijetom.

Tablica 1. Broj taksonomskih kategorija za istraživane taksonomske razine i njihova brojnost. BAP podaci bez atributa problem (vidi poglavlje 3.2.4.).

Taksonomska razina	Broj taksonomskih kategorija	Broj taksonomskih kategorija BAP	Broj jedinki	Broj jedinki BAP
Razred	128	81	6746339	3733517
Red	535	333	6741953	3733403
Porodica	2184	1229	6744844	3733440
Rod	9270	4918	6740284	3733349
Vrsta	24502	10783	6714238	3731557

3.2. Metode

3.2.1. Definiranje uzorka i njegove veličine, te njihova podjela

Kako se radi o kolekciji podataka, koja je vremenski određena kao period od 23 godine, uzorke sam definirao kao vremenske intervale od godinu dana. Interval uzorka počinje s 1. siječnja, a završava s 31. prosinca svake godine. Odlučio sam se za taj interval jer većina organizama ima životni ciklus i promjene ponašanja (čovjek također) koji se odvijaju unutar godinu dana. Također u radu Di Cecco i sur. (2021) pokazano je kako postoji velika varijabilnost broja opažanja kroz tjedne u godini, te je ta varijabilnost ponavljajuća za svaku godinu koja je povezana s ponašanjem ljudi. Veličinu uzorka definirao sam pomoću parametra broja opažanja (dalje u radu: nO).

3.2.2. Promatrani parametri

Za mjere raznolikosti promatram parametar broj vrsta (nV) kao funkciju nO . Mjere raznolikosti ne moraju biti nužno promatrane samo na razini vrste, nego je moguće promatrati i raznolikost ostalih viših taksonomskih razina (Krebs 1989). Kako promatrani podatci ne pripadaju samo jednoj skupini, taksonomski sam promatrani parametar dodatno podijelio na razine:

1. roda (dalje u radu: nR)
2. porodice (dalje u radu: nP)
3. reda (dalje u radu: nRE)
4. razreda (dalje u radu: nRA).

3.2.3. Računalne aplikacije

Podatke raznolikosti vrsta obradio sam iz .csv tablice uz pomoć računalnog jezika R u razvojnom okruženju Rstudio. Koristio sam R pakete za rad s tablicom i za računanje mjera raznolikosti: *data.table* i *vegan*; za prostornu raspodjelu: *spatstat*. Sažete prikaze rezultata izradio sam pomoću računalne aplikacije Microsoft Excel. Karte i geografsku obradu podataka izradio sam pomoću računalne aplikacije Qgis 3.14.

3.2.4. Problemi nalaza

Atribut *problem* (engl. *issue*) opisuje mogući problem s nalazom ukoliko ono postoji. To može biti problem: taksonomije u odnosu na popis vrsta područja (engl. *species checklist*); preračunavanja koordinata iz stranog koordinatnog sustava; preciznost georeferenciranja; itd. To ne znači kako su ti podaci netočni, već kako je potrebno provjeriti njihovu točnost prije korištenja u analizama. Za ovaj rad nije potrebno koristiti samo podatke bez problema jer stanje atributa problema (je li istinito ili ne) ovisi o provjeri te postoji vjerojatnost da je opažanje točno. Geografski je bitno samo da se opažanje nalazi unutar granica istraživanog područja. Zbog toga sam rezultate podijelio na: sve podatke i podatke bez problema (dalje u radu: BAP) Time sam dao raspon nalaza po točnosti za svaku analizu osim prostorne raspodjele podataka. Kako prostorna raspodjela ovisi o preciznosti georeferenciranja, korišteni su samo podatci zadovoljavajuće preciznosti manje ili jednake od 100 metara.

3.2.5. Mjere raznolikosti

3.2.5.1. Rang-abundancijski modeli

Kako bih izrazio mjeru heterogenosti, prvo sam odredio kojem tipu logaritamskih serija najbolje odgovara raspodjela podataka i za to sam koristio rang-abundancijske modele (Whittaker 1965, Krebs 1989, Wilson 1991). Za određivanje rang-abundancijskih modela koristio sam funkciju *radfit* iz paketa *vegan* (Oksanen i sur. 2022). Parametre funkcije *radfit* postavio sam na sljedeći način: obitelj je Poissonova raspodjela (*family = „poisson“*); AIC sam koristio kako bih odabrao najbolji model (*BIC = FALSE*) (Oksanen i sur. 2022). Ostali parametri su nepromijenjeni od unaprijed zadanih stanja (Oksanen i sur. 2022). Za grafički prikaz rezultata koristio sam funkciju *plot* s unaprijed zadanim stanjima (Oksanen i sur. 2022). Rezultati su također prikazani kao vrijednosti devijacije, AIC i BIC za modele: LNS, Mandelbrot, Preemption, Zipf i Null (engl. brokenstick; dalje u radu: Slomljena grana) (Oksanen i sur. 2022).

3.2.5.2. Logaritamski normalna raspodjela

Za podatke koji su rang-abundancijskim modelima pokazali logaritamski normalnu raspodjelu koristio sam Prestonovu logaritamski normalnu krivulju kako bih dobio bolji uvid u potpunost raspodjele (Preston 1948, Krebs 1989). Za logaritamski normalnu raspodjelu koristio sam funkciju *prestonfit* iz paketa *vegan* (Oksanen i sur. 2022). Parametre funkcije *prestonfit* postavio sam na sljedeći način: polovica vrijednosti oktava se prelijevaju u iduću veću oktavu (*tiesplit = TRUE*) (Oksanen i sur. 2022). Ostali parametri su nepromijenjeni od unaprijed zadanih stanja (Oksanen i sur. 2022). Oktave prate slijed binarnog logaritma ($\log_2 n$) (Oksanen i sur. 2022). Vrijednosti odstupanja opaženih od izračunatih vrijednosti oktava prikazao sam u obliku *Z* vrijednosti kako bi bile usporedive između različitih taksonomskih razina.

Jednadžba 2.

Jednadžba Z vrijednosti preuzeta iz Kreyszig (1979):

$$Z = \frac{x - \mu}{\sigma}$$

x = Vrijednost

μ = *Srednja vrijednost*

σ = *Standardna devijacija*

3.2.5.3. Akumulacijska krivulja

Akumulacijskom krivuljom odredio sam koliko se broj novih vrsta akumulira nasumičnim povećavanjem broja uzoraka (Gotelli i Colwell 2001, Oksanen i sur. 2022). Vrijednosti nasumične akumulacijske krivulje usporedio sam s vrijednostima akumulacijske krivulje koja promatra akumulaciju vrsta s uzorcima koji su kronološki poredani. Za obje akumulacijske krivulje koristio sam funkciju *specaccum* iz paketa *vegan* (Oksanen i sur. 2022). Parametre za nasumičnu akumulacijsku krivulju postavio sam na sljedeći način: metoda je nasumična (*method = random*), broj permutacija je 1000 (*permutations = 1000*), metoda za ekstrapolaciju broja vrsta je *bootstrap* (*gamma = „bootstrap“*) (Oksanen i sur. 2022).

Jednadžba 3.

Jednadžba ekstrapolacije vrsta bootstrap (Smith i van Belle 1984) preuzeta iz (Oksanen i sur. 2022):

$$S_P = S_0 + \sum_{i=1}^{S_0} (1 - p_i)^N$$

S_P = ekstrapoliran broj vrsta u ukupnom setu vrsta

S₀ = opaženi broj vrsta u kolekciji

p_i = frekvencija vrste

N = broj uzoraka u kolekciji

Ostali parametri su nepromijenjeni od unaprijed zadanih stanja (Oksanen i sur. 2022). Parametre za kronološku akumulacijsku krivulju postavio sam na sljedeći način: metoda je kronološka (*method = collector*) (Oksanen i sur. 2022). Ostali parametri su nepromijenjeni od unaprijed zadanih stanja (Oksanen i sur. 2022). Vrijednosti odstupanja kronoloških od nasumičnih vrijednosti krivulja prikazao sam u obliku Z vrijednosti (jednadžba 2., vidi poglavlje 3.2.5.2.) kako bi bile usporedive između različitih taksonomskih razina.

3.2.5.4. Rarefakcija

Rarefakcijom sam odredio broj vrsta pri minimalnoj, srednjoj, medijanoj i maksimalnoj vrijednosti broja opažanja za cijeli skup uzoraka. Za rarefakciju koristio sam funkciju *rarefy* iz paketa *vegan* (Oksanen i sur. 2022), koja koristi jednadžbu iz Hurlbert (1971) i Simberloff (1972). Parametre funkcije *rarefy* postavio sam na sljedeći način: veličinu poduzorka odredio sam kao minimalnu, srednju, medijanu i maksimalnu vrijednost broja opažanja za skup veličina svih uzoraka (tablica 2.) (*sample = c(min, mean, median, max)*) (Oksanen i sur. 2022). Ostali parametri su nepromijenjeni od unaprijed zadanih stanja (Oksanen i sur. 2022). Raspone pri minimalnoj, srednjoj, medijanoj i maksimalnoj vrijednosti prikazao sam u obliku Z vrijednosti (jednadžba 2., vidi potpoglavlje 3.2.5.2.) kako bi bile usporedive između različitih taksonomskih razina.

Tablica 2. Veličine minimalnog, srednjeg, medijanog i maksimalnog poduzorka za sve promatrane taksonomske razine. BAP - podatci bez atributa problem.

Taksonomska razina	Broj opažanja			
	Maksimalni	Srednji	Medijani	Minimalni
Razred	1306147	293319	70041	15469
Red	1305898	293128	69940	15458
Porodica	1305923	293254	70001	15451
Rod	1305384	293056	69965	15449
Vrsta	1300563	291923	69595	14991
Razred BAP	901029	162327	43890	9244
Red BAP	901008	162322	43883	9244
Porodica BAP	901021	162323	43891	9244
Rod BAP	901014	162320	43887	9241
Vrsta BAP	900430	162242	43872	9240

3.2.6. Ripleyjeva K-funkcija

K-funkcijom odredio sam prostornu raspodjelu svih nalaza zadovoljavajuće preciznosti koordinata, odnosno provjerio sam jesu li podatci sakupljeni nasumično ili se grupiraju po nekoj osnovi (primjerice oko prometnica ili gradova jer su najdostupniji ljudima) (Ripley 1976, Haase 1995, Baddeley i sur. 2016, Geldmann i sur. 2016). Za korištenje RK-funkcije odabrao sam nalaze koje imaju preciznost koordinata u metrima (engl. *coordinate precision in meters*) manju ili jednaku 100 metara. Zbog toga što RK-funkcija koristi metričku udaljenost, koordinate nalaza reprojicirao sam iz koordinatnog sustava WGS84 (EPSG: 4326) u ETRS89-prošireni / LAEA Europe (EPSG: 3035). Za to sam koristio računalnu aplikaciju Qgis. Pomoću funkcije *dodaj geometriju* (engl. *add geometry*) dodao sam x i y koordinate EPSG: 3035 sustava točkastom vektorskom sloju nalaza. Masku istraživanog područja izradio sam pomoću vektorskih slojeva kopnenih granica država i njihovih pripadajućih ekskluzivnih ekonomskih zona. Kopnene granice država preuzeo sam s mrežne stranice www.gadm.org, a ekskluzivne ekonomske zone s mrežne stranice www.marineregions.org. Cjelovito istraživano područje izradio sam pomoću funkcija spajanje vektorskih slojeva (engl. *merge vector layers*), rastapanje vektorskih slojeva (engl. *dissolve*), izbriši rupe (engl. *remove holes*) i reprojiciraj sloj (engl. *reproject layer*). Za izvođenje RK-funkcije koristio sam funkciju *Kest* iz paketa *spatstat* u računalnom okruženju R (Baddeley i sur. 2016). Parametre funkcije *kest* postavio sam na sljedeći način: za parametar korekcija odabrao sam granica (*correction* = „border“; $d_{ij} = r$; jednačba 1., vidi potpoglavlje 1.4.1.) (Baddeley i sur. 2016). Ostali parametri su

nepromijenjeni od unaprijed zadanih stanja (Baddeley i sur. 2016). Kako bih provjerio postoji li trend grupiranja opažanja po osnovi naseljenih područja, koristio sam točkasti vektorski sloj svih opažanja i pretvorio sam ga u toplinsku kartu koristeći funkciju *toplinska karta* (engl. *heatmap*). Veličinu ćelije rastera odredio sam na 10 x 10 kilometara. Toplinsku kartu sam preklopio s točkastim vektorskim slojem naseljenih područja preuzetog s mrežne stranice www.naturalearthdata.com koji sam koristio kao kovarijant.

4. REZULTATI

4.1. Logaritamske serije

4.1.1. Rang-abundancijski ili dominacija / raznolikost modeli

Za taksonomsku razinu vrsta LNS najbolje pristaje kao model raspodjele podataka, s odstupanjem (Dev. = 1012427; AIC = 1110190; BIC = 1110206) od izračunate logaritamski normalne serije. Raspodjela podataka BAP za taksonomsku razinu vrsta BAP također najbolje pristaje LNS modelu, ali s većim odstupanjem od LNS modela za podatke (Dev. = 1869119; AIC = 1906475; BIC = 1906490). Za podatke i podatke BAP model Slomljena grana (engl. *brokenstick* ili *Null*) ima najveće odstupanje, dok Mandelbrot model uopće ne pristaje podacima i podacima BAP (slika 15.). Za taksonomsku razinu roda LNS najbolje pristaje kao model raspodjele podataka, s odstupanjem (Dev. = 1043005; AIC = 1086109; BIC = 1086124) od izračunate logaritamski normalne serije. Raspodjela podataka BAP za taksonomsku razinu roda BAP također najbolje pristaje LNS modelu, s većim odstupanjem od LNS modela za podatke (Dev. = 1863005; AIC = 1882318; BIC = 1882331). Za podatke BAP LNS i Preemption modeli imaju slične vrijednosti odstupanja. Za podatke i podatke BAP, Slomljena grana model ima najveće odstupanje, dok Mandelbrot model uopće ne pristaje podacima i podacima BAP. Za taksonomsku razinu porodice LNS najbolje pristaje kao model raspodjele podataka, s odstupanjem (Dev. = 1804998; AIC = 1816668; BIC = 1816679) od izračunate logaritamski normalne serije. Raspodjela podataka BAP za taksonomsku razinu porodice BAP najbolje pristaje Preemption modelu, s manjim odstupanjem od LNS modela za podatke (Dev. = 680342; AIC = 685962; BIC = 685967). Za podatke LNS i Preemption modeli imaju najbližnje vrijednosti odstupanja. Za podatke i podatke BAP, Slomljena grana model ima najveće odstupanje, dok Mandelbrot model uopće ne pristaje podacima i podacima BAP. Za taksonomsku razinu reda LNS najbolje pristaje kao modelu raspodjele podataka, s odstupanjem (Dev. = 783408; AIC = 786798; BIC = 786807) od izračunate logaritamski normalne serije. Raspodjela podataka BAP za taksonomsku razinu reda BAP također najbolje pristaje LNS modelu, sa manjim odstupanjem od LNS modela za podatke (Dev. = 262539; AIC = 264294; BIC = 264301). Za podatke BAP Mandelbrot i Zipf modeli imaju slične vrijednosti odstupanja. Za podatke i podatke BAP Slomljena grana model ima najveće odstupanje, dok Mandelbrot model uopće ne pristaje podacima. Za taksonomsku razinu razreda model Zipf najbolje pristaje kao model raspodjele podataka, s odstupanjem (Dev. = 321324; AIC = 322127; BIC = 322133) od izračunatog Zipf modela. Raspodjela podataka BAP za taksonomsku razinu razreda BAP

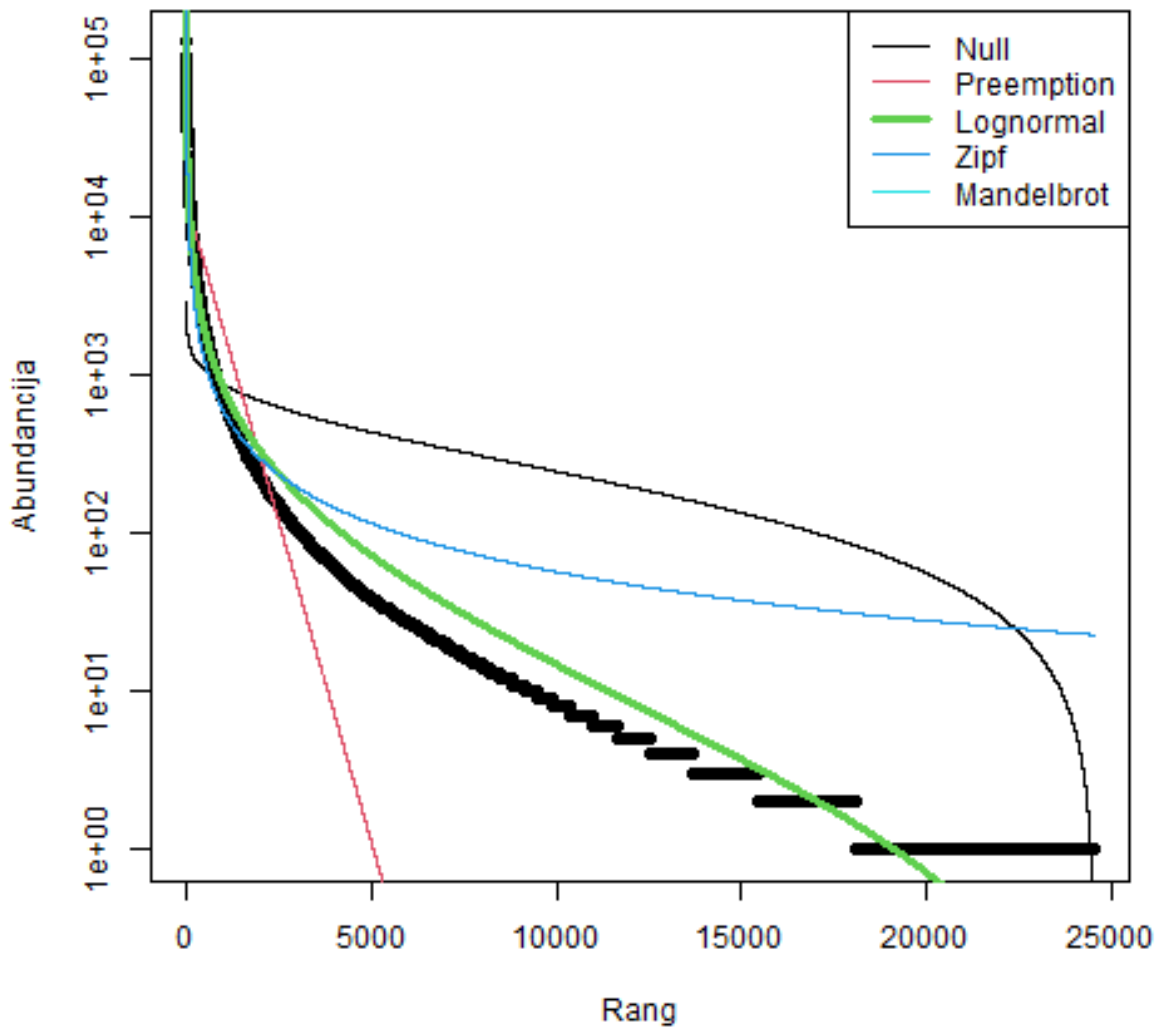
najbolje pristaje LNS modelu ($Dev. = 152424$; $AIC = 152842$; $BIC = 152847$). Za podatke BAP LNS, Mandelbrot i Zipf modeli imaju slične vrijednosti odstupanja. Za podatke LNS i Zipf modeli imaju slične vrijednosti odstupanja. Za podatke i podatke BAP, Slomljena grana model ima najveće odstupanje, dok Mandelbrot model uopće ne pristaje podacima, a za podatke BAP Preemption model uopće ne pristaje (tablica 3.).

Povećavanjem taksonomske razine podatci sve manje pristaju ponuđenim rang-abundancijskim modelima, što se vidi topološkim pregledom slikovnih rezultata (slika 15., prilog 3.-6.), iako su im vrijednosti odstupanja manje (tablica 3.), kao i prilikom razdvajanja podataka na podatke BAP (prilog 7.-11.). Za korisnike LNS model najbolje pristaje podacima od svih rang-abundancijskih modela (slika 16.).

Tablica 3. Pristajanje podataka rang-abundancijskim modelima podijeljenih na taksonomske razine i na podatke bez problema. Podebljane vrijednosti označavaju modele koji najbolje pristaju. Logaritamski normalni model (lognormal) najbolje pristaje većini podjelama podataka, osim porodici BAP kojoj najbolje pristaje Preemption model i razredu kojem pristaje Zipf. Svim podacima najmanje pristaje model Slomljene grane (sl. grana). NP- ne pristaje.

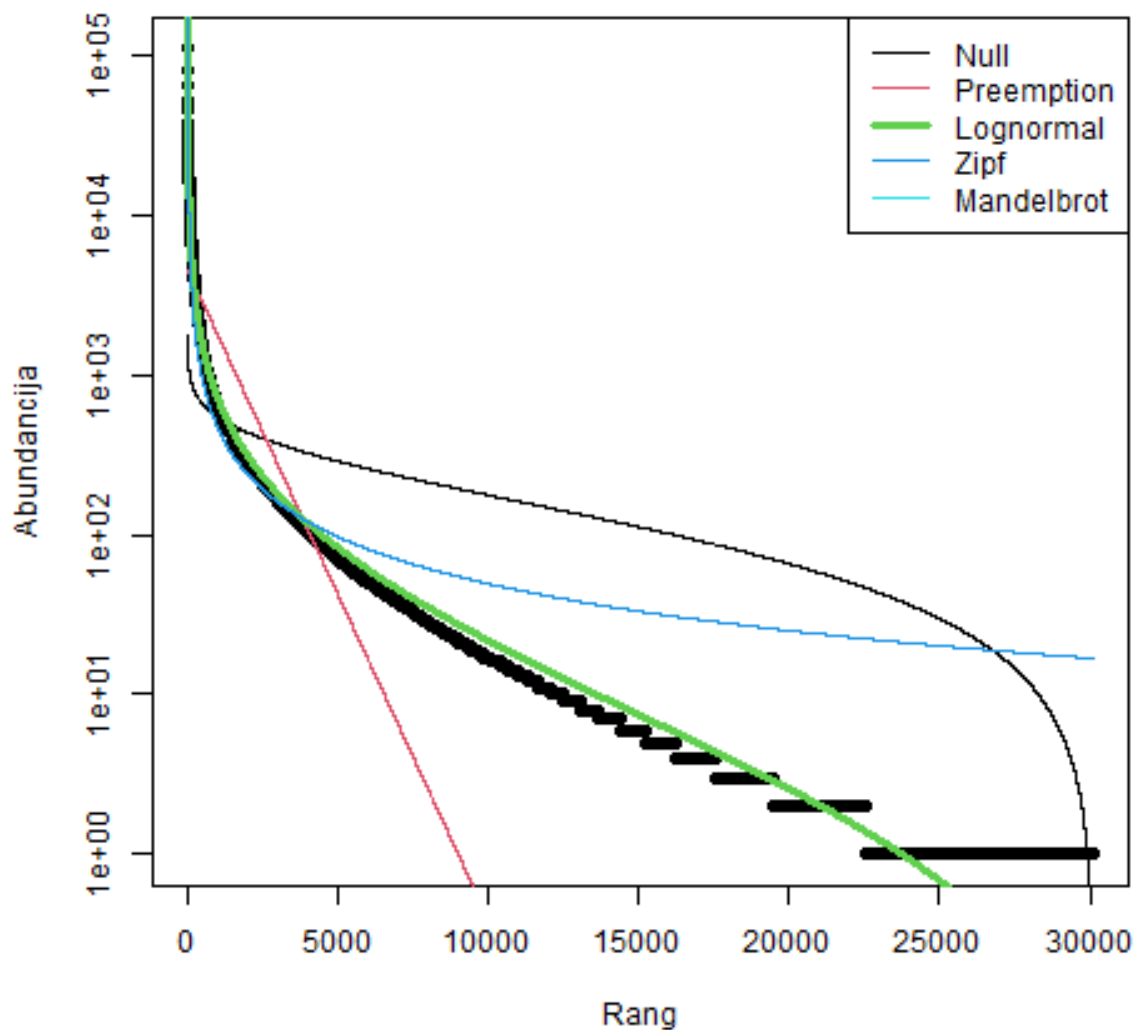
	Model	Podatci			Podatci BAP		
		Devijacija	AIC	BIC	Devijacija	AIC	BIC
Razred	Sl. grana	28284895	28285694	28285694	19385942	19386357	19386357
	Preemption	1779004	1779805	1779808	NP	NP	NP
	Lognormal	388509	389312	389318	152424	152842	152847
	Zipf	321324	322127	322133	178821	179240	179245
	Mandelbrot	NP	NP	NP	178821	179242	179249
Red	Sl. grana	21010388	21013774	21013774	15430296	15432047	15432047
	Preemption	4825485	4828873	4828877	2301431	2303184	2303188
	Lognormal	783408	786798	786807	262539	264294	264301
	Zipf	1790144	1793534	1793542	482954	484708	484716
	Mandelbrot	NP	NP	NP	432997	434753	434765
Porodica	Sl. grana	19898814	19910480	19910480	13558705	13564323	13564323
	Preemption	2526565	2538233	2538239	680342	685962	685967
	Lognormal	1804998	1816668	1816679	1964976	1970598	1970608
	Zipf	4076696	4088365	4088377	2981936	2987558	2987569
	Mandelbrot	NP	NP	NP	NP	NP	NP
Rod	Sl. grana	21377862	21420962	21420962	15687867	15707177	15707177
	Preemption	5660487	5703589	5703596	2113310	2132621	2132627
	Lognormal	1043005	1086109	1086124	1863005	1882318	1882331
	Zipf	3668567	3711671	3711686	3248919	3268233	3268246
	Mandelbrot	NP	NP	NP	NP	NP	NP
Vrsta	Sl. grana	24455885	24553644	24553644	16910934	16948287	16948287
	Preemption	8527812	8625573	8625581	3409274	3446629	3446636
	Lognormal	1012427	1110190	1110206	1869119	1906475	1906490
	Zipf	3799331	3897094	3897110	3461047	3498404	3498419
	Mandelbrot	NP	NP	NP	NP	NP	NP

Rang-abundancijski modeli za taksonomsku kategoriju vrste



Slika 15. Rang-abundancijski modeli za taksonomsku razinu vrsta. Podebljane crne linije označavaju raspodjelu podataka. Logaritamski normalni model najbolje pristaje raspodjeli podataka za taksonomsku razinu vrsta. Topološki je vidljivo odstupanje od izračunatog modela logaritamski normalne raspodjele.

Rang-abundancijski modeli za korisnike

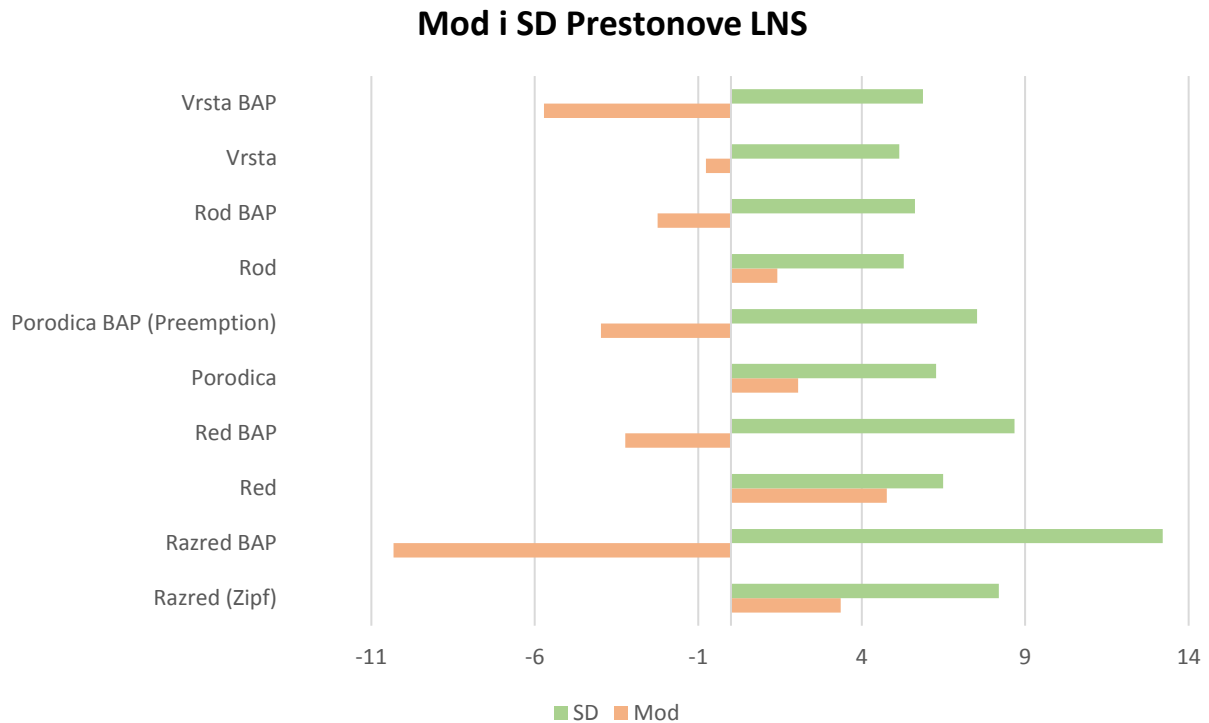


Slika 16. Rang-abundancijski modeli za raspodjelu korisnika. Podebljane crne linije označavaju raspodjelu podataka. Logaritamski normalni model najbolje pristaje raspodjeli podataka za korisnike. Topološki je vidljivo manje odstupanje od izračunatog modela logaritamski normalne raspodjele nego kod podataka za taksonomsku razinu vrsta.

4.1.2. Prestonova logaritamski normalna raspodjela

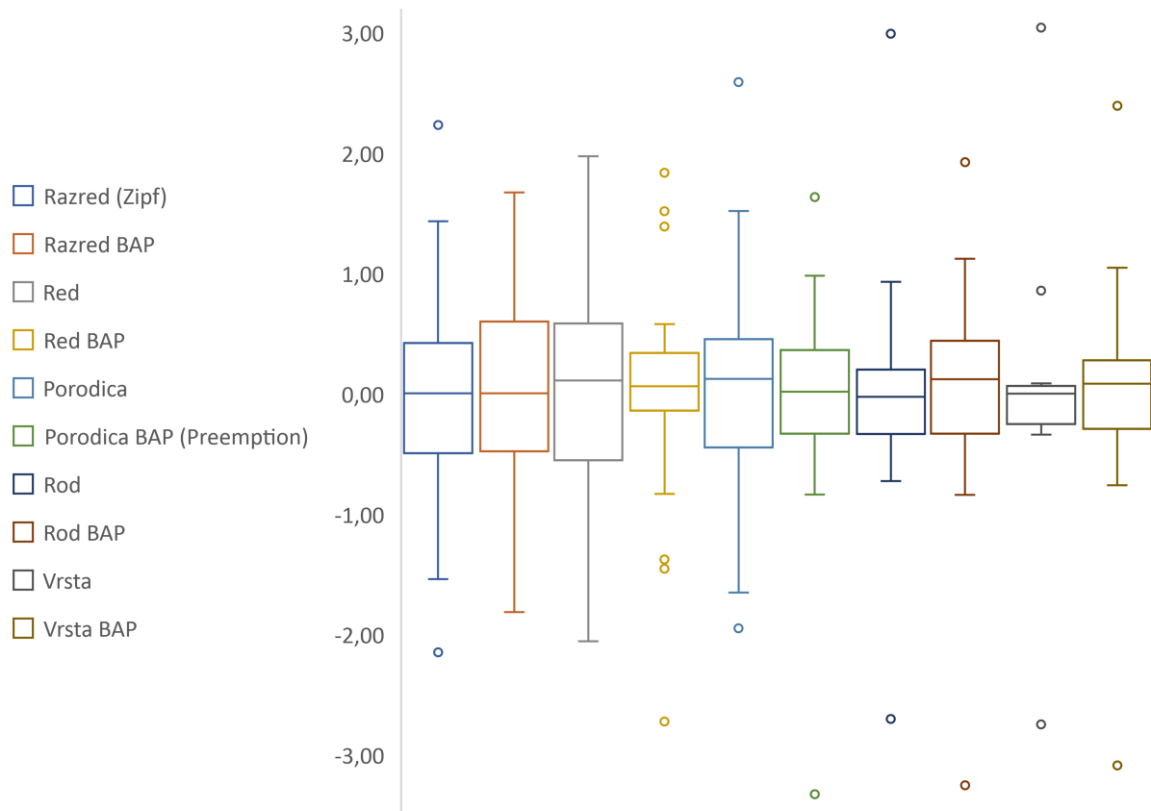
Za taksonomsku razinu vrsta moda PLNR nalazi se lijevo od oktave 1 (-0,7677118), a SD se nalazi na oktavi 16 -32 ($nV = 2137,5$). Skriveno područje nalazi se u blizini moda. Za taksonomsku razinu vrsta BAP, mod PLNR nalazi se znatno više lijevo od oktave 1 nego kod ukupnih podataka (-5,721139), a SD nalazi se na oktavi 1-2 ($nV = 2527,5$). Skriveno područje se nalazi desno od moda, blizu SD. Za taksonomsku razinu roda mod PLNR nalazi se u oktavi 1-2 (1,419412), a SD se nalazi na oktavi 64-128 ($nR = 628,5$). Skriveno područje nalazi se lijevo od moda. Za taksonomsku razinu roda BAP mod PLNR nalazi se lijevo od oktave 1 (-2,241882), a SD se nalazi na oktavi 8-16 ($nR = 544,5$). Skriveno područje nalazi se desno od moda, između mode i SD. Za taksonomsku razinu porodica mod PLNR nalazi se između oktava 2-4 i 4-8 (1,419412), a SD se nalazi na oktavi 256-512 ($nP = 121$). Skriveno područje nalazi se lijevo od moda. Za taksonomsku razinu porodica BAP, moda PLNR nalazi se lijevo od oktave 1 (-3,97843), a SD se nalazi na oktavi 8-16 ($nP = 139$). Skriveno područje nalazi se desno od moda, između moda i SD. Za taksonomsku razinu reda mod PLNR nalazi u oktavi 16-32 (4,762725), a SD se nalazi na oktavi 2048-4096 ($nRe = 12$). Skriveno područje nalazi se lijevo od moda blizu SD. Za taksonomsku razinu reda BAP mod PLNR nalazi se lijevo od oktave 1 (-3,231463), a SD se nalazi na oktavi 32-64 ($nRe = 25$). Skriveno područje nalazi se desno od moda, između moda i SD. Za taksonomsku razinu razreda mod PLNR nalazi se u oktavi 8-16 (3,360925), a SD se nalazi na oktavi 2048-4096 ($nRa = 4$). Skriveno područje nalazi se lijevo od moda, između moda i -SD. Za taksonomsku razinu razreda BAP mod PLNR nalazi se lijevo od oktave 1 (-10,32217), a SD se nalazi između oktava 8-16 i 16-32 ($nRa = 7,5$; $nRa = 10,5$). Skriveno područje nalazi se desno od moda, između moda i SD. Povećavanjem taksonomske razine skriveno područje se smanjuje, a podjelom ukupnih podataka na podatke BAP skriveno područje se povećava, te je mod udaljeniji od oktave 1 prema lijevoj strani i SD se povećava (slika 17., 19.). Najmanje odstupanje u Z vrijednostima predviđenih i opaženih oktava PLNR ima taksonomska razina vrsta, a najveće odstupanje ima taksonomska razina red. Najviše netipičnih vrijednosti ima taksonomska razina red BAP, a najmanje razred BAP i red. Najveće netipične vrijednosti imaju taksonomske razine vrsta i porodica BAP. Povećavanjem taksonomskih razina odstupanje vrijednosti se povećava, do razine razreda gdje je odstupanje manje od prethodne razine. Kod svih taksonomskih razina podjelom ukupnih podataka na podatke BAP odstupanje se povećava, osim kod reda i porodice gdje se smanjuje, ali se javlja više netipičnih vrijednosti (kod reda) ili su netipične vrijednosti veće (kod porodice) (slika 18.,

19.). Za korisnike PLNR ima mod u drugoj oktavi. Prva oktava negativno odstupa od izračunate vrijednosti, dok druga oktava odstupa pozitivno od izračunate vrijednosti (slika 20.).



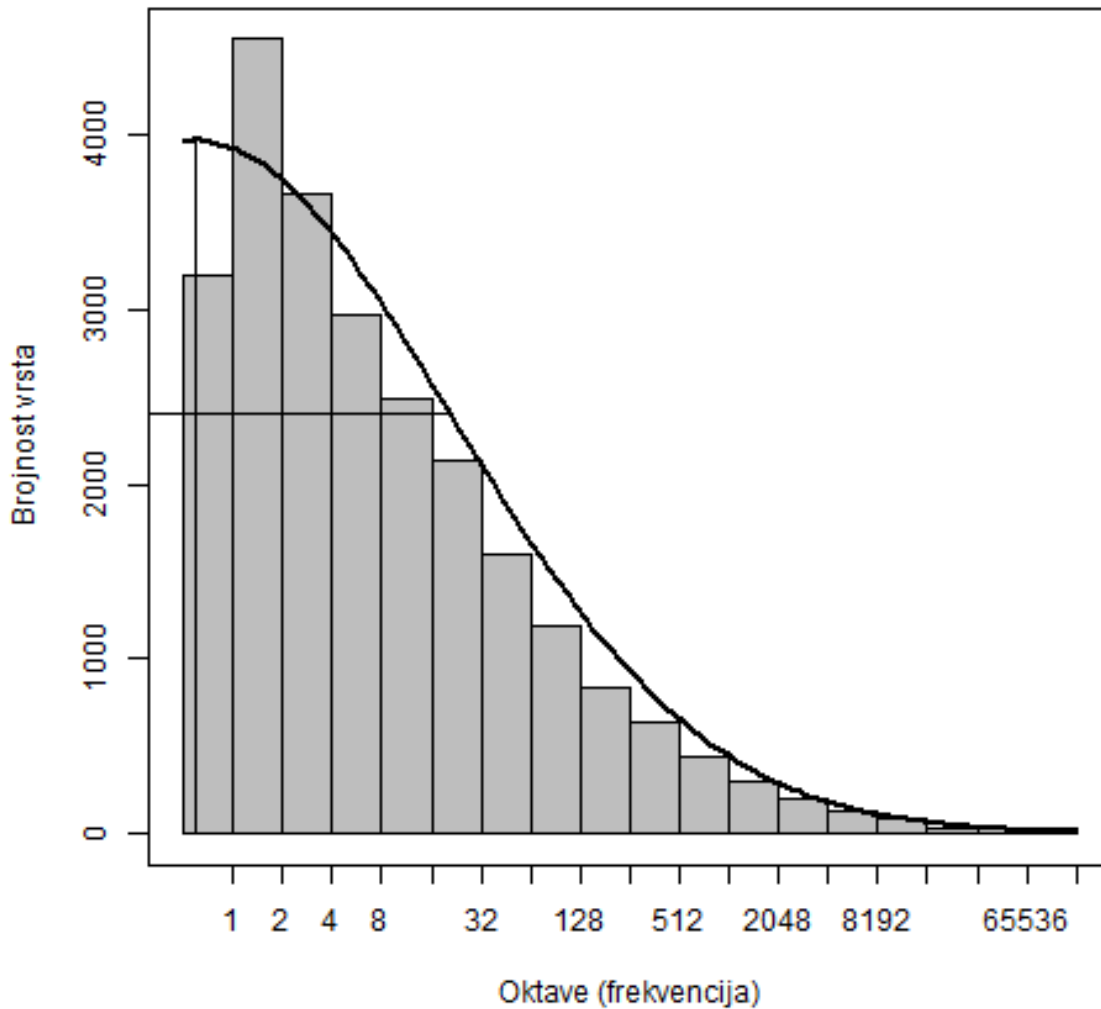
Slika 17. Položaj moda i standardne devijacije Prestonove logaritamski normalne raspodjele podataka taksonomskih razina sa i bez atributa problem (BAP). Porastom taksonomske razine mod i standardna devijacija se pomiču prema sve većoj pozitivnoj vrijednosti, tj. skriveno područje se smanjuje, do razine reda, nakon koje razred ima manju vrijednost moda. Kod taksonomskih razina BAP mod ide u negativnu vrijednost, tj. skriveno područje se povećava i ne pomicanje ne prati pravilnost. Standardna devijacija BAP se povećava porastom taksonomske razine BAP, tek nakon vrsta BAP koja ima veću vrijednost standardne devijacije od roda BAP. Taksonomska razine porodice BAP i razreda ne pripadaju modelu LNS.

Z vrijednosti odstupanja oktava predviđenih i opaženih PLNS



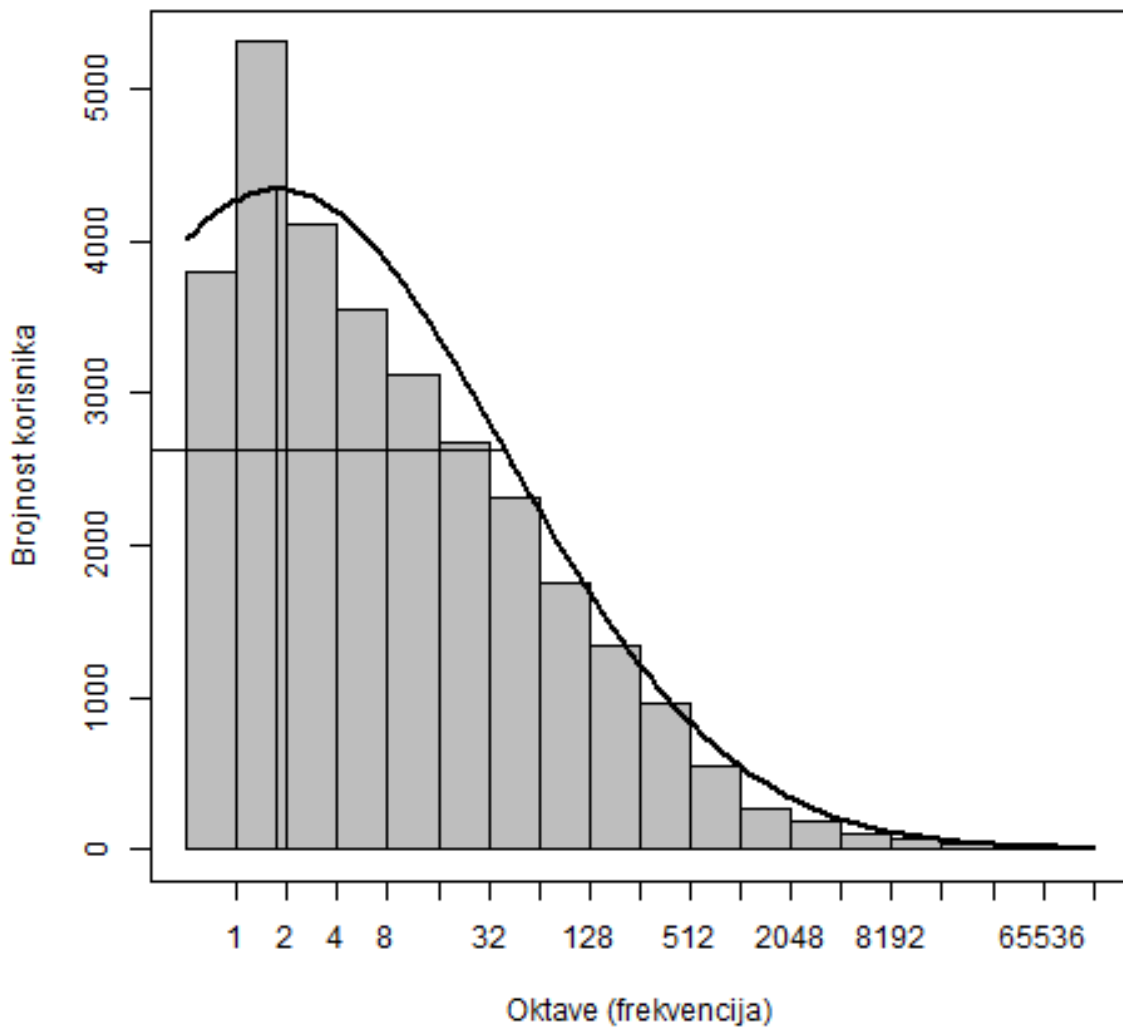
Slika 18. Odstupanja predviđenih i opaženih oktava Prestonove logaritamski normalne raspodjele podataka taksonomskih razina, sa i bez atributa problem (BAP), izražene u Z vrijednostima. Porastom taksonomske razine raspon odstupanja predviđenih i opaženih oktava se širi, do razine reda, nakon koje razred ima manji raspon odstupanja s netipičnim vrijednostima. Kod taksonomskih razina BAP raspon odstupanja predviđenih i opaženih oktava ne prati istu pravilnost. Taksonomska razine porodice BAP i razreda ne pripadaju modelu LNS.

Prestonova logaritamski normala raspodjela za taksonomsku razinu vrste



Slika 19. Prestonova logaritamska raspodjela taksonomske razine vrsta. Moda izračunate raspodjele nalazi se u prvoj oktavi (vrste s opažanjem). Prva oktava negativno odstupa od izračunate vrijednosti, dok druga oktava (vrste s dva opažanja) pozitivno odstupa, tome je razlog što prema Preston (1948) polovina broja vrsta iz prve oktave se prenosi u drugu oktavu i tako redom do zadnje oktave. To je pretpostavka kako prava raspodjela ovisi o dovoljnom broju uzoraka i polovina vrsta iz oktava pripada višim oktavama jer zajednica nije dovoljno uzorkovana.

Prestonova logaritamski normala raspodjela korisnika



Slika 20. Prestonova logaritamska raspodjela korisnika. Moda izračunate raspodjele nalazi se u drugoj oktavi (korisnici sa jednim opažanjem). Prva oktava negativno odstupa od izračunate vrijednosti, dok druga oktava (vrste s dva opažanja) pozitivno odstupa.

4.2. Akumulacijske krivulje

Srednje vrijednosti odstupanja od nasumičnih vrijednosti krivulja akumulacija, za podatke i podatke BAP, prate zasebne trendove. Trend za podatke ima veću depresiju pri uzorku 4, te nagli uzlaz pri uzorku 6 u usporedbi s trendom podataka BAP. Depresija trenda podataka BAP je pri uzorku 8 i plića je, te ima blaži uzlaz od depresije trenda podataka. Pri uzorku 17 smanjuje se razlika trendova, te se oni prate nagibom i sličnim vrijednostima. Odstupanja kronoloških od nasumičnih vrijednosti krivulja akumulacija prati trend srednje vrijednosti za gotovo sve taksonomske razine, osim za taksonomsku razinu vrste. Trend za taksonomsku razinu vrsta bliže prati trend srednje vrijednosti podataka BAP. Vrijednosti trendova za podatke imaju širu distribuciju od vrijednosti podataka BAP. Trendovi SD i SD BAP nalikuju inverznoj logističkoj funkciji. Raspodjela vrijednosti SD-a je šira od distribucije vrijednosti SD BAP-a i ima manje netipičnih vrijednosti. Najmanji raspon vrijednosti SD-a i SD BAP-a nalazi se oko točke infleksije trendova, povećavajući se prema krajevima trendova (slika 21., 22., 23.). Odstupanje kronološke i nasumične akumulacijske krivulje može se također dodatno istaknuti jednadžbama koje pristaju krivuljama. Nasumičnim vrijednostima pristaje logaritamska jednadžba:

$$y = 6317,7\ln(x) + 4213,8$$

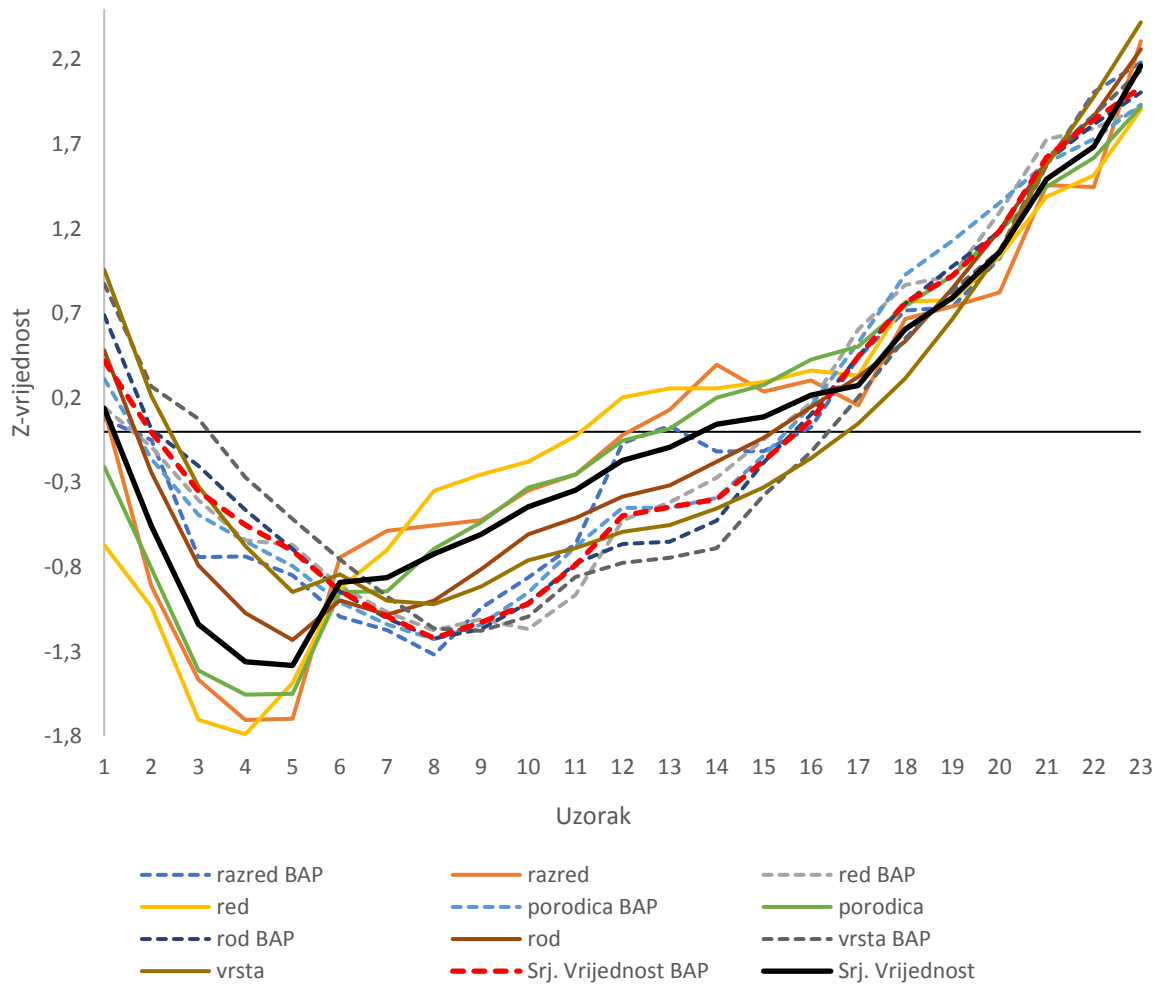
$$R^2 = 0,9937$$

dok kronološkim vrijednostima pristaje polinomna jednadžba drugog stupnja (slika 23.):

$$y = 14,083x^2 + 648,77x + 1421,1$$

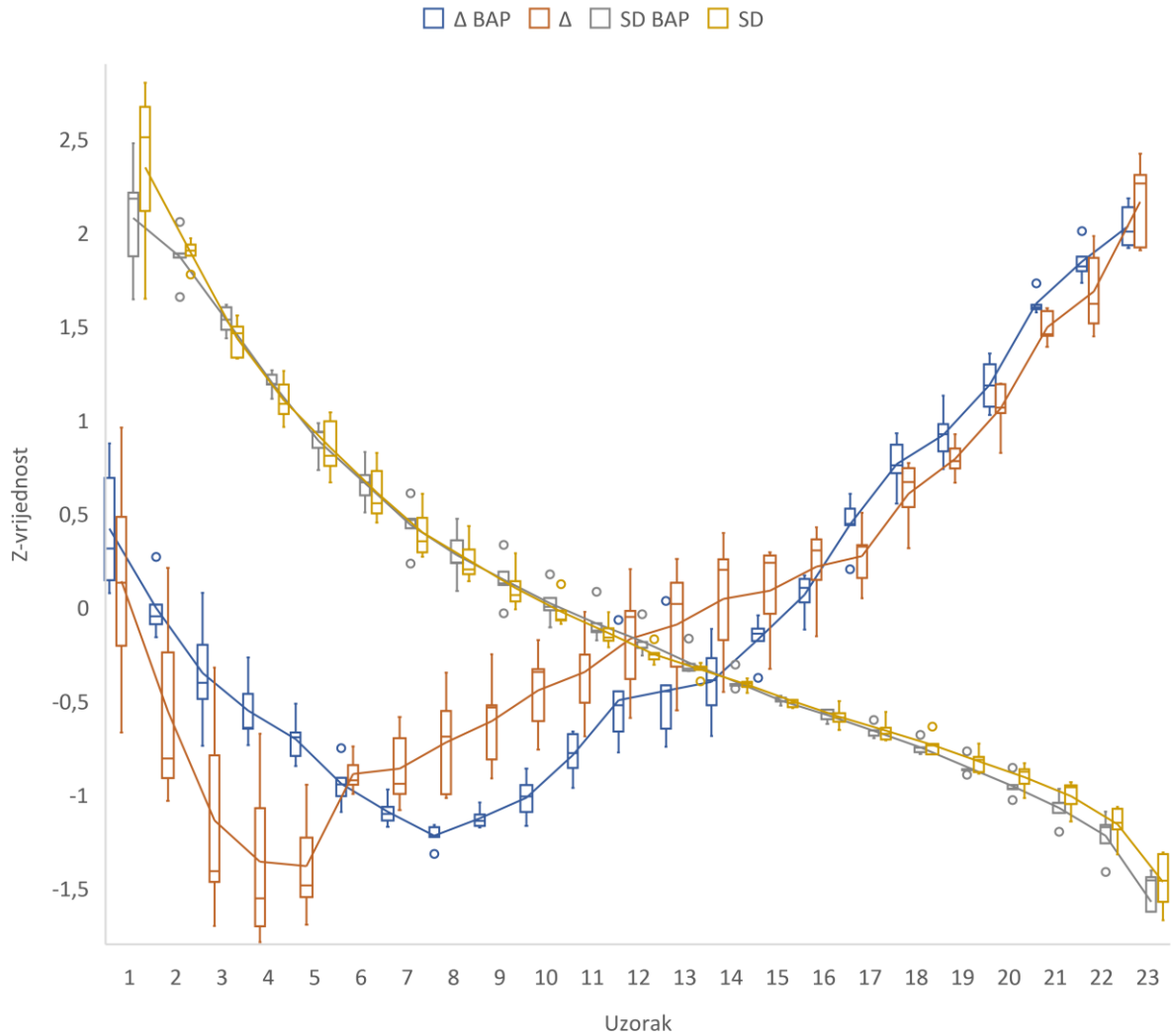
$$R^2 = 0,9958.$$

Odstupanja nasumičnih i opaženih akumulacijskih krivulja



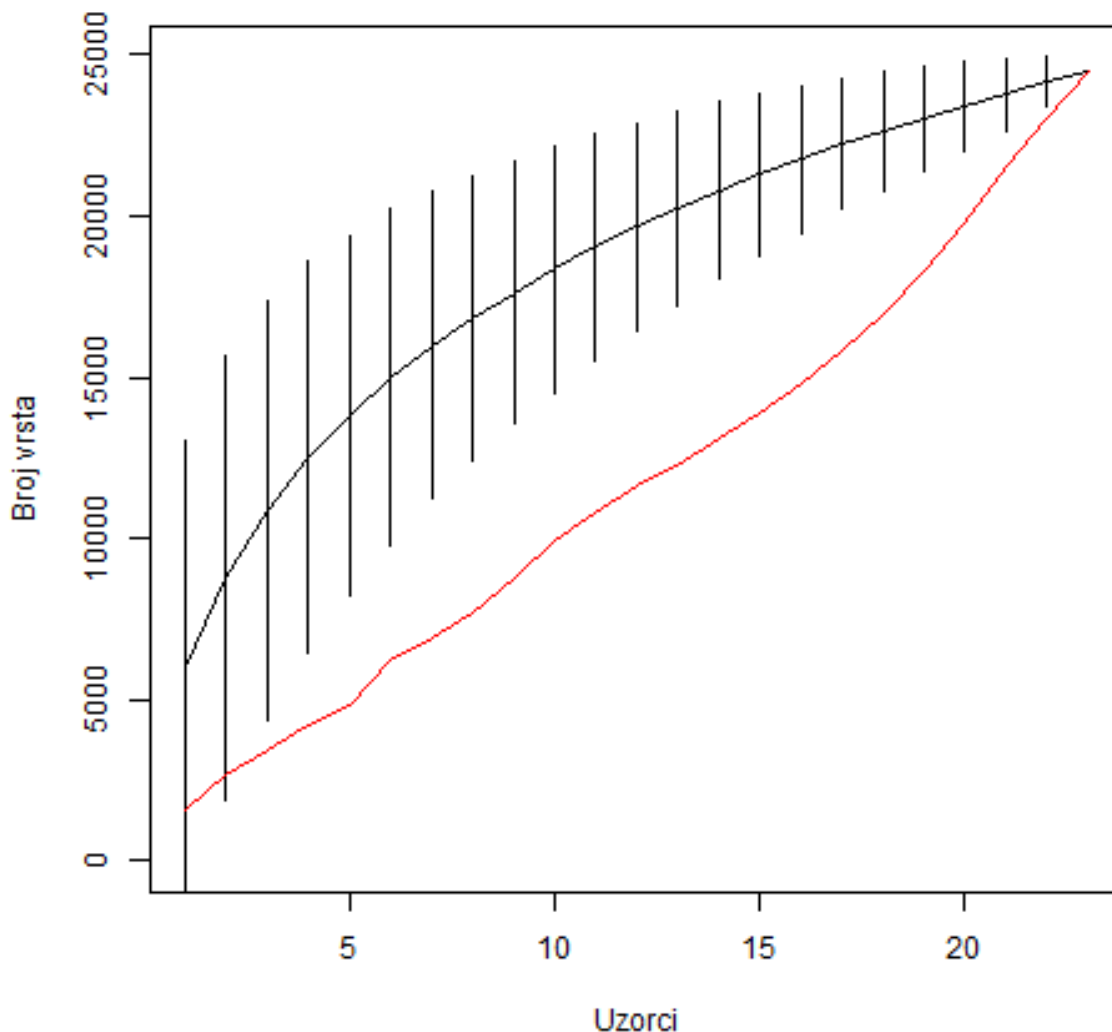
Slika 21. Odstupanja nasumičnih i opaženih akumulacijskih krivulja za sve promatrane taksonomske razine za podatke s i bez atributa problema (BAP).

Rasponi odstupanja nasumičnih i opaženih akumulacijskih krivulja i rasponi standardnih devijacija nasumičnih krivulja



Slika 22. Rasponi odstupanja (Δ) nasumičnih i opaženih akumulacija novih svojti i raspon standardnih devijacija (SD) nasumičnih krivulja za podatke sa i bez atributa problema (BAP) svih promatranih taksonomskih razina.

Nasumična i kronološka akumulacijska krivulja za taksonomsku razinu vrste



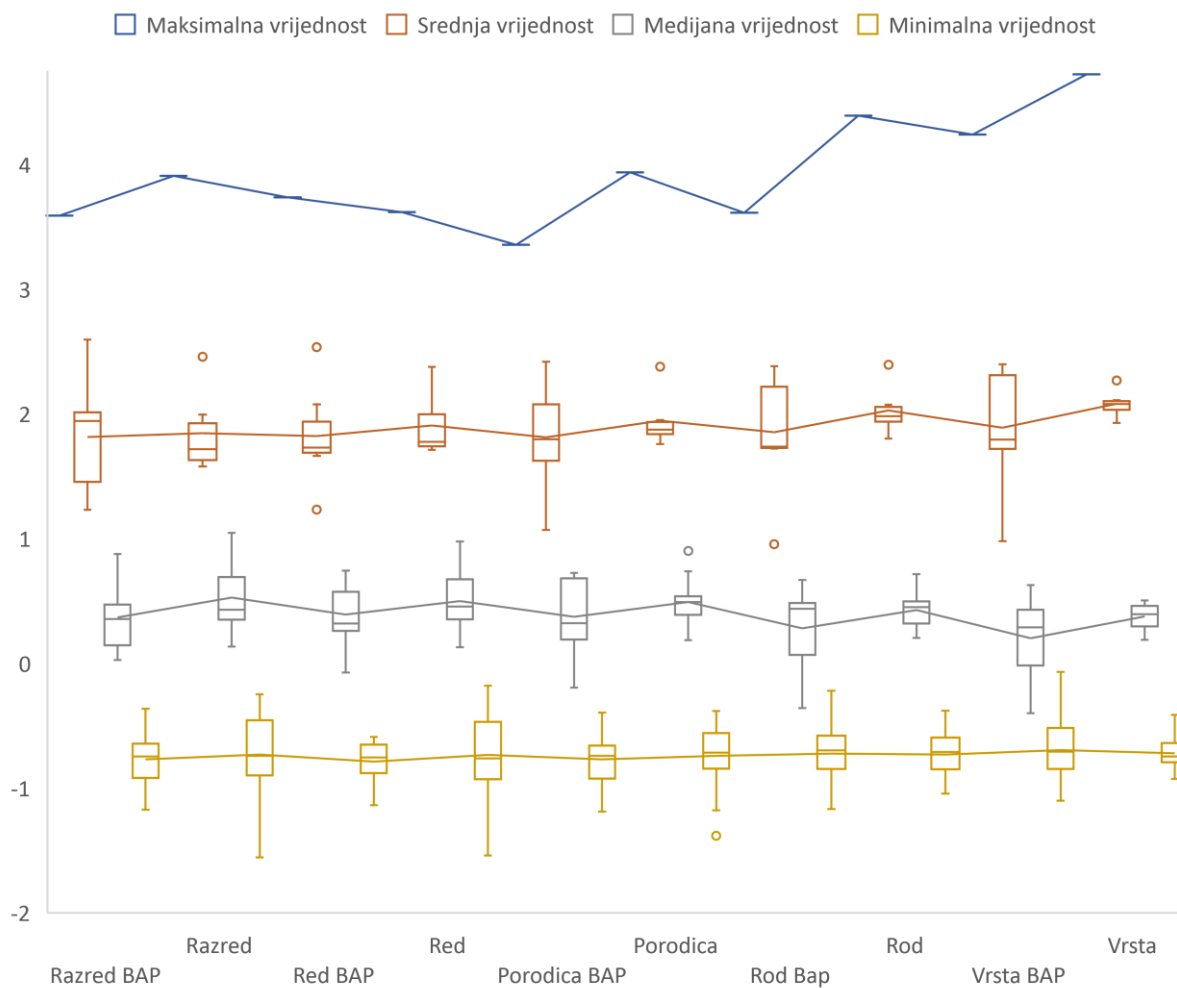
Slika 23. Nasumična (crna) i kronološka (crvena) akumulacijska krivulja novih vrsta. Standardna devijacija nasumične akumulacije predstavljene su kao okomite crne linije. Nasumičnoj krivulji pristaje logaritamska jednačzba: $y = 6317,7\ln(x) + 4213,8$ ($R^2 = 0,9937$), a kronološkoj pristaje polinomna jednačzba drugog stupnja: $y = 14,083x^2 + 648,77x + 1421,1$ ($R^2 = 0,9958$).

4.3. Rarefakcija

Za taksonomsku razinu vrsta rasponi vrijednosti nV rarefakcije pri minimalnoj, medijanoj i srednjoj vrijednosti nO se ne preklapaju. Za taksonomsku razinu vrsta BAP rasponi vrijednosti nV BAP rarefakcije pri minimalnoj i medijanoj vrijednosti nO se preklapaju. Maksimalna vrijednost nV BAP pri minimalnoj vrijednosti nO doseže donji kvartal medijane vrijednosti nO, a minimalna vrijednost nV BAP pri medijanoj vrijednosti nO skoro doseže gornji kvartal

minimalne vrijednosti nO . Minimalna vrijednost nV BAP pri srednjoj vrijednosti nO blizu je maksimalne vrijednosti nV BAP pri medijanoj vrijednosti nO . Raspodjela vrijednosti nV BAP pri minimalnoj, medijanoj i srednjoj vrijednosti nO ima veći raspon od distribucije vrijednosti nV , a maksimalni nV BAP je manji od maksimalnog nV . Za taksonomsku razinu roda rasponi vrijednosti nR rarefakcije pri minimalnoj, medijanoj i srednjoj vrijednosti nO se ne preklapaju. Za taksonomsku razinu roda BAP rasponi vrijednosti nR BAP rarefakcije pri minimalnoj i medijanoj vrijednosti nO se preklapaju. Netipična vrijednost nR BAP pri srednjoj vrijednosti nO blizu je maksimalne vrijednosti nR BAP pri medijanoj vrijednosti broja opažanja. Raspodjela vrijednosti nR BAP pri minimalnoj, medijanoj i srednjoj vrijednosti nO ima veći raspon od distribucije vrijednosti nR , a maksimalni nR BAP je manji od maksimalnog nR . Za taksonomsku razinu porodica rasponi vrijednosti nP rarefakcije pri minimalnoj, medijanoj i srednjoj vrijednosti nO se ne preklapaju. Za taksonomsku razinu porodica BAP rasponi vrijednosti nP BAP rarefakcije se ne preklapaju. Raspodjela vrijednosti nP BAP pri medijanoj i srednjoj vrijednosti nO ima veći raspon od distribucije vrijednosti nP , osim pri minimalnoj vrijednosti nO koja ima sličan raspon vrijednosti kao i kod porodice. Maksimalni nP BAP je manji od maksimalnog nP . Za taksonomsku razinu reda rasponi vrijednosti nRE rarefakcije pri minimalnoj, medijanoj i srednjoj vrijednosti nO se ne preklapaju. Za taksonomsku razinu reda BAP rasponi vrijednosti nRE BAP rarefakcije se ne preklapaju. Raspodjela vrijednosti nRE BAP pri minimalnoj i srednjoj vrijednosti nO ima manji raspon od distribucije vrijednosti nRE , osim pri medijanoj vrijednosti broja opažanja koja ima sličan raspon vrijednosti kao i kod reda. Maksimalni nRE BAP je veći od maksimalnog nRE . Za taksonomsku razinu razreda rasponi vrijednosti nRA rarefakcije pri minimalnoj, medijanoj i srednjoj vrijednosti nO se ne preklapaju. Za taksonomsku razinu razreda BAP rasponi vrijednosti nRA BAP rarefakcije se ne preklapaju. Raspodjela vrijednosti nRA BAP pri minimalnoj vrijednosti nO je manjeg raspona od distribucije vrijednosti nRA . Raspodjela vrijednosti nRA BAP pri medijanoj vrijednosti nO je slična distribuciji vrijednosti nRA . Raspodjela vrijednosti nRA BAP pri srednjoj vrijednosti nO ima veći raspon od distribucije vrijednosti nRA . Maksimalni nRA BAP je veći od maksimalnog nRA . Povećavanjem taksonomske razine, rasponi raspodjela broja taksonomskih kategorija se povećavaju, pri minimalnom, medijanom i srednjem broju opažanja. Povećavanjem taksonomske razine, vrijednost nV pri maksimalnom broju opažanja smanjuje se u relativnom odnosu na srednju vrijednost nO . Taj trend se mijenja do taksonomske razine porodice nakon koje se povećava za taksonomsku razinu reda i razreda (slika 24., 25.).

Rasponi rarefakcijskih krivulja



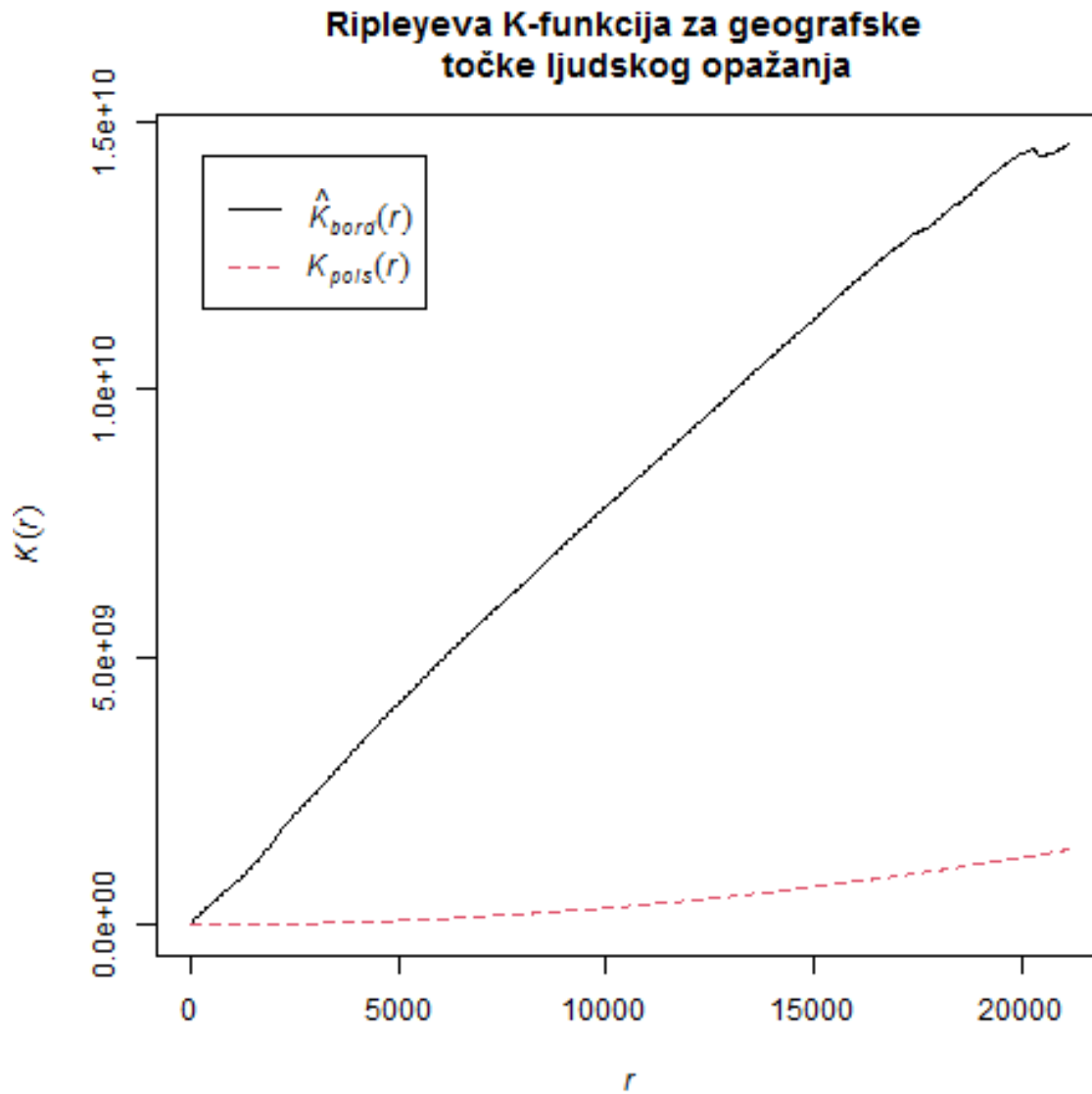
Slika 24. Rasponi Z vrijednosti rarefakcije uzoraka za sve taksonomske razine i taksonomske razine BAP pri maksimalnoj, srednjoj, medijanoj i minimalnoj vrijednosti broja opažanja za skup svih uzoraka.



Slika 25. Rarefakcija uzoraka za taksonomsku razinu vrste i njihovo bogatstvo vrsta pri poduzorku minimalne, medijane, srednje i maksimalne vrijednosti broja opažanja poduzorka.

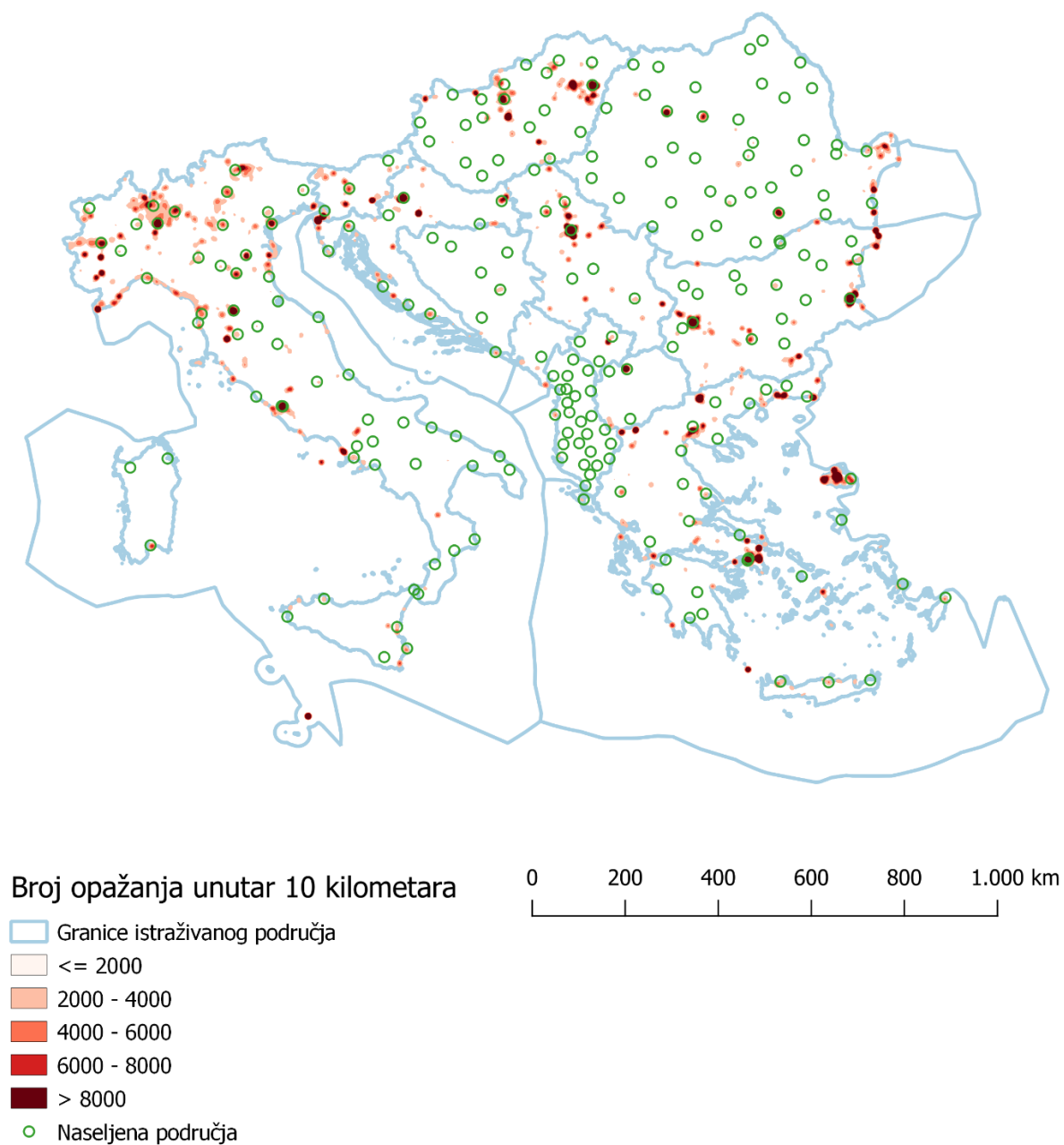
4.4. Ripleyjeva K-funkcija

Od 6714238 zapisa 1653680 ih je zadovoljavajuće točnosti (≤ 100 metara). Ripleyjeva K-funkcija za podatke znatno odstupa nagibom iznad RK-funkcije za prostorno nasumične podatke generirane za isti broj nalaza, što ukazuje na znatnu grupiranost podataka (slika 26.). Preklapanjem karte gustoće svih opažanja i karte naseljenih područja vidljiv je trend preklapanja gustoće veće od 2000 opažanja po 10 km^2 s naseljenim područjima. Pored tog trenda može se primijetiti i kako se neka područja veće gustoće nalaze na područjima obale i zaštićenim područja, npr. geopark otok Lesbos ili nacionalni park Plitvička jezera. (slika 27.).



Slika 26. Ripleyjeva K-funkcija za točke nalaza preciznosti ≤ 100 metara ($\hat{K}_{bord}(r)$) i nasumičnih točaka jednakog broja ($\hat{K}_{pois}(r)$). $\hat{K}_{bord}(r)$ znatno odstupa iznad $\hat{K}_{bord}(r)$ što opisuje iznimnu grupiranost nalaza.

Karta područja s gustoćama većim od 2000 opažanja po 10 km² i točaka naseljenih područja



Slika 27. Karta područja s gustoćama većim od 2000 opažanja po 10 km² i točke naseljenih područja. Područja s gustoćama većim od 2000 opažanja po 10 km² pokazuju trend preklapanja s naseljenim područjima.

5. RASPRAVA

Dosadašnje je shvaćanje kako se ljudsko opažanje, time i sama građanska znanost, može koristiti u istraživanjima raznolikosti vrsta. Podatci se tretiraju kao pristrani ili netočni, te se ispravljaju i prilagođavaju potrebama istraživanja raznolikosti vrsta. Takva ispravljanja i prilagodbe iz teoretskih razloga nemaju valjana opravdanja, osim što iz praktičnih razloga mogu ponuditi besplatni izvor velikog broja podataka. Rezultati svih analiza su u skladu s tvrdnjom iz rada Specht i Lewandowski (2018) kako se radi o drugačijim podacima od onih koje proizvode znanstvenici, a ne netočnim podacima koje je potrebno ispravljati, te smatram kako im je potrebno pronaći valjanu svrhu. Ljudsko opažanje ne zadovoljava osnovni uvjet metode uzorkovanja raznolikosti, a to je nasumičnost uzorkovanja (Preston 1948), što je i vidljivo iz svih rezultata mjera raznolikosti. U radu Preston (1948) već je naveden problem pojedinca kao metode uzorkovanja raznolikosti; pojedinac ide za rijetkim vrstama, a ne za učestalim. To se može smatrati točnom tvrdnjom i za skupinu ljudi (kolektiv ili prosječnu osobu), što se uočava u svim rezultatima mjera raznolikosti kao odstupanje od nasumičnih vrijednosti. Manjak nasumičnosti možemo uočiti i kada promatramo lokalitete na kojemu su opažene jedinke. Nasumični ili uređeni (sistematski) lokaliteti su također jedan od uvjeta za reprezentativan uzorak (Hankin i sur. 2019). Grupiranost lokaliteta vidljiva je iz rezultata Ripleyjeve K-funkcije za lokacije kao i trend preklapanja s naseljenim područjima, te rezultate podupiru i podatci iz rada Geldmann i sur. (2016) i Di Cecco (2021). Može se pretpostaviti kako grupiranost i trend opažanja na antropogenim staništima također utječe i na sastav vrsta u podacima, jer na takvim staništima dolazi do promjene sastava vrsta (Laurance i sur. 2007), koje građani znanstvenici opažaju. Samo vrste koje životnim stilom i prilagodbama mogu preživjeti takve uvjete naseljavaju poremećena i rubna staništa (ruderalne i pionirske vrste) (Gascon i sur. 2000), što posljedično dovodi do toga da građani znanstvenici opažaju više vrsta koje se nalaze na takvim staništima, a manje vrste koje se nalaze na prirodnim staništima. Još jedan razlog zašto prikupljeni podatci nisu primjereni za istraživanje raznolikosti je taj što oni nisu uzorci, tj. sami građani znanstvenici nisu međusobno ovisni jedni o drugima u svom sakupljanju. Podatci nisu sakupljeni u isto vrijeme, na istoj lokaciji ni u istim intervalima. Takvo uzorkovanje rezultira kolekcijom podataka, a ne uzorkom, što nije statistički reprezentativan (nasumičan) uzorak veće populacije jedinki (Pielou 1966). Iz navedenih razloga ljudsko opažanje bi se trebalo koristiti za opisivanje interakcije ljudi s raznolikošću, tj. promatrati ju kao metodu samouzorkovanja korisnika. Promotrimo svojstva koje se mogu pretpostaviti za ljudsko opažanje kao metode samouzorkovanje korisnika:

1. Prostorno nije nasumična – Grupira se po antropocentričkoj osnovi
2. Taksonomski nije nasumična – Preferira skupine po antropocentričkoj osnovi
3. Vremenski nije nasumična – Preferira vrijeme prema antropocentričkoj osnovi
4. Ne stvara uzorke – Stvara kolekciju podataka aktivnosti ljudi

Iz ovih svojstava vidljivo je kako ljudsko opažanje stvara kolekciju podataka aktivnosti čovjeka povezanu Toblerovim prvim pravilom geografije sa staništima i vrstama (Tobler 1970), tj. uzorkuje kada i gdje se osoba susrela s vrstom ili skupinom koju je odlučila zabilježiti. Ako se odnosi na neko područje onda reprezentativno uzorkuje u kojem vremenskom intervalu i na kojim mjestima se međusobno neovisna skupina ljudi susreće s vrstama ili skupinama koje odlučuju zabilježiti. Takav skup podataka može opisati prosječnu interakciju skupine ljudi s organizmima i staništima na određenom području. Interakcija se može opisati mjerama raznolikosti i selektivnošću. Selektivnost su taksonomski, vremenski i prostorni interesi ljudi, a mjere raznolikosti vrsta izražavaju intenzitet i vrstu interakcije ljudi s organizmima. Promatrajući svojstva podataka ljudskog opažanja iz te perspektive moguće je odrediti:

1. s kolikim udjelom ukupne opisane raznolikosti nekog područja prosječan građanin znanstvenik (ili kolektiv) može biti u interakciji (asimptota akumulacijske krivulje)
2. s koliko vrsta (bogatstvom vrsta) je u interakciji (asimptota rarefakcijske krivulje)
3. s koliko vrsta je često, a s koliko vrsta je rijetko u interakciji (logaritamske i logaritamski normalne krivulje).

Pri dovoljno velikom ili teoretski beskonačnom broju uzorkovanja ili veličini uzorka, određenom metodom na nekom prostoru ili u zajednici, rarefakcijska i akumulacijska krivulja dostiže asimptotu (Gotelli i Colwell 2001). Do tog svojstva dolazi zbog toga što svaki prostor, tj. zajednica, ima konačni broj vrsta koje se u njoj nalaze (Simberloff i Wilson 1970). Pojava asimptote u rarefakcijskim i akumulacijskim krivuljama također ukazuje istraživaču kako su sve vrste koje je korištena metoda mogla uzorkovati i koje se nalaze u istraživanom prostoru ili zajednici uzorkovane te kako je postignut reprezentativan uzorak (Gotelli i Colwell 2001). Iz rezultata je vidljivo kako ljudsko opažanje nije doseglo asimptotu rarefakcijske i akumulacijske krivulje pri taksonomskoj razini vrste te time ukazuje na nedovoljnu veličinu i broj uzoraka za uzorkovanje svih vrsta s kojima je ljudsko opažanje u interakciji. No iz rezultata je vidljivo kako je za taksonomski više razine potrebna manja veličina i broj uzoraka za uzorkovanje svih skupina s kojima je ljudsko opažanje u interakciji. Ljudsko opažanje, za razliku od drugih metoda uzorkovanja raznolikosti, nije selektivno prema određenim skupinama, što je vidljivo pregledom carstava koja su opažena (Search 2022b). Jedina metoda uzorkovanja koja konkurira

ljudskom opažanju obzirom na taksonomsku širinu uzorkovanja je metagenomika (Bohmann i sur. 2014). Samim time, pojava asimptote kod rarefakcijske i akumulacijske krivulje, hipotetski, zahtjeva iznimno velike uzorke. Na prostorima koji se nalaze na kontinentima, pogotovo tropskim ili suptropskim područjima, upitno je bi li došlo do asimptota tih krivulja kod ljudskog opažanja (s trenutnim brojem opažanja koji se nalaze u bazi GBIF) zato što su to područja veće biološke raznolikosti od ostalih (Krebs 1989). Hipotetski idealno područje za promatranje pojave asimptote kod tih krivulja bilo bi izolirano područje umjerene klime s relativno malim bogatstvom vrsta i relativno velikim brojem opažanja (Krebs 1989). Pronalazak takvog područja mogao bi se pretpostaviti pomoću teorije otočne biogeografije (MacArthur i Wilson 1967) i time bi idealno područje za pronalazak asimptota kod rarefakcijske i akumulacijske krivulje bio otok umjerene klime, s velikim brojem opažanja. Osim pronalaska asimptote kod rarefakcijske i akumulacijske krivulje, pretpostavljam kako bi podatci s otoka s velikim brojem opažanja iz istog razloga prikazali potpuniju logaritamski normalnu raspodjelu vrsta. Osim na geografskim otocima, asimptote je možda primjerenije tražiti u stanišnim i urbanim otocima za koje su ljudi više vezani, za koje također vrijede pravilnosti otočne biogeografije (Duelli i Obrist 2003, Angold i sur. 2006, Matthews i sur. 2016).

Ako pristupimo ovim podacima kao pokazateljima interakcije ljudi s raznolikosti vrsta, moguća je usporedba različitih područja koja su od interesa ljudima kao što su: urbane sredine, zaštićena područja ili slabo istražena područja. Moguće je također izraziti analoge α , β i γ -raznolikosti (Whittaker 1960, 1972):

1. α – interakcija na specifičnom lokalitetu, staništu
2. β – interakcija po gradijentu: biogeografskih regija, staništa, zaštićenih područja ili baza podataka
3. γ – interakcija po skupu: cijelog svijeta, biogeografskih regija ili staništa.

U ovom diplomskom radu izrađen je analog γ -raznolikosti, tj. γ – interakcija za skup staništa cjelovitog istraživanog područja. Iako utjecaj taksonoma na svojstva podataka nije uključen u ovaj rad, moramo uzeti u obzir njihov potencijalni utjecaj. Taksonomi imaju ulogu u stvaranju samih podataka koji su dostupni za preuzimanje, oni su zaduženi za identifikaciju organizama i time utječu na ukupni broj vrsta i brojnost jedinki koje se mogu naći u uzorku. Građanin znanstvenik može opaziti organizam, no to nužno ne znači kako će njegovo opažanje biti preuzeto u bazu GBIF. To također ovisi o primarnim izvorima informacija potrebnima za identifikaciju vrsta kao što su: uređene i kompletirane zbirke vrsta pojedinih država, dostupni

ključevi za identifikaciju vrsta i o samim florističkim, faunističkim, fungističkim itd. istraživanjima kojima je cilj popisati vrste koje se nalaze na nekom području. Pored toga, nove vrste se konstantno opisuju i procijenjeno je da je od ukupno oko 8,7 milijuna eukariotskih vrsta na svijetu opisano tek oko 1.2 milijuna (Mora i sur. 2011). Stoga se može pretpostaviti kako je interakcija građana znanstvenika s bioraznolikosti veća nego što je prikazana podacima u bazi GBIF. Isto treba prihvatiti činjenicu kako ovi podaci ne mogu biti reprezentativni uzorak cjelokupnog ljudskog opažanja, što proizlazi iz logičke pretpostavke u ovom radu: Svi građani znanstvenici su ljudi; svi ljudi nisu građani znanstvenici. Ovi podaci su reprezentativni samo za građane znanstvenike koji su međusobno povezani samo korištenjem aplikacija za unos podataka o bioraznolikosti. Rezultati su pokazali kako se korisnici mogu razaznati po broju opažanja u logaritamski normalnu raspodjelu što ukazuje na različite interakcije unutar same skupine građana znanstvenika, što je u skladu s radom August i sur. (2020). Ljudi također opažaju više organizama za svog života nego što je zabilježeno jer bilježenje opažanja je dodatna interakcija čovjeka s aplikacijom koja iziskuje određeni utrošak vremena na unos, te je i tu interakciju potrebno dodatno istražiti.

6. ZAKLJUČAK

1. Ljudskom opažanju vrsta i raspodjeli korisnika najbolje pristaje logaritamski normalan model rang-abundancije, dok povećavanjem taksonomske razine ono sve više odstupa zbog taksonomske pristranosti.
2. Ljudsko opažanje nije nasumično zbog taksonomske i vremenske pristranosti i prostorne grupiranosti.
3. Podatci prikupljeni ljudskim opažanjem imaju svrhu opisivanja interakcije građana znanstvenika s biološkom raznolikosti za područja gdje ljudi obitavaju i ne mogu se koristiti za određivanje raznolikosti vrsta.

7. LITERATURA

- About the OECD - OECD [mrežna stranica], 2021. Dostupna na: <https://www.oecd.org/about/> [Pristupljeno: 16. 8. 2021.].
- Aceves-Bueno, E., Adeleye, A.S., Feraud, M., Huang, Y., Tao, M., Yang, Y., and Anderson, S.E., 2017. The Accuracy of Citizen Science Data: A Quantitative Review. *The Bulletin of the Ecological Society of America*, 98 (4), 278–290.
- Angold, P.G., Sadler, J.P., Hill, M.O., Pullin, A., Rushton, S., Austin, K., Small, E., Wood, B., Wadsworth, R., Sanderson, R., and Thompson, K., 2006. Biodiversity in urban habitat patches. *Science of The Total Environment*, 360 (1–3), 196–204.
- August, T., Fox, R., Roy, D.B., and Pocock, M.J.O., 2020. Data-derived metrics describing the behaviour of field-based citizen scientists provide insights for project design and modelling bias. *Scientific Reports*, 10 (1), 11009.
- Baddeley, A., Rubak, E., and Turner, R., 2016. *Spatial point patterns: methodology and applications with R*. Boca Raton ; London ; New York: CRC Press, Taylor & Francis Group.
- Bar-On, Y.M., Phillips, R., and Milo, R., 2018. The biomass distribution on Earth. *Proceedings of the National Academy of Sciences*, 115 (25), 6506–6511.
- Beck, J., Böller, M., Erhardt, A., and Schwanghart, W., 2014. Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecological Informatics*, 19, 10–15.
- Bird, T.J., Bates, A.E., Lefcheck, J.S., Hill, N.A., Thomson, R.J., Edgar, G.J., Stuart-Smith, R.D., Wotherspoon, S., Krkosek, M., Stuart-Smith, J.F., Pecl, G.T., Barrett, N., and Frusher, S., 2014a. Statistical solutions for error and bias in global citizen science datasets. *Biological Conservation*, 173, 144–154.
- Bird, T.J., Bates, A.E., Lefcheck, J.S., Hill, N.A., Thomson, R.J., Edgar, G.J., Stuart-Smith, R.D., Wotherspoon, S., Krkosek, M., Stuart-Smith, J.F., Pecl, G.T., Barrett, N., and Frusher, S., 2014b. Statistical solutions for error and bias in global citizen science datasets. *Biological Conservation*, 173, 144–154.
- Boakes, E.H., Gliozzo, G., Seymour, V., Harvey, M., Smith, C., Roy, D.B., and Haklay, M., 2016. Patterns of contribution to citizen science biodiversity projects increase understanding of volunteers' recording behaviour. *Scientific Reports*, 6 (1), 33051.

- Bohmann, K., Evans, A., Gilbert, M.T.P., Carvalho, G.R., Creer, S., Knapp, M., Yu, D.W., and de Bruyn, M., 2014. Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution*, 29 (6), 358–367.
- Bonnet, X., Shine, R., and Lourdais, O., 2002. Taxonomic chauvinism. *Trends in Ecology & Evolution*, 17 (1), 1–3.
- Bonney, R., Shirk, J.L., Phillips, T.B., Wiggins, A., Ballard, H.L., Miller-Rushing, A.J., and Parrish, J.K., 2014. Next Steps for Citizen Science. *Science*, 343 (6178), 1436–1437.
- Boria, R.A., Olson, L.E., Goodman, S.M., and Anderson, R.P., 2014. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecological Modelling*, 275, 73–77.
- Bossu, R., Roussel, F., Fallou, L., Landès, M., Steed, R., Mazet-Roux, G., Dupont, A., Frobert, L., and Petersen, L., 2018. LastQuake: From rapid information to global seismic risk reduction. *International Journal of Disaster Risk Reduction*, 28, 32–42.
- Bowker, G.C., 2000. Biodiversity Datadiversity. *Social Studies of Science*, 30 (5), 643–683.
- Cavalier-Smith, T., 1981. Eukaryote kingdoms: Seven or nine? *Biosystems*, 14 (3), 461–481.
- Clark, J.A., 2002. Taxonomic Bias in Conservation Research. *Science*, 297 (5579), 191b–1192.
- Constitution - TDWG [mrežna stranica], 2021. Available from: <https://www.tdwg.org/about/constitution/> [Pristupljeno: 16. 08. 2021.].
- Core terms defined by Darwin Core [mrežna stranica], 2021. Dostupna na: <http://rs.tdwg.org/dwc/terms.htm> [Pristupljeno: 14. 8. 2021.].
- Darwin Core - TDWG [mrežna stranica], 2021. Dostupna na: <http://www.tdwg.org/standards/450> [Pristupljeno: 16. 08. 2021.].
- Data standards [mrežna stranica], 2021. Dostupna na: <https://www.gbif.org/standards> [Pristupljeno: 16. 08. 2021.].
- DCMI: About DCMI [mrežna stranica], 2021. Dostupna na: <https://dublincore.org/about/> [Pristupljeno: 14. 08. 2021.].
- Di Cecco, G.J., Barve, V., Belitz, M.W., Stucky, B.J., Guralnick, R.P., and Hurlbert, A.H., 2021. Observing the Observers: How Participants Contribute Data to iNaturalist and Implications for Biodiversity Science. *BioScience*, 71 (11), 1179–1188.
- Donaldson, M.R., Burnett, N.J., Braun, D.C., Suski, C.D., Hinch, S.G., Cooke, S.J., and Kerr, J.T., 2017. Taxonomic bias and international biodiversity conservation research. *FACETS*, 1 (1), 105–113.
- Duelli, P. and Obrist, M.K., 2003. Regional biodiversity in an agricultural landscape: the contribution of seminatural habitat islands. *Basic and Applied Ecology*, 4 (2), 129–138.

- Emlen, J.T., 1977. Estimating Breeding Season Bird Densities from Transect Counts. *The Auk*, 94 (3), 455–468.
- Fanelli, D., Costas, R., and Ioannidis, J.P.A., 2017. Meta-assessment of bias in science. *Proceedings of the National Academy of Sciences*, 114 (14), 3714–3719.
- Fisher, R.A., Corbet, A.S., and Williams, C.B., 1943. The Relation Between the Number of Species and the Number of Individuals in a Random Sample of an Animal Population. *The Journal of Animal Ecology*, 12 (1), 42.
- Gascon, C., Williamson, G.B., and da Fonseca, G.A.B., 2000. Receding Forest Edges and Vanishing Reserves. *Science*, 288 (5470), 1356–1358.
- GBIF [mrežna stranica], 2021. Dostupna na: <https://www.gbif.org/> [Pristupljeno: 16. 08. 2021.].
- Geldmann, J., Heilmann-Clausen, J., Holm, T.E., Levinsky, I., Markussen, B., Olsen, K., Rahbek, C., and Tøttrup, A.P., 2016. What determines spatial bias in citizen science? Exploring four recording schemes with different proficiency requirements. *Diversity and Distributions*, 22 (11), 1139–1149.
- Good, I.J., 1953. The Population Frequencies of Species and the Estimation of Population Parameters. *Biometrika*, 40 (3/4), 237.
- Gotelli, N.J. and Colwell, R.K., 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, 4 (4), 379–391.
- Haase, P., 1995. Spatial pattern analysis in ecology based on Ripley's K-function: Introduction and methods of edge correction. *Journal of Vegetation Science*, 6 (4), 575–582.
- Hankin, D.G., Mohr, M.S., and Newman, K.B., 2019. *Sampling theory for the ecological and natural resource sciences*. First edition. Oxford ; New York, NY: Oxford University Press.
- Hurlbert, S.H., 1971. The Nonconcept of Species Diversity: A Critique and Alternative Parameters. *Ecology*, 52 (4), 577–586.
- Kosmala, M., Wiggins, A., Swanson, A., and Simmons, B., 2016. Assessing data quality in citizen science. *Frontiers in Ecology and the Environment*, 14 (10), 551–560.
- Krebs, C.J., 1989. *Ecological methodology*. New York: Harper & Row.
- Kreyszig, E., 1979. *Advanced engineering mathematics*. 4th ed. New York: Wiley.
- Laurance, W.F., Nascimento, H.E.M., Laurance, S.G., Andrade, A., Ewers, R.M., Harms, K.E., Luizão, R.C.C., and Ribeiro, J.E., 2007. Habitat Fragmentation, Variable Edge Effects, and the Landscape-Divergence Hypothesis. *PLOS ONE*, 2 (10), e1017.
- Lloyd, M. and Ghelardi, R.J., 1964. A Table for Calculating the 'Equitability' Component of Species Diversity. *The Journal of Animal Ecology*, 33 (2), 217.

- Luo, M., Xu, Z., Hirsch, T., Aung, T.S., Xu, W., Ji, L., Qin, H., and Ma, K., 2021. The use of Global Biodiversity Information Facility (GBIF)-mediated data in publications written in Chinese. *Global Ecology and Conservation*, 25, e01406.
- MacArthur, R.H. and Wilson, E.O., 1967. *The theory of island biogeography*. Princeton: Princeton University Press.
- Matthews, T.J., Guilhaumon, F., Triantis, K.A., Borregaard, M.K., Whittaker, R.J., and Santos, A., 2016. On the form of species–area relationships in habitat islands and true islands. *Global Ecology and Biogeography*, 25 (7), 847–858.
- McIntosh, R.P., 1967. An Index of Diversity and the Relation of Certain Concepts to Diversity. *Ecology*, 48 (3), 392–404.
- Mora, C., Tittensor, D.P., Adl, S., Simpson, A.G.B., and Worm, B., 2011. How Many Species Are There on Earth and in the Ocean? *PLOS Biology*, 9 (8), e1001127.
- NCBI, 2021. Our Mission - NCBI [mrežna stranica]. Dostupna na: <https://www.ncbi.nlm.nih.gov/home/about/mission/> [Pristupljeno: 10. 08. 2021.].
- Nekola, J.C., Hutchins, B.T., Schofield, A., Najev, B., and Perez, K.E., 2019. *Caveat consumptor notitia museo*: Let the museum data user beware. *Global Ecology and Biogeography*, 28 (12), 1722–1734.
- Nyffeler, M. and Birkhofer, K., 2017. An estimated 400–800 million tons of prey are annually killed by the global spider community. *The Science of Nature*, 104 (3–4), 30.
- Gbif.Org, 2022. Occurrence Download. <https://doi.org/10.15468/dl.z35efj> [Preuzeto: 23. 06. 2022.].
- OECD Megascience Forum, 1999. Final report of the working group on biological informatics. Organisation for Economic Co-operation and Development.
- Oksanen, J., Simpson, G.L., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O’Hara, R.B., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., Caceres, M.D., Durand, S., Evangelista, H.B.A., FitzJohn, R., Friendly, M., Furneaux, B., Hannigan, G., Hill, M.O., Lahti, L., McGlenn, D., Ouellette, M.-H., Cunha, E.R., Smith, T., Stier, A., Braak, C.J.F.T., and Weedon, J., 2022. vegan: Community Ecology Package.
- Pianka, E.R., 1970. On r- and K-Selection. *The American Naturalist*, 104 (940), 592–597.
- Pielou, E.C., 1966. The measurement of diversity in different types of biological collections. *Journal of Theoretical Biology*, 13, 131–144.
- Preston, F.W., 1948. The Commonness, And Rarity, of Species. *Ecology*, 29 (3), 254–283.

- Quitoriano, V. and Wald, D.J., 2020. USGS “Did You Feel It?”—Science and Lessons From 20 Years of Citizen Science-Based Macroseismology. *Frontiers in Earth Science*, 8.
- Record [mrežna stranica], 2022. Dostupna na: <https://dictionary.cambridge.org/dictionary/english/record> [Pristupljeno: 01. 09. 2022.].
- Ripley, B.D., 1976. The second-order analysis of stationary point processes. *Journal of Applied Probability*, 13 (2), 255–266.
- Ripley, B.D., 1977. Modelling Spatial Patterns. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39 (2), 172–192.
- Rocha-Ortega, M., Rodriguez, P., and Córdoba-Aguilar, A., 2021. Geographical, temporal and taxonomic biases in insect GBIF data on biodiversity and extinction. *Ecological Entomology*, 46 (4), 718–728.
- Sæther, B.-E., Visser, M.E., Grøtan, V., and Engen, S., 2016. Evidence for r - and K -selection in a wild bird population: a reciprocal link between ecology and evolution. *Proceedings of the Royal Society B: Biological Sciences*, 283 (1829), 20152411.
- Search [mrežna stranica], 2022a. Available from: <https://www.gbif.org/occurrence/search> [Pristupljeno: 20 May 2022].
- Search [mrežna stranica], 2022b. Dostupna na: https://www.gbif.org/occurrence/taxonomy?basis_of_record=HUMAN_OBSERVATION&occurrence_status=present [Pristupljeno: 10. 05. 2022.].
- Simberloff, D., 1972. Properties of the Rarefaction Diversity Measurement. *The American Naturalist*, 106 (949), 414–418.
- Simberloff, D.S. and Wilson, E.O., 1970. Experimental Zoogeography of Islands. A Two-Year Record of Colonization. *Ecology*, 51 (5), 934–937.
- Smith, E.P. and van Belle, G., 1984. Nonparametric Estimation of Species Richness. *Biometrics*, 40 (1), 119.
- Specht, H. and Lewandowski, E., 2018. Biased Assumptions and Oversimplifications in Evaluations of Citizen Science Data Quality. *The Bulletin of the Ecological Society of America*, 99 (2), 251–256.
- Stearns, S.C., 1977. The Evolution of Life History Traits: A Critique of the Theory and a Review of the Data. *Annual Review of Ecology and Systematics*, 8 (1), 145–171.
- Strien, A.J. van, Swaay, C.A.M. van, and Termaat, T., 2013. Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. *Journal of Applied Ecology*, 50 (6), 1450–1458.

- Telenius, A., 2011. Biodiversity information goes public: GBIF at your service. *Nordic Journal of Botany*, 29 (3), 378–381.
- Titley, M.A., Snaddon, J.L., and Turner, E.C., 2017. Scientific research on animal biodiversity is systematically biased towards vertebrates and temperate regions. *PLOS ONE*, 12 (12), e0189577.
- Tobler, W.R., 1970. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46 (sup1), 234–240.
- TrouDET, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R., and Legendre, F., 2017. Taxonomic bias in biodiversity data and societal preferences. *Scientific Reports*, 7 (1), 9132.
- Vohland, K., Land-Zandstra, A., Ceccaroni, L., Lemmens, R., Perelló, J., Ponti, M., Samson, R., and Wagenknecht, K., eds., 2021. *The Science of Citizen Science*. Cham: Springer International Publishing.
- Waller, J., Volik, N., Mendez, F., and Hahn, A., 2021. GBIF Data Processing and Validation. *Biodiversity Information Science and Standards*, 5, e75686.
- What is GBIF? [mrežna stranica], 2021. Dostupna na: <https://www.gbif.org/what-is-gbif> [Pristupljeno: 14. 08. 2021.].
- Whittaker, R.H., 1960. Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs*, 30 (3), 279–338.
- Whittaker, R.H., 1965. Dominance and Diversity in Land Plant Communities. *Science*, 147 (3655), 250–260.
- Whittaker, R.H., 1972. Evolution and Measurement of Species Diversity. *TAXON*, 21 (2–3), 213–251.
- Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T., and Vieglais, D., 2012. Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLoS ONE*, 7 (1), e29715.
- Wilson, J.B., 1991. Methods for fitting dominance/diversity curves. *Journal of Vegetation Science*, 2 (1), 35–46.

9. PRILOG

Sadržaj:

1. UV svjetlosna piramida.....	i
Prilog 1.	i
Prilog 2.	ii
2. Rang-abundancijski modeli za istraživane taksonomske razine.....	iii
Prilog 3.	iii
Prilog 4.	iv
Prilog 5.	v
Prilog 6.	vi
3. Rang-abundancijski modeli za istraživane taksonomske razine bez atributa problem.....	vii
Prilog 7.	viii
Prilog 8.	ix
Prilog 9.	x
Prilog 10.	xi
Prilog 11.	xii

1. UV svjetlosna piramida



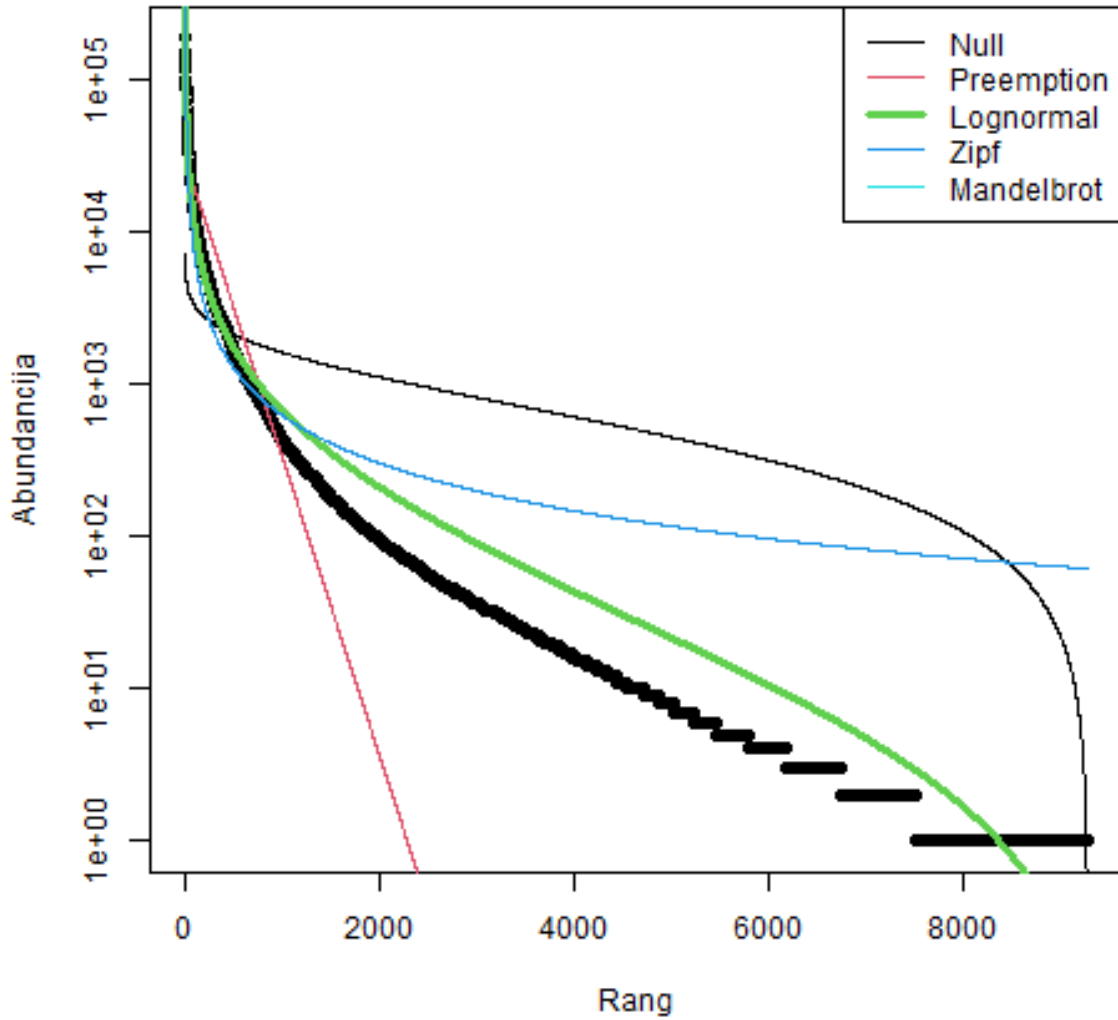
Prilog 1. UV svjetlosna piramida.



Prilog 2. Fotografija s produljenom ekspozicijom UV svjetlosne piramide. Produljena ekspozicija otkriva prostor koji obasjava UV svjetlost.

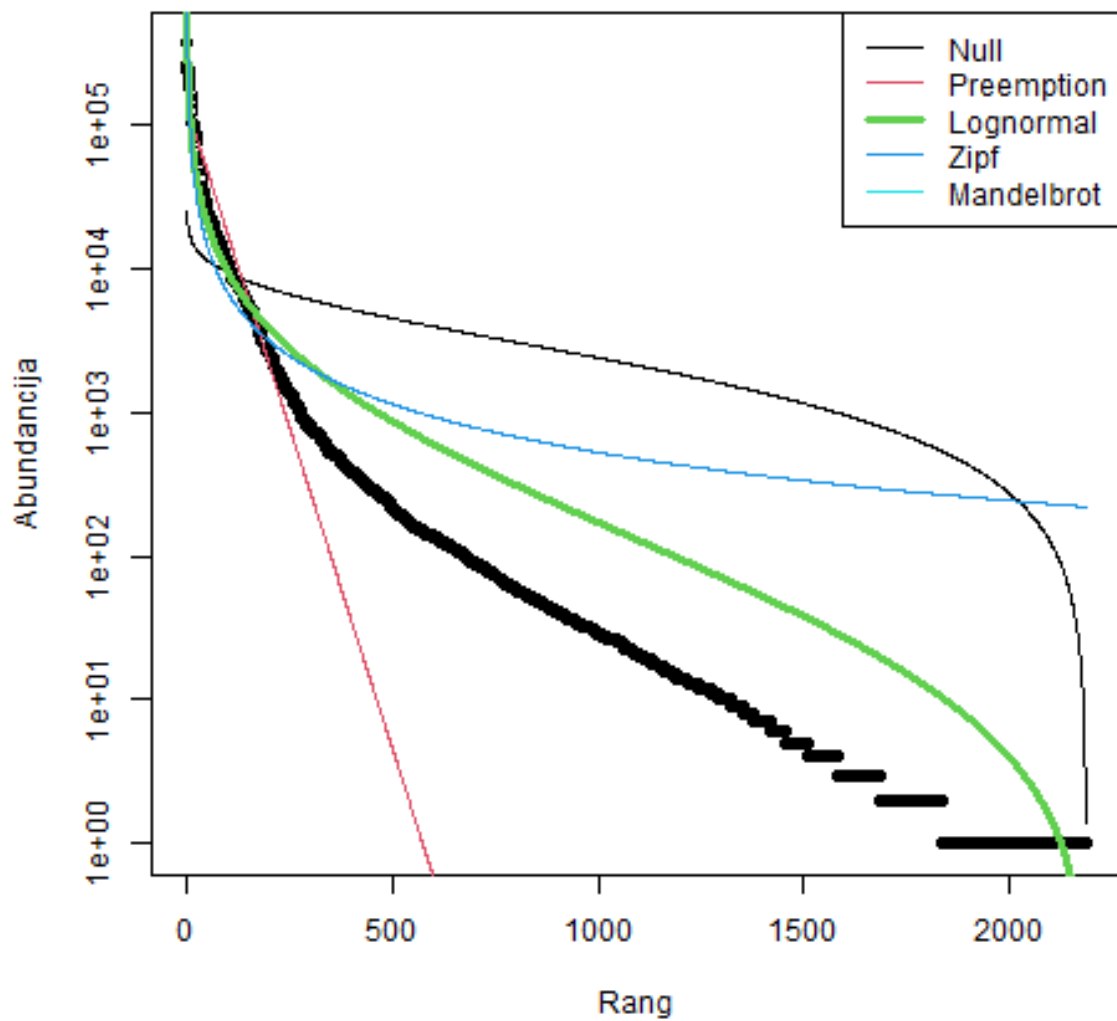
2. Rang-abundancijski modeli za istraživane taksonomske razine

Rang-abundancijski modeli za taksonomsku kategoriju roda



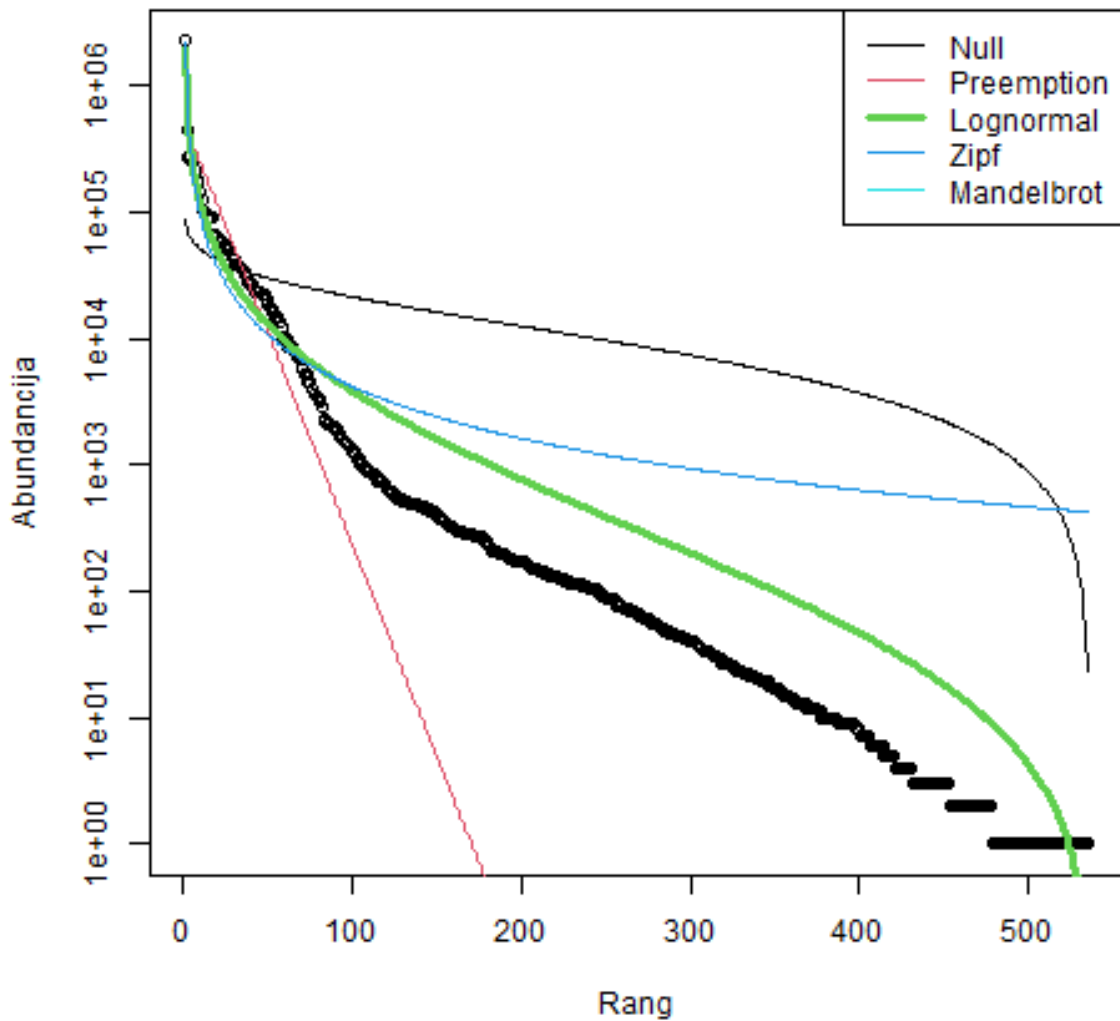
Prilog 3. Rang-abundancijski modeli za taksonomsku kategoriju roda. Podebljane crne linije označavaju raspodjelu podataka. Za najbolje pristajanje je izabran logaritamski normalan model.

Rang-abundancijski modeli za taksonomsku kategoriju porodice



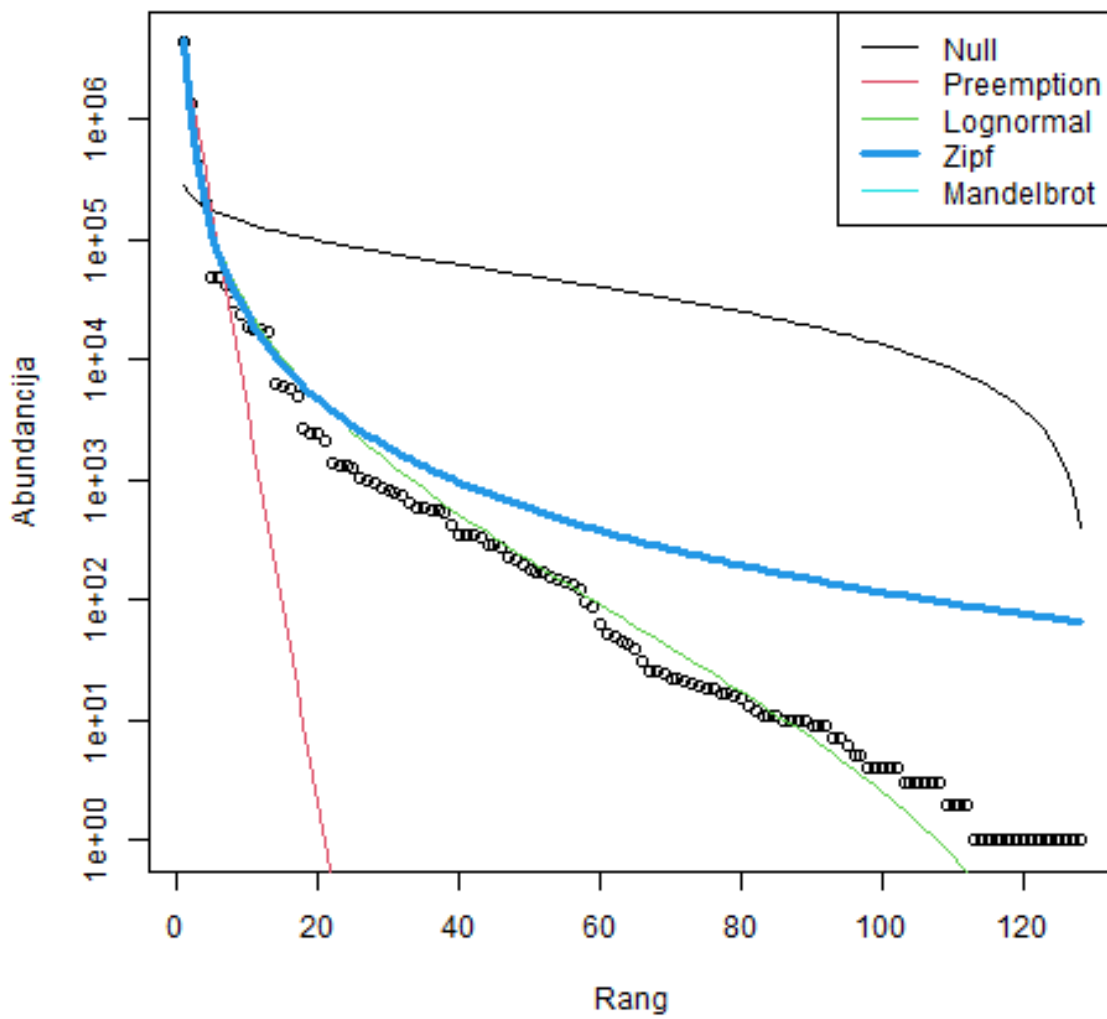
Prilog 4. Rang-abundancijski modeli za taksonomsku kategoriju porodice. Podebljane crne linije označavaju raspodjelu podataka. Za najbolje pristajanje je izabran logaritamski normalan model.

Rang-abundancijski modeli za taksonomsku kategoriju reda



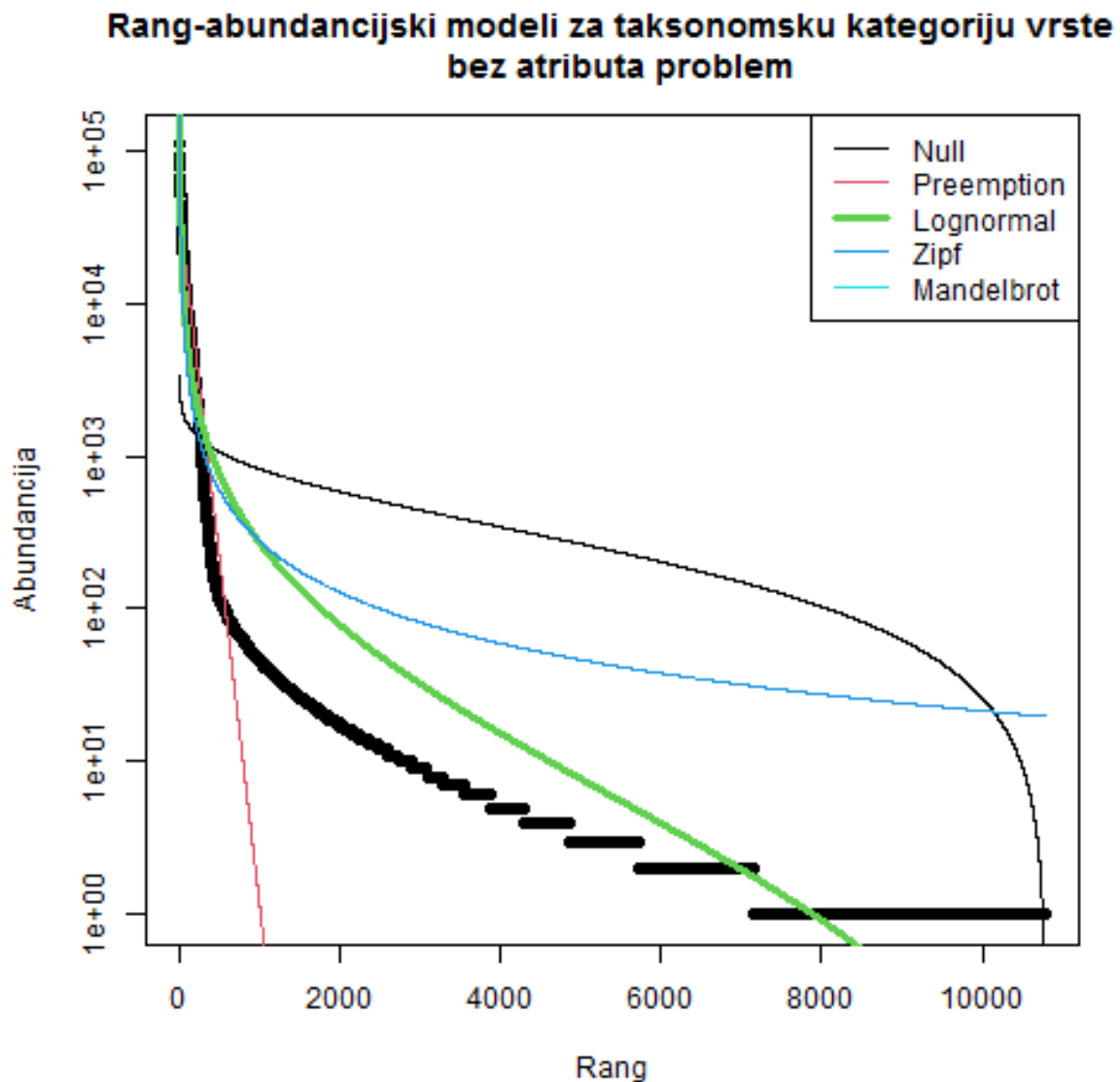
Prilog 5. Rang-abundancijski modeli za taksonomsku kategoriju reda. Podebljane crne linije označavaju raspodjelu podataka. Za najbolje pristajanje je izabran logaritamski normalan model.

Rang-abundancijski modeli za taksonomsku kategoriju razreda



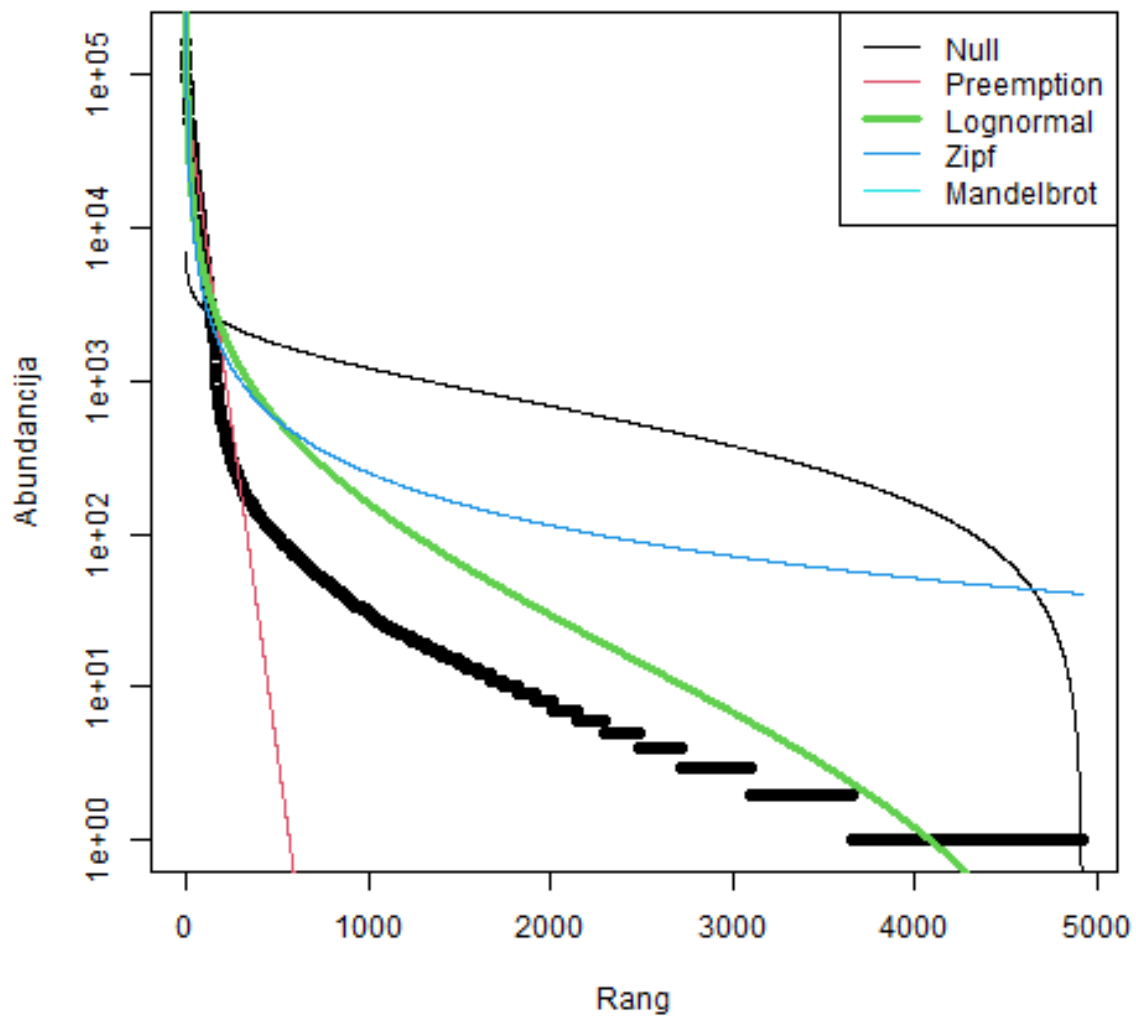
Prilog 6. Rang-abundancijski modeli za taksonomsku kategoriju razreda. Crni krugovi označavaju raspodjelu podataka. Za najbolje pristajanje je izabran Zipf model.

3. Rang-abundacijski modeli za istraživane taksonomske razine bez atributa problem



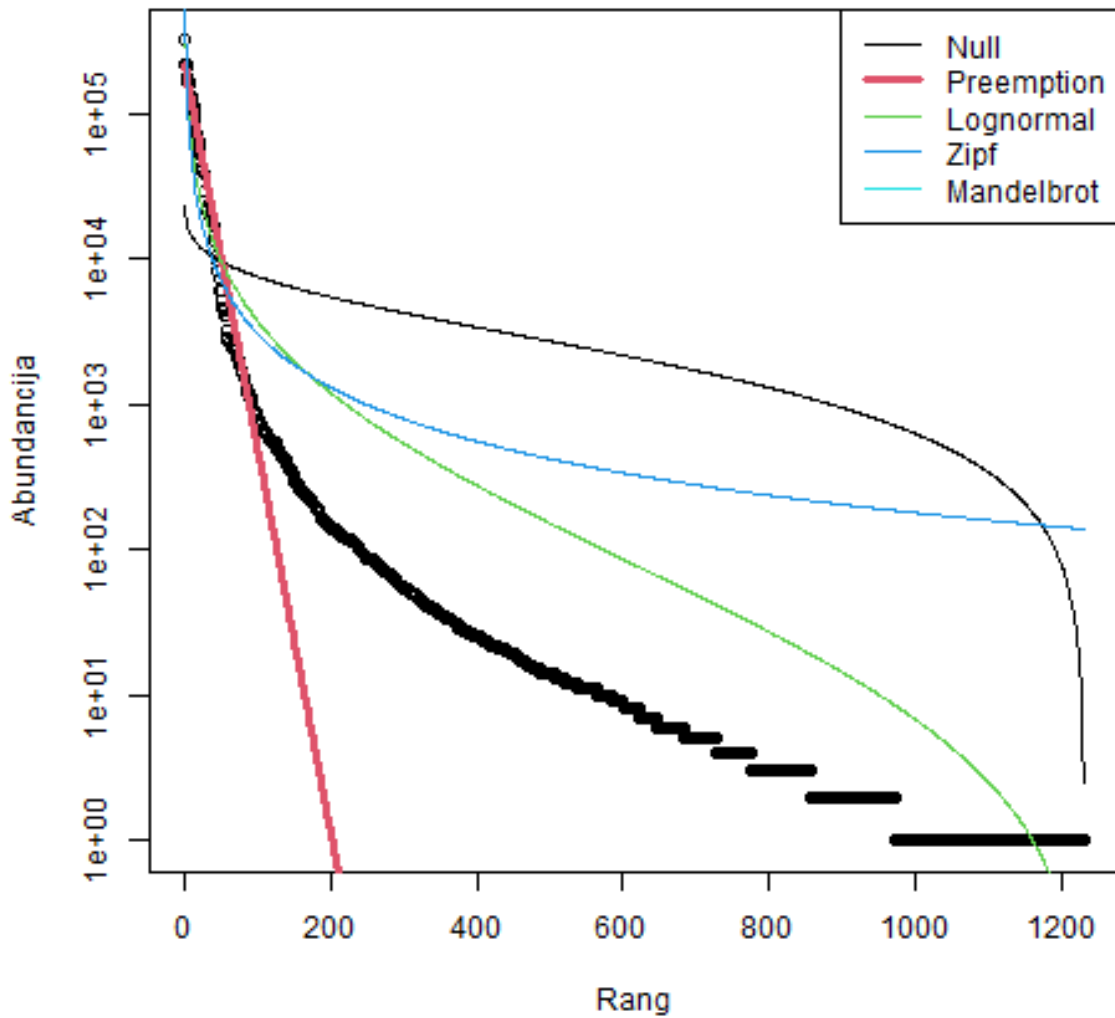
Prilog 7. Rang-abundancijski modeli za taksonomsku kategoriju vrste bez atributa problem. Podebljane crne linije označavaju raspodjelu podataka. Za najbolje pristajanje je izabran logaritamski normalan model.

Rang-abundancijski modeli za taksonomsku kategoriju roda bez atributa problem



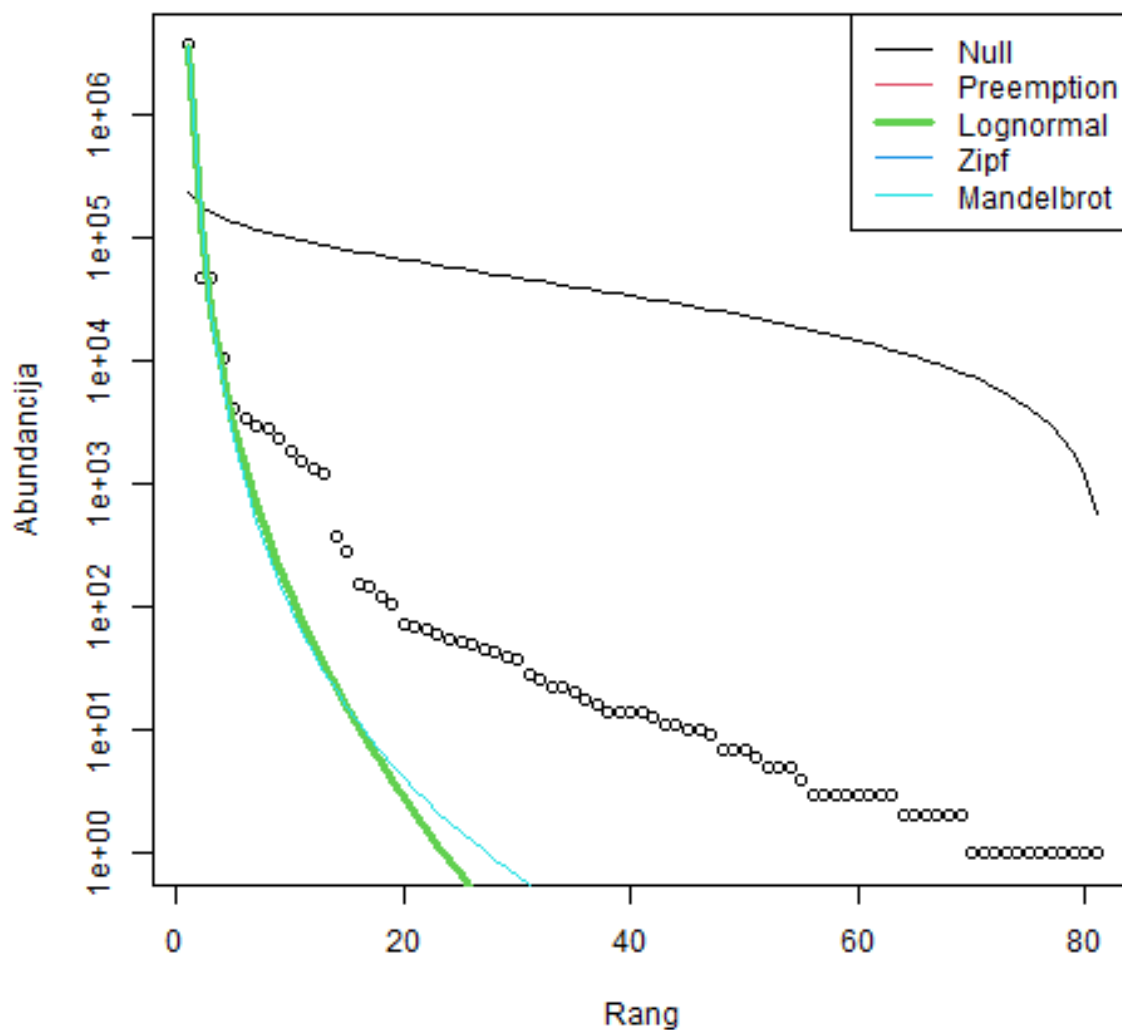
Prilog 8. Rang-abundancijski modeli za taksonomsku kategoriju roda bez atributa problem. Podebljane crne linije označavaju raspodjelu podataka. Za najbolje pristajanje je izabran logaritamski normalan model.

Rang-abundancijski modeli za taksonomsku kategoriju porodice bez atributa problem



Prilog 9. Rang-abundancijski modeli za taksonomsku kategoriju porodice bez atributa problem. Podebljane crne linije označavaju raspodjelu podataka. Za najbolje pristajanje je izabran Preemption model.

Rang-abundancijski modeli za taksonomsku kategoriju razreda bez atributa problem



Prilog 10. Rang-abundancijski modeli za taksonomsku kategoriju razreda bez atributa problem. Crni krugovi označavaju raspodjelu podataka. Za najbolje pristajanje je izabran logaritamski normalan model.

8. ŽIVOTOPIS

Rođen sam 31. 05. 1993. godine u Zadru. Od 2008. do 2012. godine. pohađao sam Opću gimnaziju Juraj Baraković u Zadru. Od 2012 do 2018. godine pohađao sam preddiplomski studij Znanosti o okolišu na Biološkom odsjeku Prirodoslovno-matematičkog fakulteta u Zagrebu, te stekao zvanje prvostupnika struke znanosti o okolišu (*univ. bacc. oecol.*). Od 2018. do danas pohađao sam diplomski studij Znanosti o okolišu na Biološkom odsjeku Prirodoslovno-matematičkog fakulteta u Zagrebu. Kao studentski posao držao sam demonstrature od 2013. do 2017. na kolegijima: Protista, Opća zoologija, Zoologija, Osnove biologije, Kralježnjaci i Histologija i embriologija na Biološkom odsjeku Prirodoslovno-matematičkog fakulteta u Zagrebu. Ukupno sam održao 495 sati demonstracija. Također sam od 2015. do 2018. i 2021. do 2022. godine bio edukator na pripremama srednjoškolaca i osnovnoškolaca za Prirodoslovnu olimpijadu Europske Unije EUSO i Međunarodnu prirodoslovnu olimpijadu mladih –IJSO. Van nastavno iskustvo sam steko: 2015. godine kao član tehničke podrške 17. Europskog karabidološkog kongresa u organizaciji Hrvatskog ekološkog društva; od 2015. do 2016. kao član organizacije Simpozija studenata bioloških usmjerenja (SiSB) Prirodoslovno-matematičkog fakulteta u Zagrebu; 2016. godine kao član tehničke podrške 2. Hrvatskog simpozija o invazivnim vrstama u organizaciji Hrvatskog ekološkog društva; od 2014. do 2017. godine kao voditelj tima Kukci na manifestaciji Noć biologije Prirodoslovno-matematičkog fakulteta u Zagrebu; od 2015. do 2017. godine kao studentski predstavnik u Vijeću Biološkog odsjeka Prirodoslovno-matematičkog fakulteta u Zagrebu; od 2016. do 2017. godine kao studentski predstavnik na Kolegiju Biološkog odsjeka Prirodoslovno-matematičkog fakulteta u Zagrebu. Sudjelovao sam na istraživačkim projektima: kao član Sekcije za pauke na istraživačko-edukacijskom projektu i terenu Udruge studenata biologije-BIUS „Papuk 2015“, Papuk i od 2016. do 2019. godine kao voditelj sekcije za Anatomiju, morfologiju i prepariranje na istraživačko-edukacijskim projektima i terenima Udruge studenata biologije-BIUS: Mura-Drava 2016, Insula Tilagus 2017, Šuma Žutica 2018, Zlarin 2019. Sudjelovao sam na skupovima sa: Šerić Jelaska, L., Šapina, I. i Franjević, M. (2017) Field predation of winter moth defoliator by soil arthropods revealed using PCR. Ecological Networks and Molecular Analysis of Trophic Interactions 2017 (sažetak), Uppsala, Švedska. (sažetak); Šapina, I. i Šerić Jelaska, L. (2018) The brown marmorated stink bug *Halyomorpha halys* (Stål, 1855) continues to invade Croatia. Knjiga sažetaka 3. simpozij o invazivnim vrstama (sažetak i poster); Skejo, J. i Šapina, I. (2017) Pygmy Blackhoppers from Wallacea (Orthoptera: Tetrigidae): *Hirrius* from Mindanao and *Zvierckia* gen. nov. from Sulawesi. Knjiga sažetaka Drugog Simpozija studenata bioloških usmjerenja-SiSB (sažetak) i Šapina, I. i Šerić Jelaska, L. (2017) First Record of the Invasive Brown Marmorated Stink Bug *Halyomorpha halys* (Stål, 1855) in Croatia. Knjiga sažetaka Drugog Simpozija studenata bioloških usmjerenja-SiSB (sažetak). Objavio sam znanstveni rad: Šapina, Ivan i Šerić Jelaska, Lucija. (2018). First report of invasive brown marmorated stink bug *Halyomorpha halys* (Stal, 1855) in Croatia. Bulletin OEPP/EPP Bulletin 48(1), 138-143. Objavio sam studentski terenski izvještaj: Majnarić Stefan i Šapina Ivan. (2017). Istraživanje Ravnokrilaca (Insecta: Orthoptera) Dugog Otoka s naglaskom na usporedbu zajednica suhих i vlažni staništa. Zbornik radova Istraživačko - edukacijskog projekta „Insula Tilagus 2017. “Udruge studenata biologije-BIUS“. Objavio sam studentski rad: Tenšek, Anđela i Šapina, Ivan. (2019). Istraživanje pojave invazivne vrste *Ailanthus altissima* (Mill. Swingle) na Arheološkoj zoni Vučedol. Fragmenti 7, 87-101.