

UNIVERZA V LJUBLJANI  
BIOTEHNIŠKA FAKULTETA

Roman LUŠTRIK

MODELI S PROSTORSKO OMEJITVIJO ZA OCENJEVANJE  
GOSTOT SUPERPOPULACIJ

DOKTORSKA DISERTACIJA

Šentvid nad Ljubljano, 2018

UNIVERZA V LJUBLJANI  
BIOTEHNIŠKA FAKULTETA

Roman LUŠTRIK

MODELI S PROSTORSKO OMEJITVIJO ZA OCENJEVANJE  
GOSTOT SUPERPOPULACIJ

DOKTORSKA DISERTACIJA

SPATIALLY EXPLICIT MODELING OF SUPERPOPULATION  
DENSITY

DOCTORAL DISSERTATION

Šentvid nad Ljubljano, 2018

Na podlagi Statuta Univerze v Ljubljani ter po sklepu Senata Biotehniške fakultete in sklepa Komisije za doktorski študij Univerze v Ljubljani z dne XX. XX. XXXX je bilo potrjeno, da kandidat izpolnjuje pogoje za opravljanje doktorata znanosti na Interdisciplinarnem doktorskem študijskem programu Statistika. Za mentorja je bil imenovan doc. dr. Tomaž Skrbinšek.

Komisija za oceno in zagovor:

Predsednik: prof. dr. Peter TRONTELJ  
Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za  
biologijo

Član: prof. dr. Vladimir BATAGELJ  
Fakulteta za matematiko in fiziko, recenzent

Delo je rezultat lastnega raziskovalnega dela.

Podpisani se strinjam z objavo svoje naloge v polnem tekstu na spletni strani Digitalne knjižnice Biotehniške fakultete. Izjavljam, da je naloga, ki sem jo oddal v elektronski obliki, identična tiskani verziji.

Datum zagovora:

Roman LUŠTRIK

KLJUČNA DOKUMENTACIJSKA INFORMACIJA

ŠD	Dd
DK	
KG	gostota, ocena velikosti populacije, učinek roba, metoda ulova in ponovnega ulova
AV	LUŠTRIK, Roman
SA	SKRBINŠEK, Tomaž (mentor)
KZ	SI-1000 Ljubljana, Jamnikarjeva 101
ZA	Univerza v Ljubljani, Biotehniška fakulteta, Interdisciplinarni doktorski študijski program Statistika
LI	2018
IN	MODELI S PROSTORSKO OMEJITVIJO ZA OCENJEVANJE GOSTOT SUPERPOPULACIJ
TD	Doktorska disertacija
OP	X, 46 str., 1 pregl., 10 sl., 1 pril., 45 vir.
IJ	sl
JI	sl/en
AI	<izvleček do približno 200 besed>

**KEYWORD DOCUMENTATION**

DN Dd  
DC  
CX density, population size estimate, edge effect, mark-recapture  
AU LUŠTRIK, Roman  
AA SKRBINŠEK, Tomaž (mentor)  
PP SI-1000 Ljubljana, Jamnikarjeva 101  
PB University of Ljubljana, Biotechnical faculty, Department of  
biology  
PY 2018  
TI SPATIALLY EXPLICIT MODELING OF SUPERPOPULATION  
DENSITY  
DT Doctoral dissertation  
NO XIII, 37 p., 6 tab., 13 fig., 7 ann., 44 ref.  
LA sl  
AL sl/en  
AB <abstract>

## KAZALO VSEBINE

<b>1 Uvod</b>	<b>1</b>
1.1 Pomen ocene velikosti populacije kot biološki parameter . . . . .	1
1.2 Načini ocenjevanja številčnosti . . . . .	1
1.3 Metoda lova in ponovnega ulova . . . . .	2
1.3.1 Osnovne značilnosti metode lova in ponovnega ulova . . . . .	2
1.3.2 Prepoznavanje osebkov . . . . .	3
1.3.3 Območje vzorčenja . . . . .	3
1.3.3.1 Domač okoliš . . . . .	4
1.3.3.2 Superpopulacija . . . . .	4
1.3.4 Predpostavke metod ulova in ponovnega ulova . . . . .	4
1.3.5 Zanesljivost označevanja . . . . .	4
1.3.6 Enaka ulovljivost . . . . .	5
1.3.7 Zaprtost populacije . . . . .	5
1.4 Cenilka velikosti populacije za dva odlovnna intervala . . . . .	5

---

1.5 Hugginsov model za zaprte populacije . . . . .	8
1.5.1 Model CAPWIRE . . . . .	9
1.6 Učinek roba . . . . .	10
1.6.1 Poskusi reševanja posledic učinka roba . . . . .	10
1.7 Naš pristop k reševanju posledic učinka roba . . . . .	11
1.7.1 Simulacije kot orodje za raziskovanje . . . . .	13
1.8 Cilji disertacije . . . . .	13
<b>2 Materiali in metode</b>	<b>15</b>
2.1 Simuliranje gibanja osebkov . . . . .	15
2.1.1 Odlov . . . . .	15
2.1.2 Izračun individualne spremenljivke . . . . .	16
2.1.2.1 Dvorazsežnostna normalna porazdelitev . . . . .	16
2.1.2.2 Empirična porazdelitev . . . . .	16
2.2 Ocenjevanje velikosti populacij . . . . .	17
2.2.1 Hugginsov model . . . . .	17

2.2.2 Po modelu CAPWIRE . . . . .	18
2.3 Statistike za primerjavo učinkovitosti delovanja popravka . . . . .	18
<b>3 Rezultati</b>	<b>20</b>
3.1 Simulacije . . . . .	20
3.2 Parameter p modelov $M_0$ in $M_{sp}$ po Hugginsu . . . . .	21
3.2.1 Primerjava simulirane in ocenjene ulovljivosti glede na porazdelitev	21
3.2.1.1 Empirična porazdelitev . . . . .	21
3.2.1.2 Normalna porazdelitev . . . . .	22
3.2.2 Primerjava ulovljivosti glede na število simuliranih osebkov . . .	23
3.2.3 Primerjava ulovljivosti glede na razmerje velikosti domačega okoliša in območja vzorčenja . . . . .	25
3.3 Gostota . . . . .	29
3.3.1 Razlike v gostoti glede na število simuliranih osebkov in funkcij za izračun individualne spremenljivke . . . . .	29
3.3.2 Razlike v gostoti glede na funkcijo za izračun individualne spremenljivke, števila odlovnih intervalov ter število simuliranih osebkov . . . . .	32
3.4 AICc . . . . .	36

<b>4 Razprava</b>	<b>39</b>
4.1 Pristranskost ocene ulovljivosti . . . . .	40
4.2 Primerjava AICc za modela $M_0$ in $M_{sp}$ . . . . .	41
4.3 Primerjava ocen gostot . . . . .	41
4.3.1 Računanje gostote s pomočjo popravljene velikosti območja vzorčenja . . . . .	42
<b>5 Sklepi</b>	<b>44</b>
<b>6 Povzetek</b>	<b>46</b>
<b>7 Viri in reference</b>	<b>47</b>

## KAZALO PREGLEDNIC

## KAZALO SLIK

1	Vzorčene točke in območje vzorčenja . . . . .	11
2	Funkcija simuliranja vzorčnih točk . . . . .	12
3	Primerjava ocenjene ulovljivosti s simulirano . . . . .	23
4	Primerjava ocenjene ulovljivosti s simulirano glede na število simuliranih osebkov . . . . .	24
5	Vpliv razmerja med ocenjeno in pravo ulovljivostjo v povezavi z razmerjem velikosti domačega okoliša in vzorčenega območja. . . . .	26
6	Prikaz pristranskosti ocene ulovljivosti ( $p$ ) glede na razmerje površin domačega okoliša in območja vzorčenja v povezavi s številom simuliranih osebkov (stolpci) in odlovnih presledkov (vrstice). . . . .	28
7	Prikaz kazalnika gostot ( $\hat{D}/D$ ) . . . . .	31
8	Prikaz vrednosti simulacij, kjer smo za izračun individualne spremenljivke uporabili empirično porazdelitev . . . . .	34
8	Prikaz vrednosti simulacij, kjer smo za izračun individualne spremenljivke uporabili normalno porazdelitev . . . . .	36
9	Prikaz razlike v AICc za modela $M_0$ in $M_{sp}$ . . . . .	37
10	Prikaz razlike v AICc za modela $M_0$ in $M_{sp}$ . . . . .	38

## 1 UVOD

Velikosti populacije je v ekologiji pomemben parameter, ki ga lahko preučujemo. Služi v pomoč pri raziskovanju drugih značilnosti populacije. Z nekaterimi populacijami živali aktivno upravljam in ocena številčnosti je lahko izhodiščni parameter za kvaliteten trajnostni odvzem osebkov iz populacije.

### 1.1 POMEN OCENE VELIKOSTI POPULACIJE KOT BIOLOŠKI PARAMETER

Populacijo lahko opredelimo kot skupino osebkov iste vrst, ki naseljujejo skupni prostor v danem času (Krebs, 2001). Kot taka je na voljo za proučevanje in upravljanje.

Proučevanje katero koli populacije se začne s poznavanjem njene biologije in ekologije. Posamično ali v skupinah proučujemo demografske parametre, ki dajo vpogled v populacijo in vrsto. Taki parametri so denimo življenjska doba, število potomcev, stopnja preživetja različnih starostnih skupin ipd. Poleg teh pa je pomembna tudi velikosti populacije, ki jo lahko imenujemo tudi številčnost oz. jo gledamo na znano velikem prostoru kot gostoto. Številčnost je eden najbolj intuitivnih parametrov, ki jih lahko pripišemo populaciji, saj predstavlja izhodišče za upravljanje. Poleg številčnosti so lahko drugi omejujoči dejavniki, denimo genetska slika. Primera takih dveh (sub) populacij lahko spremlijamo na primerih floridskega panterja (Pimm in sod., 2006) ali nenazadnje slovenskega risa.

### 1.2 NAČINI OCENJEVANJA ŠTEVILČNOSTI

Velikost populacije lahko ocenjujemo na dva načina. Kot število osebkov na območju vzorčenja ali neposredno kot gostoto (število osebkov na prostorsko enoto). V obeh primerih pa ostaja problem zaznavnosti posameznega osebka. Zaradi načina življenja ali pa velikega območja na katerem se populacija razteza, vseh živali fizično ni mogoče zaznati v izvedljivih časovnih, finančnih in drugih okvirih. Zato moramo poseči po pristopih, ki omogočajo, da zaznamo osebkom "prištejemo" tiste, ki smo jih spregledali. Na ta način lahko ocenimo vsoto zaznanega in pričakovanega dela populacije. S pomočjo znanstveno preverljivih metod lahko oceni podamo tudi interval zaupanja, s katerim lahko sporočimo tudi kako natančna je naša ocena.

Ko na določenem območju z gotovostjo zaznamo vse osebke, jih preprosto preštejemo. To imenujemo census, kjer predpostavimo, da vse osebke zagotovo zaznamo ( $p = 1$ ). Cenzusi v ekologiji niso pogosti, ker določen delež osebkov navadni ni mogoče zaznati. Metode, ki upoštevajo dejstvo, da vsi osebki med vzorčenjem niso zaznani s popolno gotovostjo ( $p < 1$ ), so osnovane na podlagi štetja na ploskvah, oddaljenosti od opazovalca in lova in ponovnega ulova (Williams in sod., 2002).

Metoda štetja na ploskvah deluje na način, da območje, kjer želimo oceniti številčnost neke populacije, razdelimo na prostorske podenote. Na vzorcu podenot izvedemo vzorčenje in iz ocen številčnosti v podenotah lahko sklepamo o številčnosti na večjem območju. Pri tem moramo upoštevati več predpostavk, saj ima sklepanje iz vzorca na večje območje več pasti.

Metode, kjer se beleži oddaljenost osebka od točke opazovanja, v prvi vrsti podajo oceno gostote. Opazovanje lahko poteka na naključno izbranih točkah ali transektilih, vhodni podatki za računanje pa je razdalja osebka do transektov. Predpostavlja se, da je na transektilih, ki jih popisovalec pregleduje, zaznavnost 1, z oddaljenosti od transektov pa zaznavnost pada po krivulji. Tip krivulje lahko izberemo na podlagi teorije ali prakse.

Do sedaj opisani načini za oceno številčnosti niso upoštevali identitete zaznanega osebka, kar pa je domena metod ulova in ponovnega ulova.

### 1.3 METODA LOVA IN PONOVNEGA ULOVA

#### 1.3.1 Osnovne značilnosti metode lova in ponovnega ulova

Metode ulova in ponovnega ulova se od zgoraj naštetih metod razlikujejo po tem, da za informacijo uporabljajo identiteto osebka. Zanesljiva določitev identitete osebka je najvišja kvaliteta podatka, ki ga lahko pridobimo za oceno številčnosti populacije. Kljub temu, da so bile nekatere matematične osnove metode lova in ponovnega ulova razvite že pred več stoletji, se je paleta modelov razširila in prešla v splošno uporabo v začetku dvajsetega stoletja (Pollock, 2000).

Poznamo več matematičnih modelov in vsak ima svoje značilnosti. V eni od delitev lahko modele razvrstimo na dve večji skupini, ki se razlikujeta po tem, da so primerne

za odprte oz. zaprte populacije. O odprtih populacijah govorimo takrat, ko osebki med vzorčenjem prihajajo in odhajajo iz populacije. Zaprte populacije tega prehoda ne predpostavljajo in je zato že v osnovi pojem "velikost populacije" lažje opredeliti. Pregled metod za zaprte populacije v svojem klasičnem delu opišejo Otis in sod. (1978), kasneje pa še drugi (tudi za odprte), kot so White in sod. (1982), Seber (1986) in White in Burnham (1999).

### 1.3.2 Prepoznavanje osebkov

Da lahko osebek prepoznamo kot edinstvenega, ga moramo na nekakšen način ujeti oz. zaznati. V preteklosti je to predstavljalo fizičen ulov, nakar so žival nedvoumno označili. Osebke se označuje na načine, da je oznaka trajna ali začasna (npr. označevanja robnih lusk ščita pri želvah (Pike in sod., 2005), ščipanje ušes ali nohtov pri manjših sesalcih (Wiewel in sod., 2007), pisanje številk na krila metuljev (Jugovic in sod., 2017), ščipanje prstov pri dvoživkah (Campbell in sod., 2009), opremljanje z ušesnimi značkami, barvanje kožuha, tetoviranje idr.).

Z napredovanjem molekularnih tehnik in večanjem računalniške moči, je danes možen ulov brez fizičnega stika z osebki. Tako je manj verjetno verjetno, da bo spremenila vedenje in posledično spremenila njen ulovljivost. Iz neinvazivno nabranih vzorcev (npr. sluzi, sline, dlake, iztrebka, perja) preko DNK z relativno visoko zanesljivostjo določimo identiteto osebka. Pri živalih, ki imajo značilno obarvano kožo ali dlako (npr. trebuh nekaterih dvoživk, obarvanost kožuha pri risu), lahko s pomočjo fotografij in računalniških programov ločimo med različnimi osebki.

### 1.3.3 Območje vzorčenja

Ulov in označevanje osebkov poteka na geografsko določenem območju, ki za živali z velikim območjem gibanja le deloma zaobjame proučevano populacijo. Velikost območja vzorčenja je navadno omejeno zaradi velikih razdalj potovanj, fizičnimi mejami (otoki) ali pa s političnimi mejami (parkov, držav ali drugih enot). Iz tega sledi, da osebke zaznavamo (lovimo), njihovo gibanje ni nujno omejeno na naše arbitralno določeno območje vzorčenja.

### **1.3.3.1 Domač okoliš**

Območje domačega okoliša osebka predstavlja prostor, kjer se zadržuje osebek med dnevnimi aktivnostmi. Raba prostora navadno sovpada s trenutnimi potrebami osebkov (iskanje hrane, partnerjev) in se v času lahko spreminja. Velikost območja je odvisna od velikosti organizma ter razpoložljivosti hrane in drugih virov potrebnih za življenje.

### **1.3.3.2 Superpopulacija**

Populacija je skupina osebkov, ki se nahajajo znotraj območja vzorčenja. Poleg osebkov, ki se neprestano zadržujejo znotraj območja vzorčenja, lahko zaznamo tudi osebke, ki prehajajo njegove meje, ker se njihov domač okoliš le deloma prekriva z našim arbitarnim območjem vzorčenja. Njihova zaznavnost je v povprečju nižja od osebkov, ki se v prostoru in času stalno nahajajo v območju vzorčenja. Skupaj z ostalimi osebki tvorijo t.i. superpopulacijo, za katero pa ne vemo natančno kako je prostorsko omejena.

### **1.3.4 Predpostavke metod ulova in ponovnega ulova**

Kot velja za vse modele, tudi pri modelih ulova in ponovnega ulova veljajo predpostavke. Kršitev le-teh ima lahko za posledico blage ali hude posledice. V splošnem se trudimo, da predpostavkam zadostimo. V primerih simuliranih vrednosti lahko nekatere predpostavke zanemarimo ali pa jih celo nadzorujemo, kar nam pomaga proučevati kaj se dogaja pri kršenju le-teh.

### **1.3.5 Zanesljivost označevanja**

Da osebek nedvoumno in zanesljivo označimo pomeni, da se oznaka v prostoru in času ne izgubi ali spremeni in da jo raziskovalec lahko zanesljivo prebere. V preteklosti se je uporabljalo neposredne načine označevanja (ušesne oznake, tetovaže, rezanje uhljev, barvanje kremljev...) danes pa se za vsaj nekatere skupine živali (npr. večji sesalci) uporabi pristope s pomočjo genetike. Ker za nabiranje nekaterih tipov vzorcev (iztrebkov, dlak, peres) ne potrebujemo neposrednega stika z osebkom, to imenujemo neinvazivno vzorčenje. Pri genetskem vzorcu je zaradi napak genotipizacije ali identičnih dvojčkov mogoče, da genotip določimo narobe, vendar so metode kljub temu relativno zanesljive v smislu označevanja osebkov (Waits in Paetkau, 2005). Za živali, ki imajo značilne vzorce kožuha se uporablja lahko foto pasti, ki so doživele razmah uporabe v zadnjih desetih letih.

### 1.3.6 Enaka ulovljivost

Predpostavlja se, da imajo vse živali enako verjetnost, da jih bomo zaznali oz. ulovili. Do neenake ulovljivosti lahko pride na več načinov. Če imajo živali različne ulovljivosti govorimo o heterogenosti ulovljivosti. Zvišanje (t.i. trap-happy) ali znižanje ulovljivosti (i. trap-shy) je posledica pristranskega vzorčenja, bodisi zaradi opazovalca ali pa preučevanega subjekta (npr. privabljanje ali odganjanje od detektorja). Avtorji opozarjajo, da kršenje te predpostavke vpliva na pristranskost cenilke velikosti populacije. Smer pristranskosti je znana. Ko je ulovljivost podcenjena, je ocena velikosti populacije precenjena. Če ulovljivost precenimo, je ocena velikosti populacije podcenjena (Williams in sod., 2002). Na žalost pa del, ki bi preučevali ta fenomen tudi kvantitativno in tako pomagal praktikom oceniti velikost napake zaradi ponesrečenega vzorčenja, skoraj ni.

### 1.3.7 Zaprtost populacije

Zaprta populacije je tista, pri kateri v času vzorčenja ne prihaja do sprememb v smislu emigriranja, imigriranja, rojstev ali smrti. Drugače rečeno, populacija se strukturno ne spreminja. Tudi za to predpostavko obstajajo testi, kot so jih predlagali Otis in sod. (1978) in Stanley in Burnham (1999). Teste pesti preobčutljivost ali pa prenizka moč, da zaznajo nekatere tipe zaprtosti populacije.

Modeli za zaprte populacije lahko predvidevajo heterogenost ulovljivosti, vedenjski odziv ali spremembo ulovljivosti v času (Otis in sod., 1978) zaradi česar so lahko kljub učinku roba uporabni v smislu, da niso pristranski. Kljub temu pa to prehajanje lahko zmanjša ulovljivost do te mere, da je parametre modela težko oceniti oz. so le-ti nesmiselni (npr. interval zaupanja od  $-\infty$  do  $\infty$ ). Pri organizmih, ki imajo že zaradi svoje biologije nizko zaznavnost (npr. ris), je to lahko za raziskavo pogubno.

## 1.4 CENILKA VELIKOSTI POPULACIJE ZA DVA ODLOVNA INTERVALA

Cenilka za velikost populacije, ko vzorčimo v dveh odlovnih intervalih, imenujemo Lincoln-Petersenova cenilka. Ime dobila po Lincolnu in Petersenu, ki sta jo uporabili pri svojem delu, vendar nista avtorja. Pred njima jo je uporabil že vsaj Laplace, ko je v 18. stoletju ocenil število prebivalcev v Franciji. Lincoln-Petersonovo cenilko je

mogoče izpeljati na več načinov, spodaj pa bomo predstavili le eno, ker se navezuje na Hugginsov model za zaprte populacije, ki ga uporabimo v tej raziskavi.

Pri ulovu nedvoumno označimo osebke in jih spustimo, hkrati pa zabeležimo vse že prej označene osebke. Lovno zgodovino si lahko predstavljamo kot matriko velikosti  $n \times m$ . Vrstice predstavljajo osebke, stolci pa odlovnne presledke. V primeru ulova in prepoznavanje osebka v določenem odlovnem presledku je vrednost matrike na tem mestu 1, sicer pa 0. Spodnja matrika prikazuje tri osebke, ki so bili ujeti v dveh odlovnih presledkih. Prvi osebek je bil ujet (in označen) samo v prvem presledku, drugi osebek samo v zadnjem, tretji osebek pa je bil zabeležen v obeh odlovnih presledkih.

$$M = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$$

Lincoln-Petersenova cenilka (Williams in sod., 2002) je prizanjena za dva odlovna intervala in temelji na razmerju med številom ujetih živali v prvem intervalu ( $n_1$ ) in populacijo ( $N$ ) ter razmerju med številom ujetih živali v obeh intervalih ( $m_2$ ) in številom živali ujetih v drugem intervalu ( $n_2$ ).

$$\frac{n_1}{N} = \frac{m_2}{n_2}$$

in če izpostavimo N, dobimo cenilko

$$\hat{N} = \frac{n_1 n_2}{m_2}$$

V primeru, da imamo dva odlovna intervala, je verjetnost ulovljivosti  $p^*$  produkt verjetnosti ulovljivosti v prvem ( $p_1$ ) in drugem ( $p_2$ ) odlovnem intervalu s sledečo enakostjo

$$p^* = 1 - (1 - p_1)(1 - p_2)$$

Parameter  $\hat{p}^*$  lahko izrazimo tudi s pomočjo pogojnega multinomskega modela

$$P(x_{ij} \mid r, p_1, p_2) = \frac{r!}{x_{11}!x_{10}!x_{01}!} \cdot \left(\frac{p_1 p_2}{\hat{p}^*}\right)^{x_{11}} \left(\frac{p_1 q_2}{\hat{p}^*}\right)^{x_{10}} \left(\frac{q_1 p_2}{\hat{p}^*}\right)^{x_{01}}$$

kjer je  $r$  unikatno število ujetih živali. Tako dobimo oceno parametrov največjega verjetja

$$\hat{p}_1 = \frac{x_{11}}{x_{11} + x_{01}}$$

ter

$$\hat{p}_2 = \frac{x_{11}}{x_{11} + x_{10}}$$

S tem lahko izrazimo  $p^*$

$$\begin{aligned} p^* &= 1 - (1 - \hat{p}_1)(1 - \hat{p}_2) \\ &= \frac{rx_{11}}{(x_{11} + x_{10})(x_{11} + x_{01})} \end{aligned}$$

,

ki ga lahko uporabimo v kanonični cenilki za  $\hat{N}$

$$\hat{N} = \frac{r}{\hat{p}^*} = \frac{(x_{11} + x_{10})(x_{11} + x_{01})}{x_{11}} = \frac{n_1 n_2}{m_2}$$

Ker je Lincoln-Petersenova cenilka pristranska, in to inverzno v primerjavi z velikostjo vzorca, je popravek

$$\hat{N} = \frac{(n_1 + 1)(n_2 + 1)}{m_2 + 1} - 1$$

Varianca za  $\hat{N}$  je  $[(n_1 + 1)(n_2 + 1)(n_1 - m_2)(n_2 - m_2)]/[(m_2 + 1)(m_2 + 2)]$ , interval zaupanja pa je mogoče izpeljati na različne načine (Williams in sod. (2002), str 291).

Za modeliranje heterogenost ulovljivosti so razvili model  $M_h$  (Otis in sod., 1978)

$$P(f_1, \dots, f_K \mid F) = \frac{N!}{(\prod_{j=1}^K f_j!)(N - M_{K+1})} \pi_0^N - M_{K+1} \prod_{j=1}^K \pi_j^{f_j}$$

kjer je parameter  $\pi_j = \int_0^1 \frac{K!}{(K-j)!j!} p^j (1-p)^{K-j} dF(p)$

## 1.5 HUGGINSOV MODEL ZA ZAPRTE POPULACIJE

Huggins-Alho model (bolje znan kot Hugginsov model kot ga tudi imenujemo v tem delu, Huggins (1989), formula (1)) je podoben temu kar smo videli že za model, ki predvideva heterogenost ulovljivosti ( $M_h$ ). Verjetnostna funkcija je

$$L = \prod_{i=1}^n \prod_{j=1}^t p^{I_{ij}(1-p_{ij})(1-I_{ij})} \quad (1)$$

Ključni dodatek je ocena ulovljivosti z uporabo spremenljivk na nivoju posameznika s pomočjo že prej omenjene logistične funkcije.

$$p_i = \frac{e^{b_0 + \beta_1 x_i}}{1 + e^{b_0 + \beta_1 x_i}}$$

,

V tem primeru predstavljata  $\beta_0$  in  $\beta_1$  parametra, ki jih ocenimo s pomočjo individualnih spremenljivk (logistična regresija). Hugginsov model ne vsebuje parametra za velikost populacije  $\hat{N}$ , zato se ga oceni s pomočjo cenilke

$$\hat{N} = \sum_{i=1}^{M_K+1} \frac{1}{\hat{p}_i^*}$$

in

$$\hat{p}_i^* = 1 - \prod_{j=1}^K (1 - \hat{p}_i) = 1 - (1 - \hat{p}_i)^K$$

$\hat{p}_i^*$  je ocenjen parameter verjetnosti, da je osebek  $i$  ujet vsaj enkrat.

### 1.5.1 Model CAPWIRE

Modeli paketa CAPWIRE (Miller in sod., 2005) gradijo na osnovi modelov z mešano ulovljivostjo. Za vhodne podatke uporabi skupno število ulovov posameznega osebka, ne glede na število odlovnih presledkov ali koliko vzorcev je bilo nabranih v posameznem presledku. Modele parametra oceni s pomočjo multinomske porazdelitve.

Na voljo sta dva modela, ECM in TIRM. Prvi predpostavlja, da imajo vsi osebki enako verjetnost, da so zaznani. Model TIRM predpostavlja, da obstajata dve skupini osebkov, ki imata dve različni ulovljivosti. Skupino modelov, ki predpostavljajo več skupin, imenujemo modeli mešane ulovljivosti.

## 1.6 UČINEK ROBA

Prehajanje roba vzročenega območja povzroča kršitev predpostavke o zaprtosti populacije. To zaznamo kot heterogenost ulovljivosti, kar ima za posledico, da je ocena velikosti populacije vzorčenega območja pristranska. Vpliv tega pojava smo opazili pri naknadni analizi podatkov, ki smo jih uporabili za oceno številčnosti medvedov v Sloveniji (Skrbinšek s sod. 2018), saj nekateri osebki prehajajo robove območja vzorčenja preko državne meje Slovenija-Hrvaška. Kendall (1999) je pokazal, da je verjetnost ulovljivosti osebkov za populacije, ki prehajajo rob vzorčenega območja  $E(\hat{p}_i) = \tau_i p_i$ .  $\tau_i$  je verjetnost, da se osebek nahaja znotraj vzorčenega območja, pi pa ulovljivost. Ocena velikosti populacije bo nepristranska, če je  $\tau_i = 1$ . V primeru, da je  $\tau_i$  manjši od 1 pa bo ocena pristranska. To imenujemo učinek roba (Hansson, 1969; White in sod., 1982; Wilson in Anderson, 1985) in je v študijah ulova-ponovnega ulova nezaželen učinek.

### 1.6.1 Poskusi reševanja posledic učinka roba

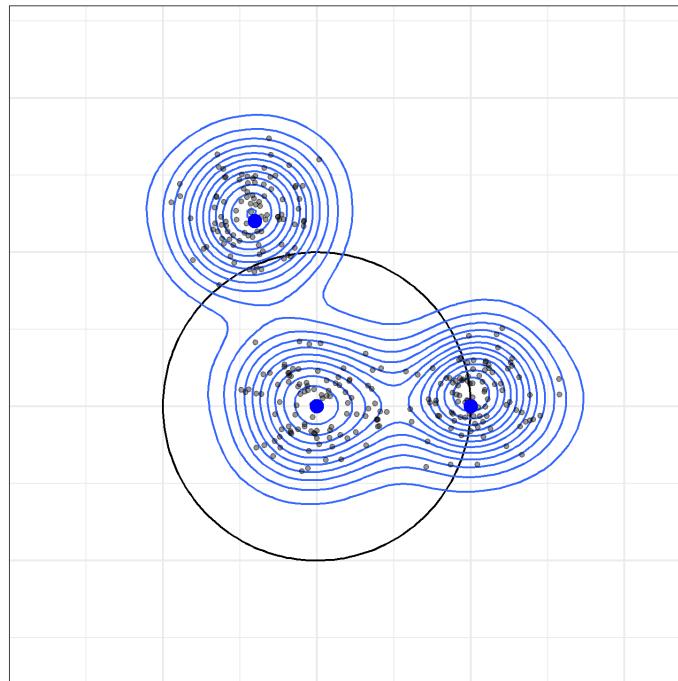
Učinek roba je že dolgo proučevan problem (Efford, 2004). Prvič se v literaturi s tem ukvarja (Dice, 1938, 1941), ki predлага, da se območje vzorčenja poveča za polmer domačega okoliša proučevane vrste. Dosedanje raziskave se osredotočajo predvsem na vzorčenje s pomočjo pasti nameščenih v obliki mreže ali sita (Williams in sod., 2002). V delih, v katerih se ukvarjajo s popravljanjem učinka roba, navadno uporabljajo vzorčenje na situ (Parmenter in sod., 2003). Drugi so predlagali podobne popravke, kot so na povprečni najdaljši premik (MMDM, angl. mean maximum distance moved, Wilson in Anderson (1985)), razdalje med pastmi, povprečna razdalja premikanja med pasti in druge (Miller in sod., 1997; Boulanger in McLellan, 2001; Royle in sod., 2013).

Drug možni pristop se razlikuje od prejšnjih po tem, da se oceni kakšen odstotek časa preživijo posamezniki zunaj oz. znotraj vzorčenega območja. Ivan in sod. (2013) so to želeli oceniti s pomočjo telemetrije. Royle in sod. (2013) poročajo, da sta White in Shenk (2001) uporabila podoben pristop, kjer sta s pomočjo telemetrije ocenila verjetnost ulovljivosti ( $\Psi$ ) in ocenila gostoto kot  $\hat{D} = \frac{\hat{N}\Psi}{A}$  kjer je A območje vzorčenja,  $\hat{N}$  pa ocenjena velikost populacije s pomočjo modelov za zaprte populacije. Kritika teh pristopov je, da morajo biti osebki, s pomočjo katerih ocenjujemo  $\Psi$ , opremljeni s telemetrijskimi ovratnicami reprezentativno glede na njihovo dejansko ulovljivost. Omenjeni pristopi k reševanju posledic učinka roba so ad hoc. Rešitve ali poskusi njih, ki so nastale z namenom reševanja praktičnih problemov, s katerimi so se v preteklosti srečevali biologi in drugi raziskovalci naravoslovja. V praksi to pomeni, da nimajo splošne oblike in jih zato ni mogoče nadgraditi (Royle in sod., 2013).

V nekatere modele lahko vključimo spremenljivke na nivoju posameznika. S pomočjo dodatne informacije lahko dodatno pojasnimo razlike v ulovljivosti med osebkami ali skupinami osebkov. Z uporabo logistične funkcije je mogoče oceniti verjetnosti ulovljivosti  $p$  in ponovne ulovljivosti  $c$  kot jih opredeli Hugginsov model (Boulanger in McLellan, 2001; Boulanger in sod., 2004). Ta pristop je nekje med modeli, ki prostora ne upoštevajo, in pristopom, ki vgradi prostorsko komponento neposredno v model.

### 1.7 NAŠ PRISTOP K REŠEVANJU POSLEDIC UČINKA ROBA

Na podlagi preliminarnih analiz podatkov za oceno številčnosti medvedov v Sloveniji leta 2007 (Skrbinšek s sod., 2018) bomo v tem delu raziskali učinkovitost možnega popravka posledic učinka roba. V poglavju ”Učinek robašmo omenili, da sta Boulanger in McLellan (2001) v modelu uporabila mero, ki je povezana z razdaljami premikanja med centroidom domačega okoliša in robom območja vzorčenja. S tem naj bi v modelu upoštevala razliko v ulovljivosti osebkov in izboljšala oceno parametrov. V tem delu predlagamo drugo mero, za katero menimo, da bi lahko bolje opisala gibanje osebka znotraj vzorčenega območja in s tem bolje prispeva k popravku učinka roba.



Slika 1: Primer pojavljanja treh osebkov in območja vzorčenja. Modre pike predstavljajo centroide domačega okoliša, sive točke pa točke, kjer se osebki (potencialno) pojavljajo. Modre črte povezujejo točke enake verjetnosti pojavljanja osebka.

Predpostavimo, da imamo območje, kjer se nahajajo osebki in je praviloma večje od



Slika 2: Levo je ponazorjena verjetnost pojavljanja osebka oz. verjetnost, da ga zaznamo, v odvisnosti od oddaljenosti od centroida domačega okoliša. Prikaz desno ponazarja "ulov" osebka (sive pike) v prostoru, krivulje na desni in zgornji stranici pa sta robni gostoti.

območja vzorčenja (slika 1). Na območju so razporejeni centroidi domačih okolišev osebkov. Vsak osebek ima samo en centroid, okoli katerega se giba. V centroidu domačega okoliša je verjetnost ulovljivosti največja, z oddaljevanjem od le-tega pa pada (slika 2). Gibanje bi tako lahko opisali s simetrično dvorazsežnostno normalno porazdelitvijo ali pol-normalno porazdelitvijo  $p_{ij} = p_0 \exp^{-(\frac{1}{2\sigma^2} \cdot d)}$  (Royle in sod., 2013, stran 127) kjer je  $d$  razdalja od centroida domačega okoliša osebka do točke ulova,  $\sigma^2$  pa varianca, ki je parameter te porazdelitve. Uporabo domačega okoliša si lahko predstavljamo kot zvonec z vrhom v centru. Veliko večino točk pričakujemo na oddaljenosti treh standardnih odklonov od centroida.

Osebkom, ki imajo pretežni delež domačega okoliša znotraj vzorčenega območja, lahko pripisemo delež njihovega domačega okoliša znotraj območja vzorčenja 1. V kolikor pa se centroid osebka nahaja v bližini roba območja vzorčenja, lahko pričakujemo, da bo del časa osebek preživel zunaj območja. V tem času med vzorčenjem ne bo zaznaven (ulovljiv), zato bo delež njegovega domačega okoliša znotraj območja vzorčenja manjši od 1. Podobno velja za krivulje, ki nimajo parametrične oblike (neparametrične krivulje).

Osebke med premikanjem lovimo in beležimo lokacijo in čas, ko smo jih zaznali. Točke na sliki 2 desno predstavljajo take ulove. Predpostavljamo, da te točke predstavljajo reprezentativno gibanje osebkov. Histogram razdalj med pari točk predstavlja razdalje, ki jih osebek lahko prehodi. To porazdelitev uporabimo pri izračunu deleža, ki pripada znotraj območja vzorčenja. Delež porazdelitve znotraj območja vzorčenja uporabimo v modelu kot individualno spremenljivko.

### 1.7.1 Simulacije kot orodje za raziskovanje

V matematičnih modelih opišemo pojav s funkcijami, ki so lahko preproste ali kompleksne. S kompleksnostjo pojava pa narašča tudi kompleksnost funkcij. Drug pristop je opis pojava s stohastičnimi procesi - simulacijami. Pojav razdelimo na manjše podenote in vsako enoto simuliramo kot samostojni proces. Razpon vrednosti, ki jih zavzemajo parametri podenote določi raziskovalec, najpogosteje na podlagi teoretskih osnov in poznavanje pojava. Podenota ni samostojna ampak je odvisna od drugih podenot in parametrov, ki jih določi raziskovalec. Primer take podenote je molekula vode, ki se obnaša na svoj način v razmerju do sosednjih molekul. Ko na sistem pogledamo od daleč lahko vidimo kako se le-ta obnaša glede na različne vhodne parametre. Na ta način simulirajo npr. ladijske elise, letalska krila ipd. V ekologiji lahko na tak način simuliramo gibanje in obnašanje živali ter preučujemo interakcijo med osebki oz. višjimi enotami in okoljem (Bolker, 2008).

## 1.8 CILJI DISERTACIJE

Cilj disertacije je izboljšati računanje gostote populacij, kjer med vzorčenjem prihaja do kršitve predpostavke o zaprtosti populacij.

Na podlagi razdalj med vzorčenimi lokacijami osebkov v območju vzorčenja bomo ocenili delež, ki ga osebek prebije znotraj območja vzorčenja. To informacijo bomo vključili v model za zaprte populacije kot individualno spremenljivko.

Ker bomo v modelu, ki upošteva individualno spremenljivko, v modeliranje vključili več informacij pričakujemo, da bo boljši od tistega, kjer te spremenljivke ne bomo uporabili. Za boljši model bomo smatrali tistega, ki bo imel nižji kriterij AICc.

Z vključitvijo informacije o morebitnem prehajanju roba vzorčenega območja v model, bomo izboljšali oceno ulovljivosti, saj bomo v model vključili eno od pomembnih komponent, ki vplivajo na spremenjeno zaznavnost in posledično ulovljivost osebkov. Pričakujemo, da bo model, ki vključuje individualno spremenljivko, uspešno kompenziral za spremenjeno ulovljivost zaradi prehoda čez rob območja vzorčenja.

Izboljšana ocena gostote bo na račun boljšega poznavanja velikosti dejanskega prispevka območja osebkov v območje vzorčenja, saj bomo iz gibanja osebkov lahko raz-

brali kako so razporejeni v prostoru.

Manjše kršitve predpostavke zaprtosti populacije ne bi smele bistveno vplivati na prisstranskost ocene gostote. Pričakujemo, da bo razlika med ocenjeno in dejansko gostoto odvisna od razmerja med površino domačega okoliša in površino vzorčenega območja.

## 2 MATERIALI IN METODE

V našem primeru bo osnovni gradnik simulacije osebek (DeAngelis in Grimm, 2014). Posamezniki v teh modelih imajo lastnosti, ki so lahko povezane s prostorom in časom, fiziološke lastnosti ali pa so v povezavi z drugimi osebki iste ali druge vrste. Ker parametre določimo sami, imamo veliko kontrole nad naključnostjo (stohastične), kar je še posebej uporabno pri proučevanju vpliva variabilnosti na pojav, ki ga raziskujemo.

### 2.1 SIMULIRANJE GIBANJA OSEBKOV

Simulacijo smo postavili v svet A, ki je tako velik, da noben simuliran osebek, ki ni imel nezanemarljivo možnost biti zaznan, ne bo imel praktične možnosti, da bi dosegel rob. To je delovalo kot varovalka, da ni prišlo do popačenja rezultatov pri simulacijah s parametri na skrajnosti svojih razponov. Znotraj sveta A smo centroide osebkov v območju S razporedili naključno s pomočjo enakomerne porazdelitve. Gibanje osebkov smo opisali z dvorazsežno normalno porazdelitvijo. V simulacijah smo generirali 500, 1000, 1500 in 2000 osebkov.

Polmer domačega okoliša osebkov smo določili v enotah standardnega odlkona ( $\sigma$ ) normalne porazdelitve. V centroidu je imel osebek največjo ulovljivost, z oddaljenostjo od centroida pa verjetnost pada glede na funkcijo normalne porazdelitve.

#### 2.1.1 Odlov

Osebke smo vzorčili s ponavljanjem v K odlovnih presledkih (5, 10, 15), kjer smo v posameznem presledku zabeležili po eno lokacijo na osebek. Ulovljivost je za vse osebke enaka in določena s parametrom  $p$ . Parameter  $p$  smo simulirali med 0.1 in 0.3, kar predstavlja pogoste realne ulovljivosti, ki jih poročajo v študijah ulova-ponovnega ulova (npr. Wilson in Anderson (1985); Foster in Harmsen (2012); Chandler in sod. (2018)).

Pri ulovu smo zabeležili tudi lokacijo znotraj vzorčenega območja, točke zabeležene zunaj pa smo krnili. Za osebke, od katerih smo dobili dva ali več vzorcev, smo izračunali vse možne pare razdalj med točkami. Te razdalje so osnova za izračun individualne

spremenljivke.

### 2.1.2 Izračun individualne spremenljivke

Izračune smo izvajali na mreži celic oziroma rastru. Velikost celic smo nastavili na tako velikost, da po našem mnenju izbira velikosti ni vplivala na izračun. Glede na oddaljenost celice od centroida osebka smo s pomočjo posamezne funkcije za vsako celico rastra izračunali verjetnost pojavljanja, celice znotraj območja vzorčenja pa sešeli. Ta vrednost je povezana s časom, ki ga osebek porabi znotraj območja vzorčenja oziroma s časom, v katerem je na voljo za odlov.

#### 2.1.2.1 Dvorazsežnostna normalna porazdelitev

Gibanje na podlagi dvorazsežnostne normalne porazdelitve se ujema s tistim, ki smo ga uporabili za simuliranje osebkov. Centroid smo izračunali kot geometrično sredino vseh vzorčenih točk osebka, za standardni odklon pa smo uporabili simulirano vrednost.

#### 2.1.2.2 Empirična porazdelitev

Za prileganje empirične porazdelitve smo uporabili podatke o razdaljah med pari vseh možnih vzorčenih točk. Na podlagi števila celic in razdalje med centroidom vzorčenega območja in njegovim robom smo ustvarili primerno število razdelkov, katerih število ni vplivalo na točnost izračuna individualne spremenljivke. Vsako parno razdaljo smo uvrstili v razdelek in tako pridobili kumulativno število prehodnih razdalj (padajoča funkcija).

Razdelke smo dodatno utežili na sledeč način. V območje vzorčenja smo na podlagi enakomerne porazdelitve ustvarili toliko naključnih točk, kot je bilo vzorčenih parnih razdalj. Iz vseh empiričnih prehodnih razdalj smo s ponavljanjem vzorčili dolžine in jih postavili v prostor z izhodiščem v prej naključno vzorčenih točkah, smer daljice pa je bila naključna. Delež daljic, ki se je nahajal popolnoma znotraj območja vzorčenja, so predstavljali utež razreda. Uteženemu histogramu smo prilegli in ocenili funkcijo s tremi parametri, ki spominja na kumulativno porazdelitveno funkcije Weibullove porazdelitve.

$$f(x) = e^{-(\frac{x}{\lambda})^k} \cdot a$$

za  $x \geq 0$ . Za lažji izračun deleža domačega okoliša znotraj vzorčenega območja smo dodali parameter  $a$ , ki spremeni vrednost, kjer funkcija seká  $y$  os pri  $x = 0$ . Za vse osebke smo predpostavili enako krivuljo, ki opiše rabo domačega okoliša okoli centroida. Na podlagi krivulje in območja vzorčenja smo izračunali relativni delež, ki ga posamezen osebek prebije znotraj vzorčenega območja.

Porazdelitev smo izbrali po tem, ko smo izkustveno določili, da relativno dobro opiše porazdelitev uteženih dolžine prehojenih razdalj.

## 2.2 OCENJEVANJE VELIKOSTI POPULACIJ

Velikost populacije smo ocenili s pomočjo dveh modelov, Hugginsovega modela in modela CAPWIRE. Iz prvega smo izluščili parametra ulovljivost in velikost ocenjene populacije ter pripadajoče intervale zaupanja, iz drugega pa samo velikost populacije in pripadajoč interval zaupanja. V vseh primerih smo uporabili 95 % interval zaupanja.

### 2.2.1 Hugginsov model

Z vzorčenjem smo za vsak osebek ustvarili odlovnno zgodovino z individualno spremenljivko v K odlovnih presledkih. Delež časa, ki ga je osebek preživel znotraj območja vzorčenja glede na eno od zgoraj opisanih krivulj, smo uporabili kot individualno spremenljivko.

$M_0$ : V prvem modelu smo predpostavili, da sta parametra za prvi in vsak nadaljnji ulov enaka ( $c = p$ ), ulovljivost pa je odvisna samo od odlovne zgodovine (brez drugih spremenljivk). Ta model bi v zapisu, ki ga uporablja (Otis in sod., 1978), imenovali  $M_0$  in predpostavlja, da učinka roba ni.

$M_{sp}$ : Za drugi model smo prav tako predpostavili enakost parametrov ( $c = p$ ), ulovljivost pa je odvisna od individualne spremenljivke. Le-to smo izračunali na dva različna načina (glej poglavje Izračun individualne spremenljivke). Ta model bo predvidoma upošteval, da obstaja učinek roba in ga v nadaljevanju označujemo kot  $M_{sp}$ .

### 2.2.2 Po modelu CAPWIRE

Pri modeliranju s paketom CAPWIRE smo uporabili model TIRM, ki predpostavili, da gre za dve skupini osebkov z različnimi verjetnostmi ulovljivosti. Odlovne zgodovine smo sešeli po osebku in število ulovov uporabili v modelu.

## 2.3 STATISTIKE ZA PRIMERJAVO UČINKOVITOSTI DELOVANJA POPRAVKA

Za vsak set parametrov (npr. velikost populacije, velikost domačega okoliša, ulovljivost...) smo izračunali tri modele - po Hugginsovem modelu z in brez modifikacije ulovljivosti s pomočjo individualne spremenljivke ( $M_0$ ,  $M_{sp}$ ) in po modelu TIRM, kjer smo predpostavili dve skupini, ki imata lahko različni ulovljivosti.

Za vsak model smo zabeležili vrednosti parametrov, ki smo jih uporabili za simuliranje osebkov; število simuliranih osebkov, število odlovnih presledkov, polmer vzorčenega območja, število osebkov, ki smo jih zaznali (ujeli znotraj območja vzorčenja vsaj enkrat), standardni odklon uporabljen pri simuliranju domačega okoliša osebka, povprečni najdaljši premik in simulirana verjetnost ulovljivosti.

Od ocenjenih parametrov Hugginsovih modelov smo zabeležili oceno ulovljivosti za posamezen odlovni presledek (in pripadajoč 95% interval zaupanja) za modela  $M_0$  in  $M_{sp}$ , ocenjena velikost populacije (in pripadajoč 95% interval zaupanja), AIC ter razliko v AIC med modeloma  $M_0$  in  $M_{sp}$ . Za model TIRM smo shranili le velikost populacije in interval zaupanja, ker ocena ulovljivosti s Hugginsovim modelom ni neposredno primerljiva. Poleg populacijskih statistik, ki jih je izračunal program MARK, smo shranili tudi parametre prileganja funkcij, s pomočjo katerih smo opisali parne razdalje med vzorčenimi točkami osebkov. Za (pol)normalno porazdelitev je to standardni odklon ( $\sigma$ ), za empirično porazdelitev pa trije parametri ( $\lambda$ ,  $k$ ,  $a$ ).

Za vsako simulacijo smo izračunali sledeče statistike:

- Gostoto osebkov brez popravka, t.i. "naivna gostota". Izračunali smo jo z deljenjem števila simuliranih osebkov ( $N$ ) z velikostjo območja vzorčenja.
- Gostoto osebkov s popravki, kjer smo vzorčeno območje povečali za razdaljo enako 50., 60., 70., 80., 90., 95. in 99. percentilu normalne ali empirične porazdelitve.

Računanje percentila za normalno porazdelitev z znanim povprečjem in standarni odklon je preprosto. S pomočjo kvantilne funkcije (*qnorm*) izračunamo pri kateri verjetnosti (prehodjeni razdalji) je vrednost naključne spremenljivke manjša ali enaka dani verjetnosti. Za empirično porazdelitev kvantilna funkcija ni znana, oz. je relativno zapletena. S pomočjo simulacijskega pristopa in vzorčenjem z zavračanjem smo simulirali mnogo vrednosti, ki se prilegajo empirični porazdelitvi z znanimi parametri. Percentile smo izračunali na podlagi simuliranih vrednostih. Percentile smo uporabili kot razdaljo, za katero povesmo povečali čamо območje vzorčenja in posledično izračunali spremenjeno gostoto.

- AICc. Modela  $M_0$  in  $M_{sp}$ , ki ocenita velikost populacije brez in s popravkom, smo primerjali glede na kriterij AICc. Kriterij nam omogoča rangirati modela, pri tem pa tehta med zapletenostjo in prileganjem modela podatkom. Uporabili bomo AICc, ki se od AIC razlikuje po tem, da je bolj primeren za manjše vzorce. Model z nižjo vrednostjo AICc in za razliko vsaj 2, bo boljši od drugega modela.

Izračunane statistike gostote smo primerjali s pravo gostoto, t.i. zlatim standardom, ki smo ga izračunali tako, da smo število generiranih centroidov (osebkov) delili s površino, ki smo jo uporabili za generiranje teh osebkov. Izračunali smo kazalnik, ki prikazuje delež ocenjene gostote od prave gostote ( $I = \frac{\hat{D}}{D}$ ).

Za prikaz trendov in razlik med skupinami smo uporabili posplošen linearen model (t.i. GLM, Faraway (2006)), z normalno razporejenimi ostanki.

Simulacije smo spisali v programskem okolju R (R Core Team, 2018). Pri tem bomo uporabili pakete **sp** (Bivand in sod., 2008), **raster** (Hijmans in Graham, 2006), **rgeos** (Bivand in Rundel, 2017), **cluster** (Maechler in sod., 2018), **splancs** (Rowlingson in Diggle, 2017), **foreach** (Microsoft in Weston, 2017)) in **doParallel** (Corporation in Weston, 2017).

Za oceno parametrov Hugginsovega modela smo uporabili paket **RMark** (Laake, 2013), ki omogoča posredno povezavo s programom MARK (Cooch in White, 2012). Model **TIRM** smo ocenili s pomočjo R paketa **capwire** (Pennell in Miller, 2012). Za vse grafike smo uporabili paket **ggplot2** (Wickham, 2009).

Simulacije bomo poganjali na delovni postaji na Oddelku za biologijo Katedra za ekologijo in varstvo okolja. Postaja ima v dveh procesorjih 24 sredic in 48 niti ter več kot 100 GB delovnega spomina. Osnovni operacijski sistem je CentOS, v katerem tečejo virtualni sistemi (odprtokodni sistemu Ubuntu 18.04). Disertacija je bila oblikovana s sistemom **L<sup>A</sup>T<sub>E</sub>X**.

### 3 REZULTATI

#### 3.1 SIMULACIJE

Skupno smo izvedli 6000 simulacij. Tri tisoč simulacij smo izvedli na način, da smo individualno spremenljivko izračunali s pomočjo simuliranih vrednosti standardnega odklona (normalna porazdelitev) in na podlagi vzorčenih točk izračunanim centroidom kot povprečje ( $\mu$ ). S pomočjo programa MARK smo uspešno ocenili v 2916 simulacijah. Enako število simulacij smo izvedli za primer, kjer smo krivuljo za izračun individualne spremenljivke ocenili iz podatkov (empirična porazdelitev). Pri oceni parametrov smo bili uspešni 2681-krat (Tabela 1).

Da nekaterih modelov nismo uspešno ocenili je po pričakovanjih. Zaradi določene strukture podatkov je možno, da optimizacijski kriterij znotraj parametrskega prostora ne najde optimuma, ali pa so ocenjeni parametri na intervalu  $-\infty < x < \infty$ . Tiste simulacije, ki niso dale smiselnega rezultata, smo iz analize izključili.

Na polno obremenjenem računalniku je bila simulacija izračuna individualne spremenljivke končana v približno 48 urah. Nadaljnje analize ulova in ponovnega ulova ter priprava podatkov za analizo pa so zahtevale še dodatnih šest ur.

Zaznati je mogoče vzorec v številu uspešnih simulacij glede na uporabljeni funkcijo. Število uspešnih simulacij glede na število generiranih osebkov in število odlovnih intervalov se med temo dvema skupinama razlikuje. Večina neuspelih simulacij pri uporabi empirične funkcije je za najmanjše število odlovnih intervalov ( $K = 5$ ).

Tabela 1: Prikaz števila simulacij, kjer smo uspešno ocenili parametre, glede na funkcijo, ki je bila uporabljena za izračun individualne spremenljivke. V vrsticah je število generiranih osebkov ( $N$ ), v stolpcih pa število odlovnih intervalov ( $K$ ). Uspešno izračunanih simulacij z empirično funkcijo je nekoliko manj. Za lažjo predstavo v spodnji matriki prikazujemo razliko med zgornjima tabelama tako, da smo empirični (levo) odšteli normalno (desno) tabelo. Pretežni delež neuspelih simulacij je predvsem na račun simulacij, ki imajo 5 odlovnih intervalov in/ali manjše število simuliranih osebkov (nižja gostota).

Empirična funkcija			Normalna funkcija			Razlike med matrikama			
$K = 5$	$k = 10$	$K = 15$	$K = 5$	$k = 10$	$K = 15$	$K = 5$	$k = 10$	$K = 15$	
$N = 500$	142	187	190	188	209	197	-46	-22	-7
$N = 800$	141	192	178	179	205	183	-38	-13	-5
$N = 1000$	135	182	185	175	190	185	-40	-8	0
$N = 1300$	168	212	194	191	212	195	-23	0	-1
$N = 1500$	171	186	218	199	190	218	-28	-4	0

### 3.2 PARAMETER P MODELOV $M_0$ IN $M_{sp}$ PO HUGGINSU

Za modela  $M_0$  in  $M_{sp}$  modelu smo predpostavili, da sta parametra  $p$  in  $c$  enaka. Ulovljivost v prvem in vseh ostalih odlovnih presledkih predpostavljamo enako. Zato v analizi obravnavamo le parameter ulovljivosti  $p$ . Za lažjo primerjavo smo izračunali razmerje med ocenjeno in simulirano vrednostjo  $p$  (slika 3). Model TIRM oceni  $p$  za dve skupini, kar neposredno ni primerljivo s parametrom  $p$  po Hugginsovem modelu, zato le-tega nismo podrobneje preučili.

V povprečju je ulovljivost za vse simulacije in modele podcenjena in je približno med 0.35 in 0.4 deleža simulirane vrednosti. Ulovljivost ocenjena po modelu  $M_{sp}$  je nekoliko bolj podcenjena, ne glede na to s pomočjo katere porazdelitve izračunamo individualno spremenljivko. V povprečju je ulovljivost za primer empirične porazdelitve za oba modela nekoliko nižja (za približno 0.01). Vidimo tudi, da se razlika med modeloma rahlo zmanjša s povečanjem simulirane ulovljivosti (slika 3).

#### 3.2.1 Primerjava simulirane in ocnjene ulovljivosti glede na porazdelitev

##### 3.2.1.1 Empirična porazdelitev

Povprečje deleža simulirane vrednosti preprostega  $M_0$  modela je 0.397, modela z individualno spremenljivko  $M_{sp}$  pa za 0.035 nižja.

Call:

```
glm(formula = p.val ~ p.var, data = xep)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.36868	-0.16667	-0.07254	0.09971	1.44373

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.397081	0.001467	270.69	<2e-16 ***
p.var.p.target.sp	-0.035010	0.002075	-16.88	<2e-16 ***
---				

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.05769313)

```
Null deviance: 3109.8 on 53619 degrees of freedom
Residual deviance: 3093.4 on 53618 degrees of freedom
AIC: -786.36
```

### 3.2.1.2 Normalna porazdelitev

Povprečje deleža simulirane vrednosti preprostega  $M_0$  modela je 0.387, modela z individualno spremenljivko  $M_{sp}$  pa 0.349 (0.038 nižja).

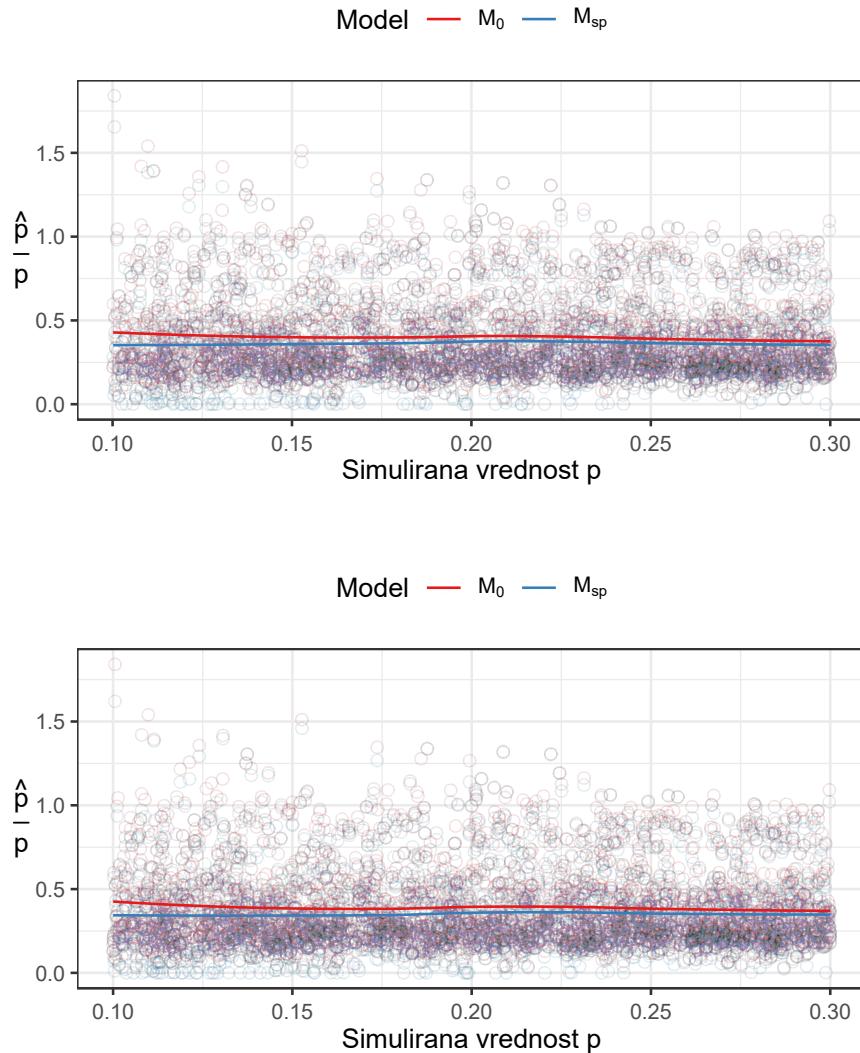
```
Call:
glm(formula = p.val ~ p.var, data = xep)

Deviance Residuals:
    Min      1Q Median      3Q     Max 
-0.35876 -0.16239 -0.07163  0.09433  1.45364 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.387167  0.001381 280.29 <2e-16 ***
p.varp.target.sp -0.037886  0.001953 -19.39 <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.05563641)

Null deviance: 3265.5 on 58319 degrees of freedom
Residual deviance: 3244.6 on 58318 degrees of freedom
AIC: -2972.7
```

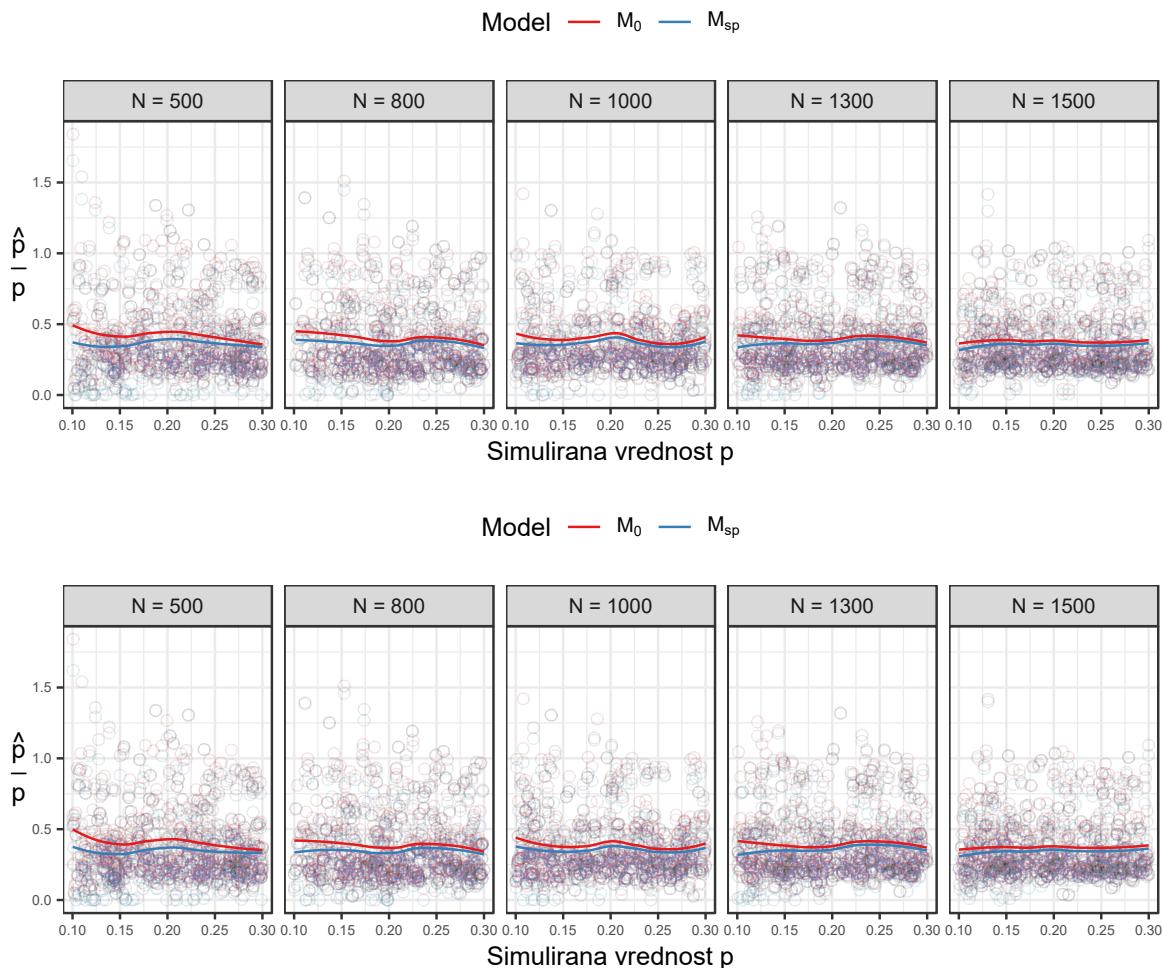


Slika 3: Povprečna ocena ulovljivosti (parameter  $p$ ) za Hugginsova modela  $M_0$  in  $M_{sp}$ . Na osi  $x$  so prikazane simulirane vrednosti  $p$ , na osi  $y$  pa razmerje med ocenjeno in simulirano vrednostjo. V primeru, da ne bi bilo kršenih predpostavk, bi morala biti ocena in simulirana vrednost približno podobni (in njuno razmerje  $\sim 1$ ). Izračun individualne spremenljivke s pomočjo empirične (zgoraj) in normalne funkcije (spodaj). Ne kaže, da izbira funkcije za izračun individualne spremenljivke bistveno vpliva na oceno parametra  $p$ .

### 3.2.2 Primerjava ulovljivosti glede na število simuliranih osebkov

Ocena ulovljivosti ( $p$ ) glede na število simuliranih osebkov kaže podobno sliko (slika 4). V povprečju model, kjer smo uporabili za izračun še individualno spremenljivko ( $M_{sp}$ ), bolj podcenjuje ulovljivost kot model brez nje ( $M_0$ ). Ta razlika je relativno majhna.

Opazili smo tudi trend, da je razkorak med modeloma večji za simulacije z manj osebkami za manjše simulirane ulovljivosti. Z večanjem simuliranega parametra  $p$  se razlika zmanjšuje tudi za primere z najmanjšim številom simuliranih osebkov. Razlike med simulacijami, kjer smo individualno spremenljivko izračunali s pomočjo empirične ali normalne porazdelitve, so relativno majhne.

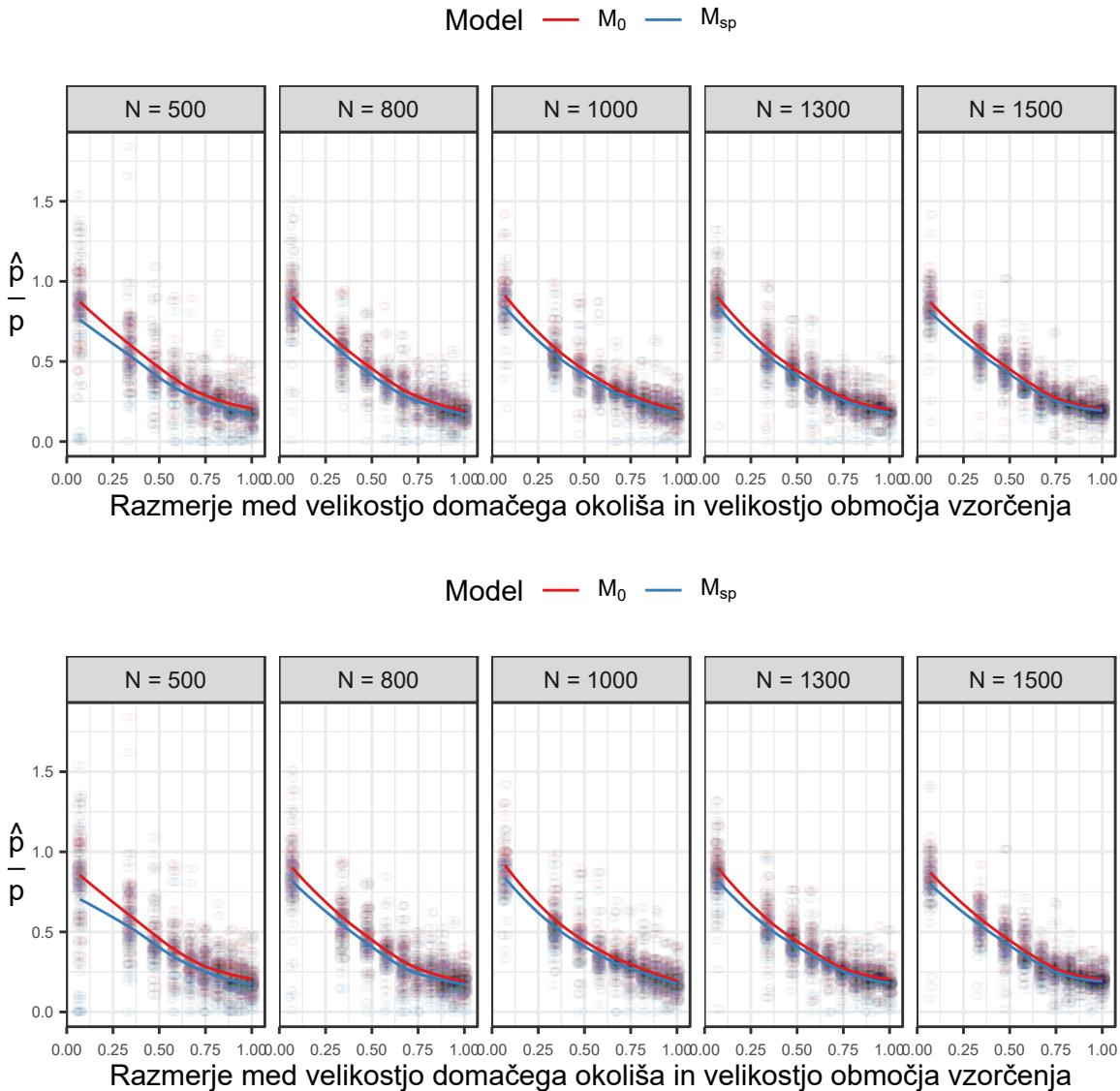


Slika 4: Zgornja slika prikazuje podatke, kjer smo individualno spremenljivko ocenili s pomočjo empirične spremenljivke, spodnja pa podatke za simulacije, kjer smo individualno spremenljivko izračunali s pomočjo normalne porazdelitve. Točke prikazujejo razmerje med oceno ulovljivosti in pravo, simulirano, vrednostjo glede na število generiranih osebkov. Na x osi so prikazane simulirane vrednosti ulovljivosti, na y osi pa razmerje med ocenjeno in simulirano vrednostjo. Okenca prikazujejo podatke glede na število simuliranih osebkov (od 500 do 1500).

### **3.2.3 Primerjava ulovljivosti glede na razmerje velikosti domačega okoliša in območja vzorčenja**

S slike 5 je razvidno, da je vpliv zelo velik. Ko je razmerje velikosti domačega okoliša in velikosti območja vzorčenja majhno (majhen domač okoliš v primerjavi z območjem zvorčenja), je ocena parametra relativno zanesljiva. Z večanjem razmerja med domačim okolišem in območjem vzorčenja pa se zanesljivost ocene poslabšuje in se celo spusti pod globalno povprečje.

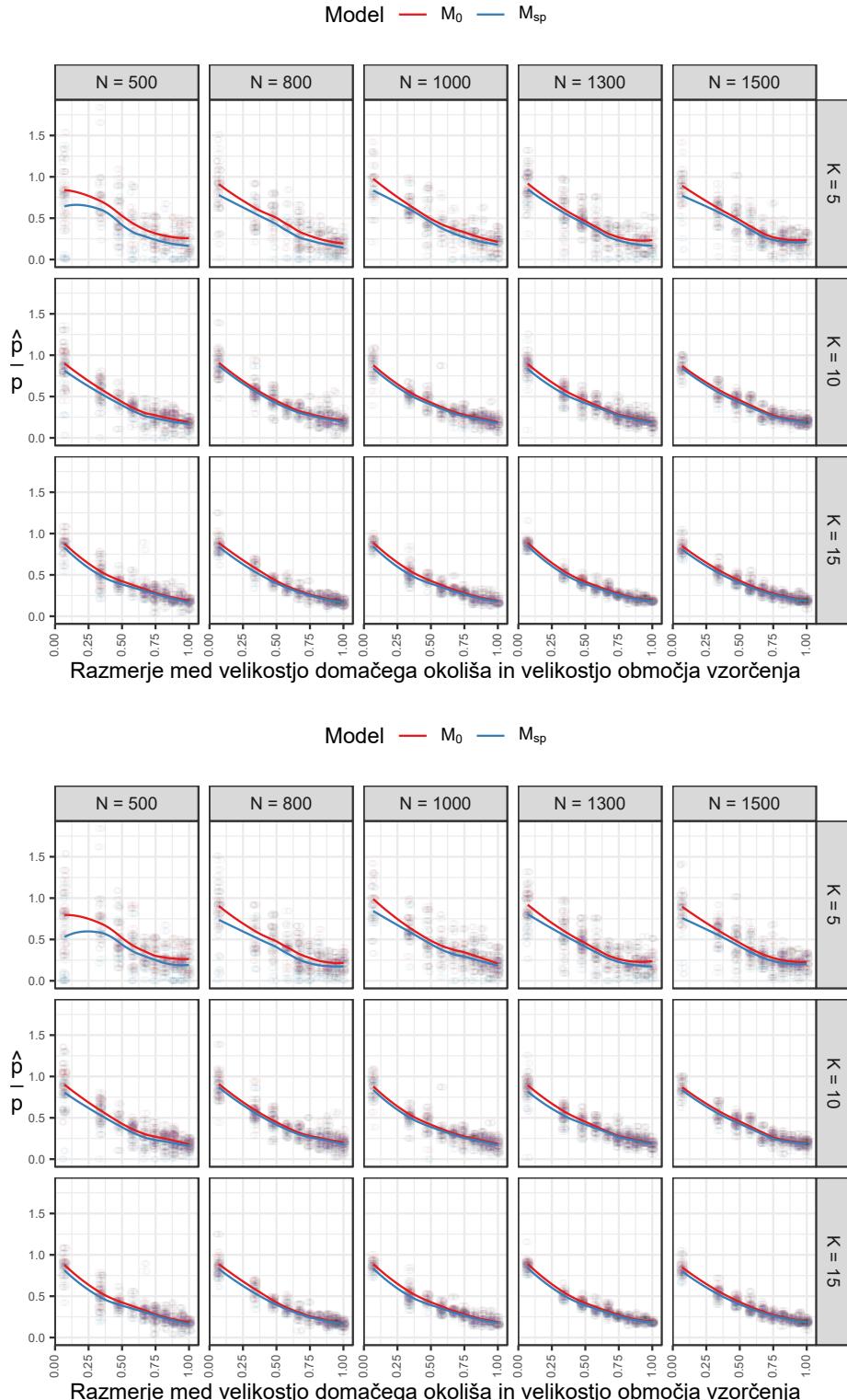
Sodeč po naših rezultatih s slike 5, ne moremo sklepati, da ima število generiranih osebkov v povezavi z razmerjem domačega okoliša in vzorčenega območja bistven vpliv na oceno ulovljivosti oz. je vpliv neznaten. Razlike med simulacijami, kjer smo spremenjali funkcijo za izračun individualne spremenljivke, so majhne oz. jih ni.



Slika 5: Vpliv razmerja med ocenjeno in pravo ulovljivostjo v povezavi z razmerjem velikosti domačega okoliša in vzorčenega območja. Zgornja slika je za simulacije, kjer smo individualno spremenljivko izračunali s pomočjo empirične porazdelitve, spodnja pa s pomočjo normalne porazdelitve.

Razmerje med površinama domačega okoliša in območja vzorčenja ima na pristransko ocene parametra  $p$  velik vpliv. Na sliki 6 je prikazano kako se pristransko ocene parametra  $p$  spreminja z razmerjem površin domačega okoliša in vzorčenega območja glede na število generiranih osebkov in odlovnih presledkov. Povprečji ocen obeh Hugginsovih modelov konvergirata z večanjem razmerja, kjer več simuliranih osebkov pomeni bolj podobno (in pristransko) oceno med modeloma. Podoben vpliv ima tudi število odlovnih presledkov, poleg tega pa le-ti vpliva tudi na variabilnost ocen. Več

kot je odlovnih presledkov, manjša je variabilnost v oceni. S slike 6 ne moremo trditi, da smo zaznali bistvene razlike med simulacijami, kjer smo za izračun individualne spremenljivke uporabili normalno ali empirično funkcijo.



Slika 6: Prikaz pristranskosti ocene ulovljivosti ( $p$ ) glede na razmerje površin domačega okoliša in območja vzorčenja v povezavi s številom simuliranih osebkov (stolpci) in odlovnih presledkov (vrstice). Na prvi sliki so prikazani rezultati simulacij, kjer smo za izračun individualne spremenljivke uporabili empirično, na spodnji pa normalno funkcijo. Na osi  $x$  je prikazano razmerje od 0 (površina domačega okoliša zelo majhna v primerjavi s površino vzorčenega območja) do 1 (površina domačega okoliša in vzorčenega območja enako velika). Pristranskost ocene parametra  $p$  narašča z večanjem razmerja velikosti domačega okoliša in območja vzorčenja. Povprečja ulovljivosti modelov po Hugginsu konvergirajo. Z večanjem števila simuliranih osebkov se manjša tudi razlika v oceni parametra. Podobno sliko vidimo z večanjem števila odlovnih presledkov (vrstice), kjer poleg konvergence povprečij modelov opazimo tudi manjšanje variance ocene.

### 3.3 GOSTOTA

Dejansko gostoto smo izračunali kot število centroidov (osebkov) znotraj površine, kjer smo jih simulirali. S pomočjo modelov Huggins in TIRM smo dobili oceno številčnosti znotraj območja vzorčenja. Ker so osebki hodili čez mejo vzorčenega območja, smo tako dejansko ocenili velikost superpopulacije katere prostorska omejenost ni nedvoumno določena. Na podlagi ocenjenega števila osebkov in velikosti vzorčenega območja smo izračunali gostoto te superpopulacije. Za izračun gostote smo uporabili tudi razširjeno območje vzorčenja. Območje smo razširili tako, da smo polmer povečali za določeno razdaljo. Razdaljo smo izbrali kot percentil parne razdalje iz funkcije, ki smo jo ali prilegli podatkom (empirična funkcija) ali pa izračunali na podlagi simuliranih parametrov (dvorazsežna normalna porazdelitev). Gostoto smo izračunali glede na območje povečano za standardni odklon polmera simuliranega območja gibanja osebka (hr), za percentil porazdelitve (50-99) oz. za najdaljšo prehojeno razdaljo (effect) v posamezni simulaciji. Spodnji rezultati so podani kot razmerje med ocenjeno in pravo gostoto. Ker gre za razmerje kjer je prava gostota v imenovalcu, se bodo v primeru nepristransko ocenjene gostote vrednosti nahajale blizu 1.

#### 3.3.1 Razlike v gostoti glede na število simuliranih osebkov in funkcij za izračun individualne spremenljivke

Na sliki 6 so prikazani podatki izračuni gostot. Na sliki uporabimo kazalnik, ki je z gostoto povezan kot razmerje med gostoto izračunano iz ocenjene velikosti populacije in simulirane velikosti populacije ( $\hat{D}/D$ ).

Opazimo razlike med simulacijami, kjer smo uporabili empirično in normalno porazdelitev za izračun individualne spremenljivke. Razlike so predvsem v primerih, kjer smo območje vzorčenja povečani za razdaljo od 50. do 99. percentila. Pristranskost kazalnika se v odvisnosti od razmerja domačega okoliša in vzorčenega območja povečuje hitreje za primere, ko smo za izračun individualne spremenljivke uporabili empirično porazdelitev.

Pristranskost kazalnika v primeru nepovečanega območje hitro narašča z večanjem razmerja med velikostjo domačega okoliša in velikostjo območja vzorčenja podobno za oba načina izračuna individualne spremenljivke.

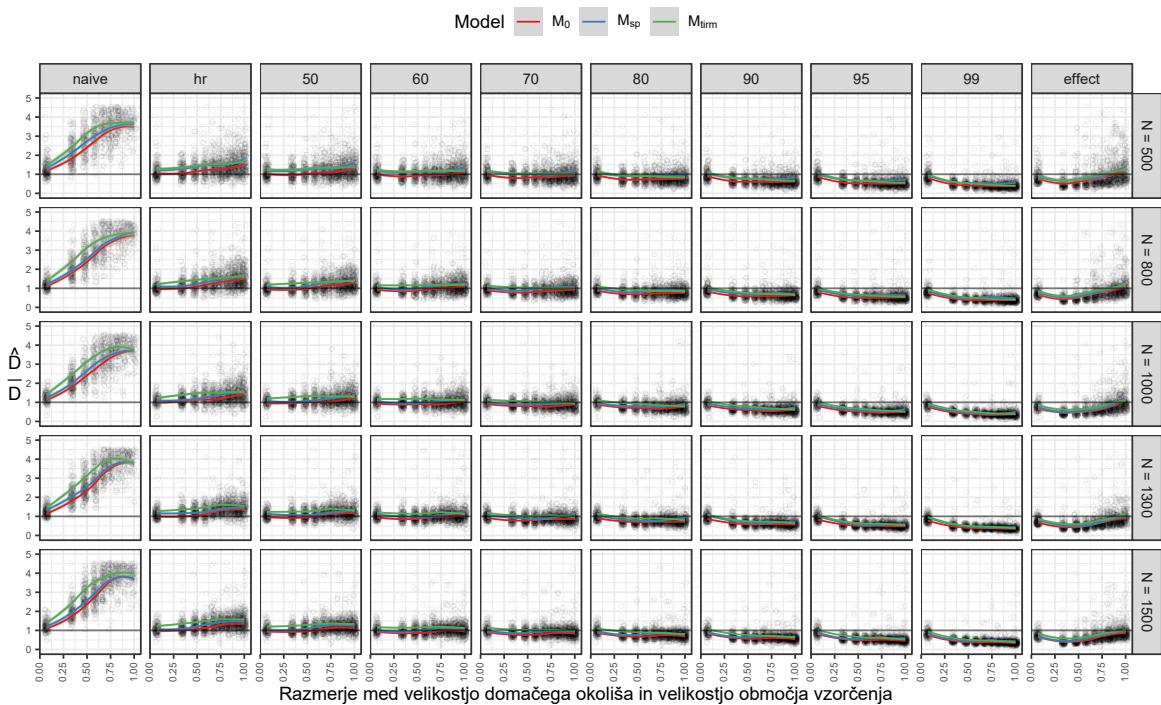
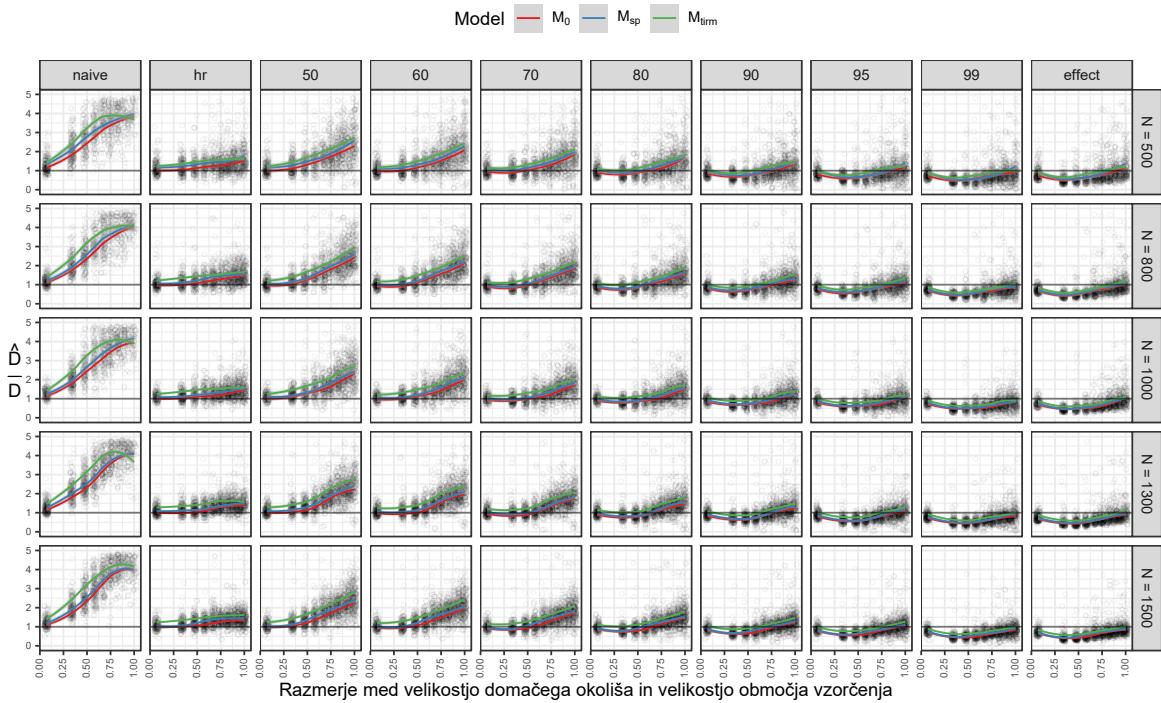
Z manjšanjem razmerja med površino domačega okoliša in območja vzorčenja se pri-

stranskost v povprečju zmanjšuje. Za primere, kjer smo za izračun individualne spremenljivke uporabili empirično funkcijo, je gladilka za večja povečanja razmerja bolj konkavne oblike, s povečevanjem površine vzorčenega območja za izračun gostote pa se pristranskost v splošnem zmanjšuje.

Pri simulacijah, kjer smo za izračun individualne spremenljivke uporabili normalno porazdelitev je konkavnosti možno opaziti le za primere, kjer smo območje vzorčenja povečali za najdaljšo prehodeno razdaljo (zadnji stolpec). Za oba načina izračuna individualnih spremenljivk pa velja, da se povečanje površine vzorčenega območja pozna najmanj za tiste simulacije, kjer je bilo razmerje med velikostjo domačega okoliša in velikostjo vzorčenega območja najmanjše.

Povečevanje števila simuliranih osebkov prispeva predvsem k zmanjšanju variabilnosti ocen velikosti populacij in posledično k manjši variabilnosti gostot. Variabilnost je za veliko večino primerov najmanjša za simulacije, kjer je razmerje med velikostjo domačega okoliša in velikostjo vzorčenega območja najmanjša.

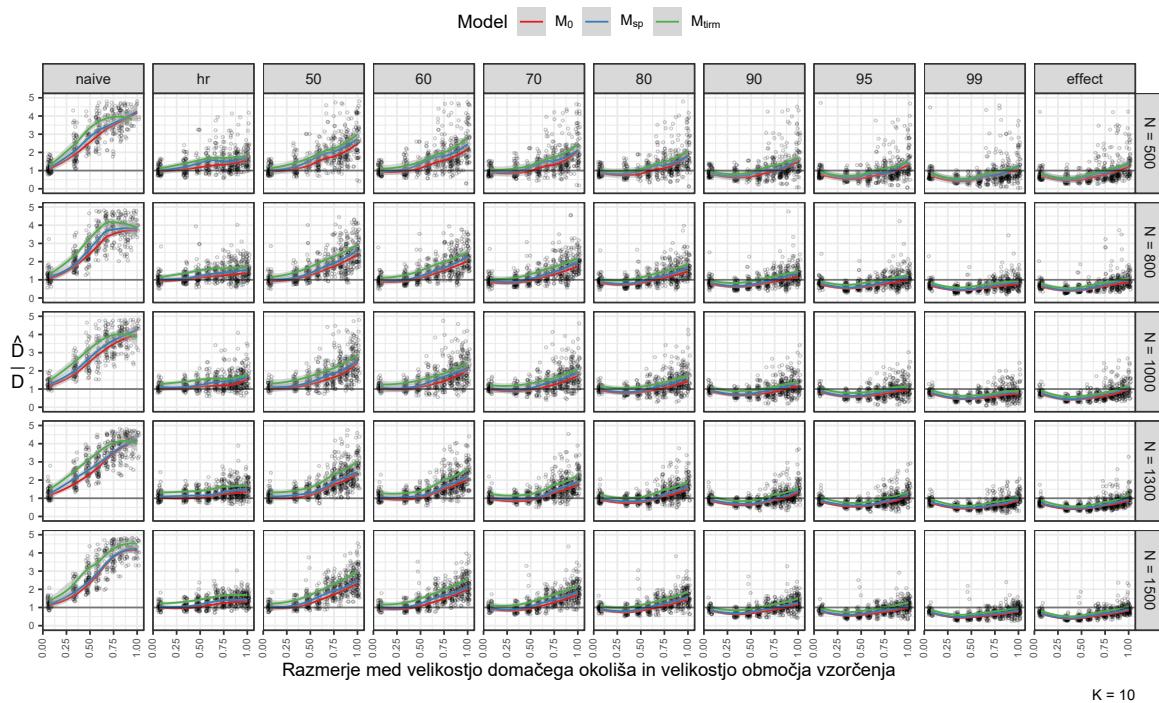
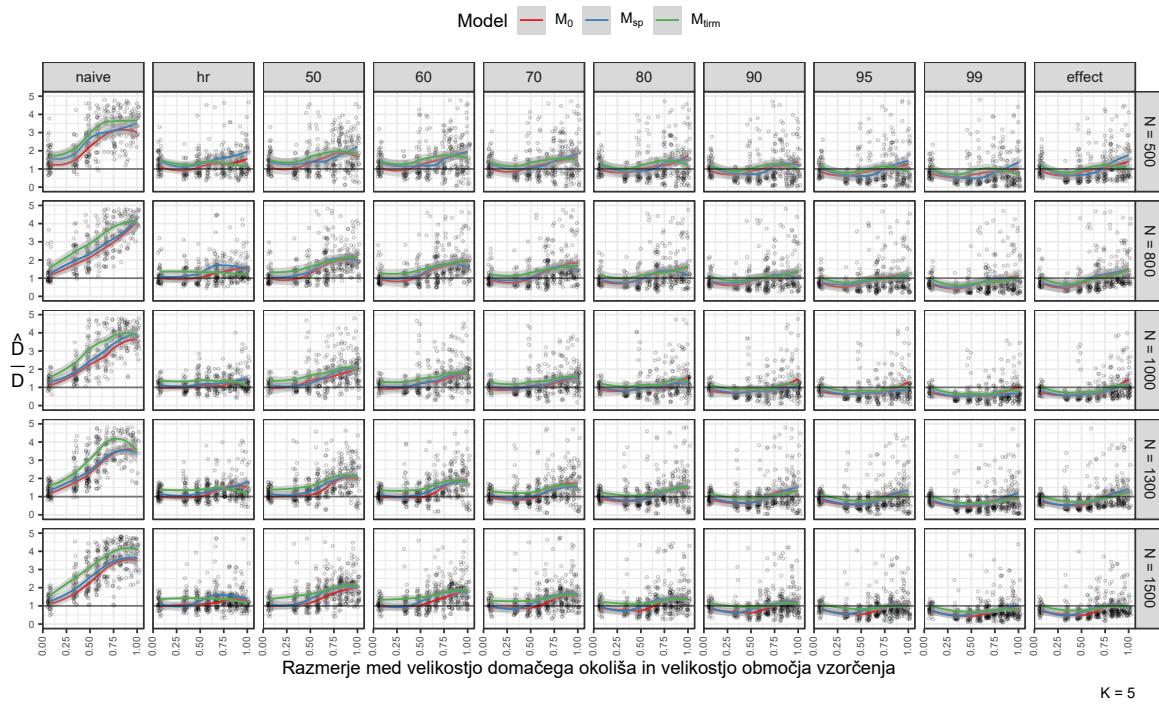
Razlike med modeli so med scenariji primerljivi. Najbolj precenjene rezultate da model TIRM, manj  $M_{sp}$  in najmanj  $M_0$ . Razlike med modeli izginejo predvsem na račun povečanja razmerja med površinama domačega okoliša in območja vzorčenja.

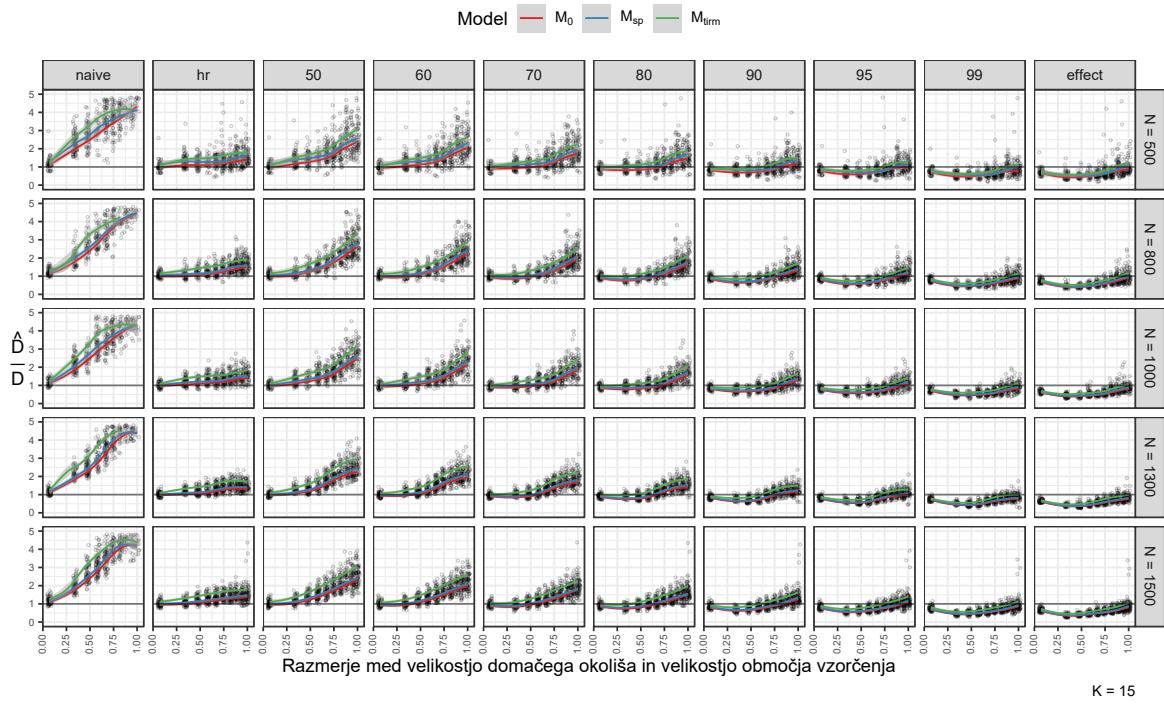


Slika 7: Prikaz kazalnika gostot ( $\hat{D}/D$ ). Na zgornji sliki so prikazani rezultati za simulacije, kjer smo za izračun individualne spremenljivke uporabili empirično porazdelitev, na spodnji pa dvorazsežno normalno porazdelitev. Stolpci oken ponazarjajo razširitev vzorčenega območja za percentil dane porazdelitve, vrstice pa število simuliranih osebkov, na podlagi katerih smo izvedli vzorčenje in posledično izračun velikosti populacije in gostote. Pri izračunu smo uporabili tri modele, Hugginsov model  $H_0$ ,  $H_{sp}$ , kjer smo ulovljivost modelirali glede na individualno spremenljivko in TIRM, ki predpostavlja dve skupini z različno ulovljivostjo.

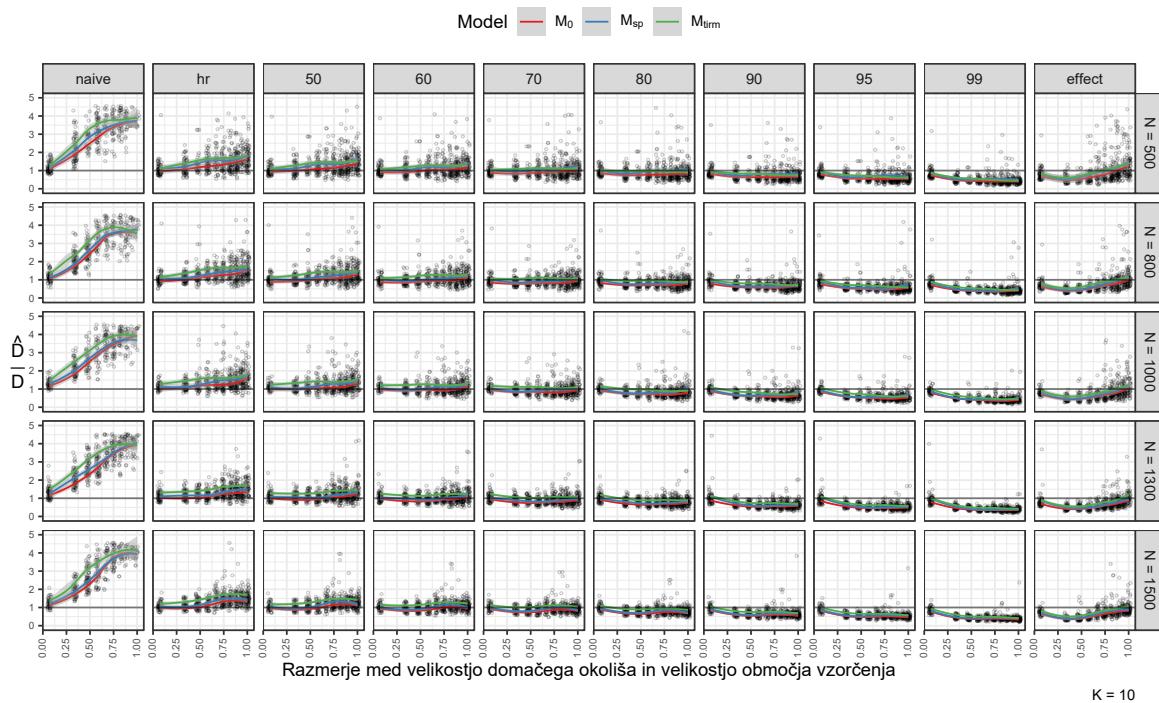
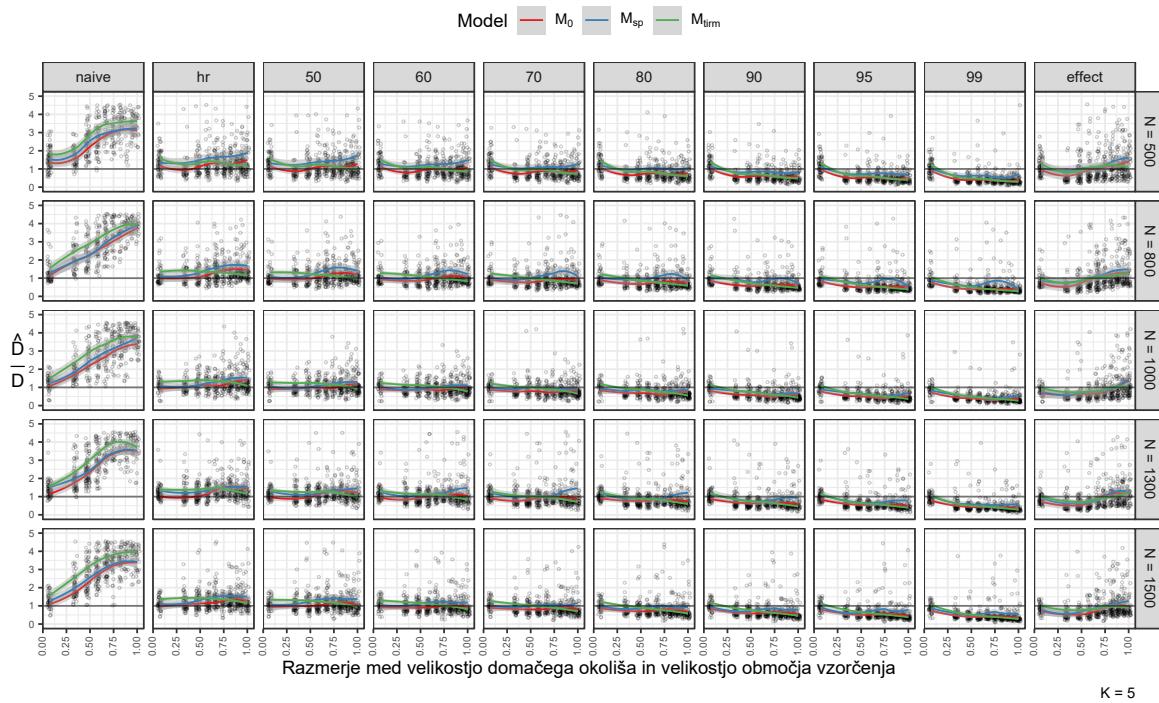
### 3.3.2 Razlike v gostoti glede na funkcijo za izračun individualne spremenljivke, števila odlovnih intervalov ter število simuliranih osebkov

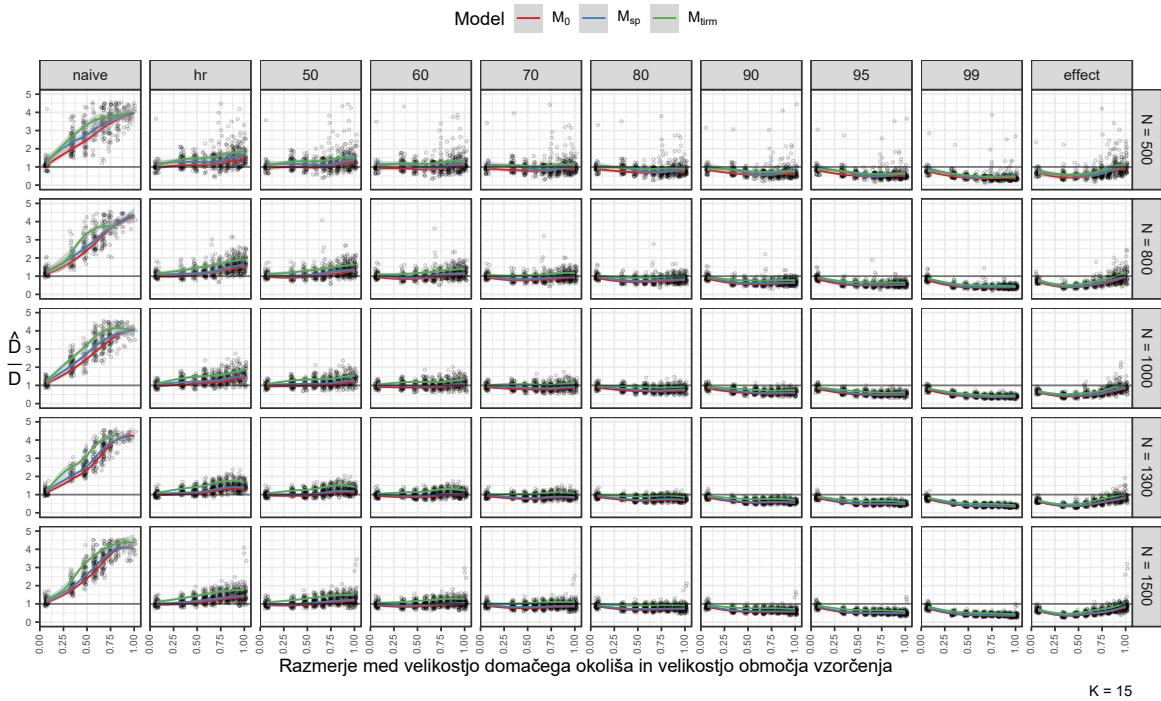
Na slikah 7 in 8 smo podatke s slike 6 prikazali tako, da so podatki razdeljeni na odlovne presledke ( $K = [5, 10, 15]$ ). Pogled glede na število odlovnih presledkov nam je dal predvsem vpogled v spremembo variabilnosti. Pri simulacijah z najnižjim  $K$  ( $K = 5$ ) je varianca večja kot pri tistih s  $K = [10, 15]$ . Pri slednjih že opazimo prevladujoč vzorec v povprečju kazalnika za vse modele glede na popravke velikosti območja vzorčenja (stolpci) in število simuliranih osebkov (vrstice). Vpliv povečanja površine vzorčenega območja se opazi predvsem za simulacije, kjer je razmerje med velikostjo domačega okoliša in velikostjo vzorčenega območja relativno velika. Na simulacije, kjer je to razmerje relativno majhno, povečanje območja vzorčenja pri pristranskosti ne kaže očitnega učinka.





Slika 8: Prikaz vrednosti simulacij, kjer smo za izračun individualne spremenljivke uporabili empirično porazdelitev. Ta slika prikazuje iste podatke, kot so na sliki 7 (zgoraj), dodatno pa smo jih razrezali glede na število odlovnih intervalov ( $K$ ). Zgoraj  $K = 5$ , v sredini  $K = 10$ , spodaj  $K = 15$ .



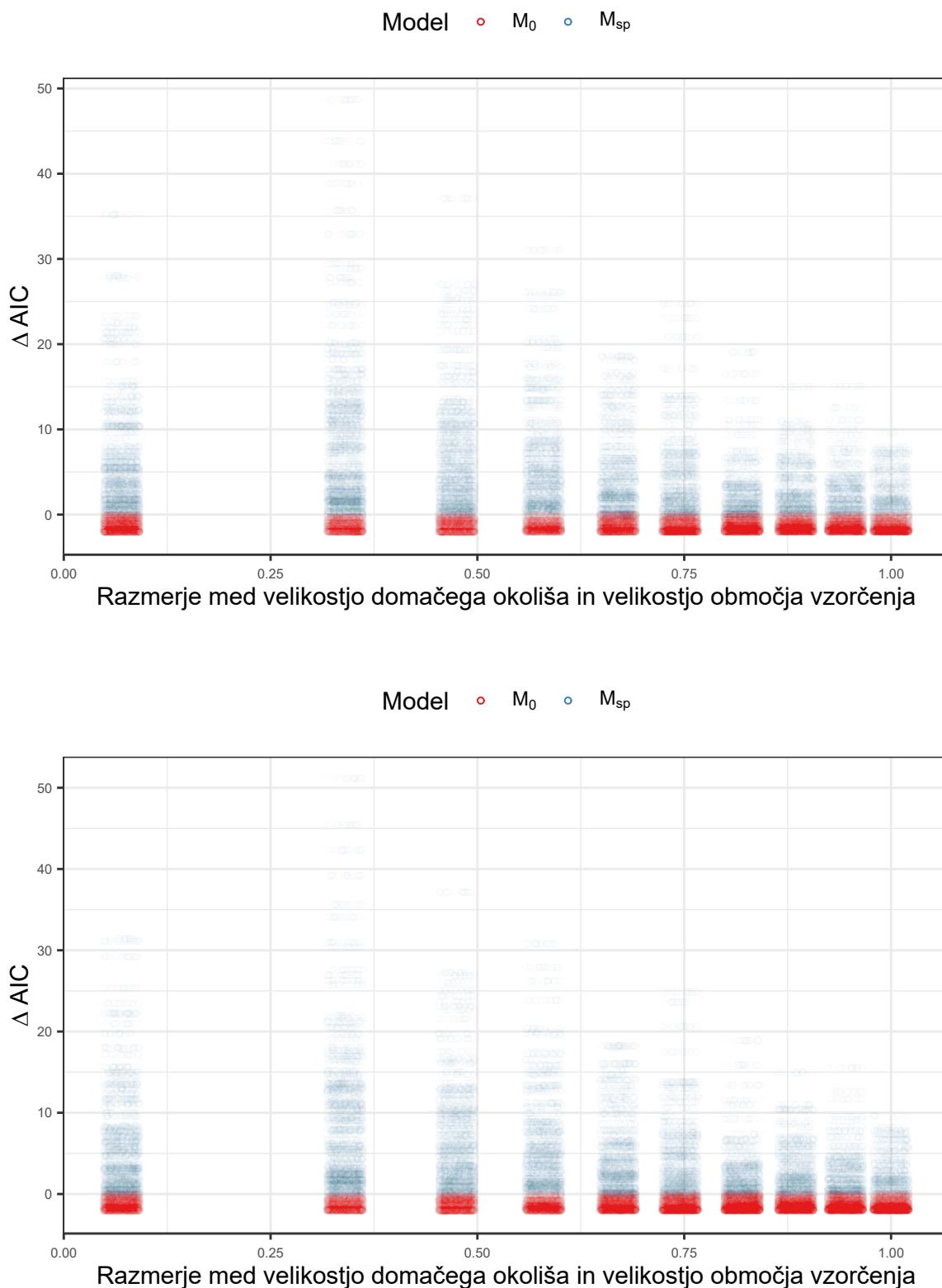


Slika 8: Prikaz vrednosti simulacij, kjer smo za izračun individualne spremenljivke uporabili normalno porazdelitev. Ta slika prikazuje iste podatke, kot so na sliki 8 (spodaj), dodatno pa smo jih razrezali glede na število odlovnih intervalov ( $K$ ). Zgoraj  $K = 5$ , v sredini  $K = 10$ , spodaj  $K = 15$ .

### 3.4 AICC

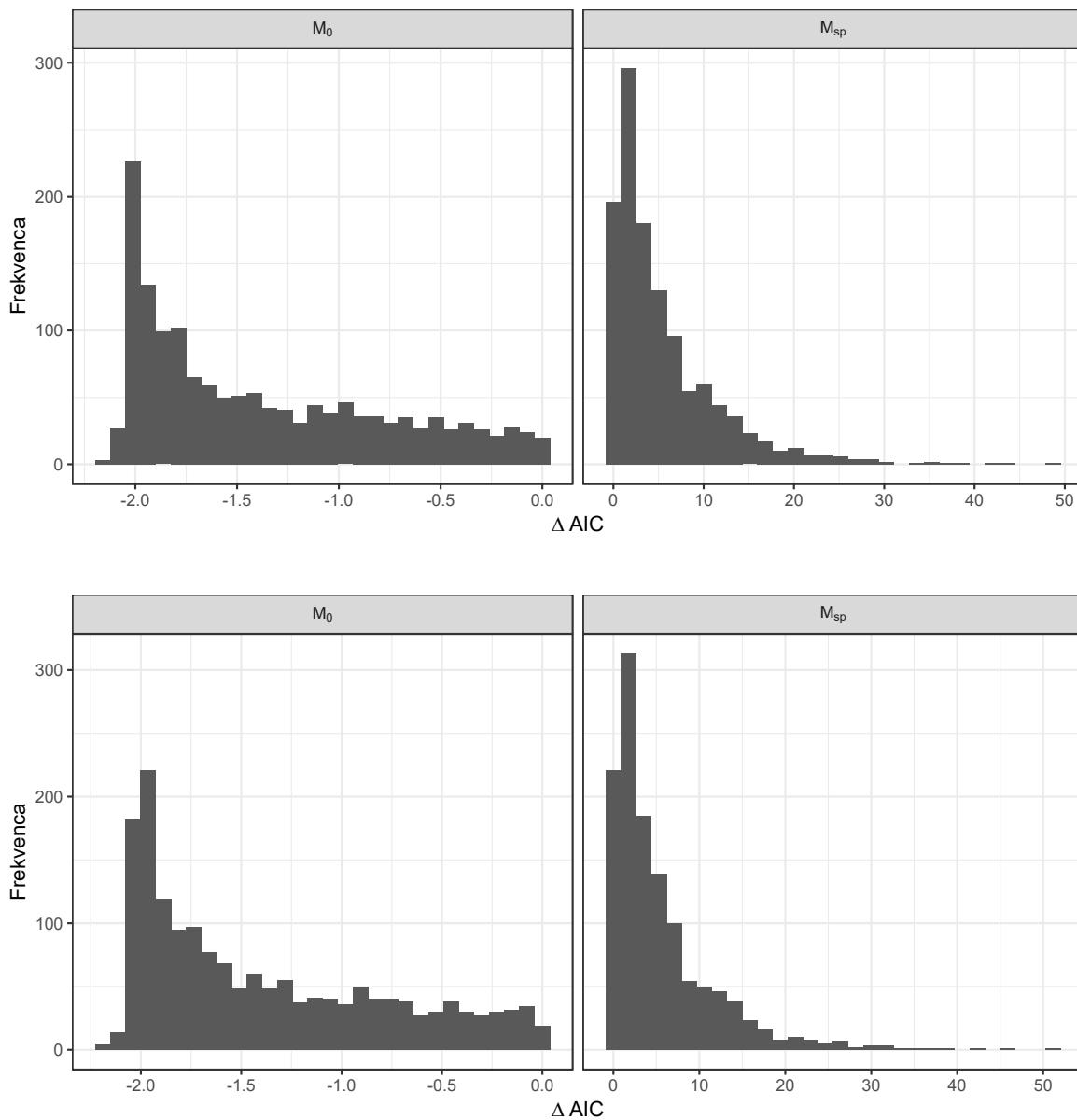
Metriko AICc smo primerjali samo za Hugginsova modela M0 in Msp. Verjetje modela TIRM ni primerljiv s Hugginsovim modelom. Poleg tega uporabi za izračun velikosti populacije drugačen set podatkov. Primerjava vrednosti AICc z ostalima dvema modeloma zato ne bi bila pravilna.

Za vsako simulacijo smo določili kateri model ima manjši AICc, nato pa razliko med modeloma prikazali na osi y. Vedno smo odšteli AICc modela M0 od AICc modela Msp. Negativne vrednosti so povezane s tistimi simulacijami, kjer je bil boljši model M0, pozitivne pa kjer je bil boljši model Msp. Če je bil boljši model Msp, je bila razlika pogosto mnogo večja (modra barva) kot v primerih, ko je bil boljši model M0 (rdeča barva). Število simuliranih osebkov ne vplivajo na razliko v AICc (ni prikazano), zato smo podatke prikazali brez te spremenljivke na sliki 10.



Slika 9: Prikaz razlike v AICc za modela  $M_0$  in  $M_{sp}$ . Na zgornji sliki so prikazani podatki, ker smo za izračun individualne spremenljivke uporabili empirično porazdelitev, na spodnji pa normalno porazdelitev. Rdeče obarvane točke predstavljajo simulacije, kjer je bil boljši model  $M_0$ , modre pa  $M_{sp}$ .

Porazdelitev vrednosti so nekoliko bolj razvidne s slike 11. Opazimo, da glede na to katero porazdelitev uporabimo za izračun individualne spremenljivke, v rezultatih ni razlik.



Slika 10: Porazdelitev razlik v AICc za modela  $M_0$  in  $M_{sp}$  s slike 10. Zgornja dva histograma predstavljata vrednosti, kjer smo za izračun individualnih spremenljivk uporabili empirično porazdelitev, spodnja dva pa normalno porazdelitev. Previdni morami biti pri razlaganju osi x, saj le-ti med modeloma nista primerljivi.

## 4 RAZPRAVA

Pomembno orodje za raziskovanje nekaterih osnovnih parametrov populacij, kot je njena velikost, so metode ulova-ponovnega ulova. Kot velja za vse metode, imajo tudi te nekatere predpostavke, ki nas do neke mere omejujejo. V tem delu smo se osredotočili na kršenje predpostavke enake ulovljivosti. Do kršenja te predpostavke pride, ker osebki prihajajo in odhajajo iz območja vzorčenja, kar imenujemo učinek roba.

S pomočjo simulacij in klasičnih modelov za zaprte populacije (po Hugginsu) smo preverili delovanje predlaganega popravka. Vsakemu osebku smo izračunali individualno spremenljivko, ki je odvisna od lege vzorcev (centroida) v prostoru in predpostavljeni funkcije gibanja. Gibanje okoli centroida smo opisali s pomočjo dveh porazdelitev. Dvorazsežno normalno porazdelitev, za katero smo uporabili simuliran standardni odklon. Ta primer predstavlja t.i. zlati standard, saj bi morala najbolje opisati gibanje osebkov. Druga porazdelitev pa je odvisna od premikov osebkov, katerim priležemo rahlo spremenjeno kumulativno porazdelitev Weibullove funkcije. Ta način ne zahteva dodatnih informacij, ki bi jih sicer lahko pridobili z dodatnimi metodami (npr. opremljanje osebkov s telemetrijskimi ovratnicami).

Opis gibanja po (pol)normalni porazdelitvi je v simulacijskih študijah relativno pogosto (Bolker, 2008; Ivan in sod., 2013), saj omogoča razumevanje gibanja in je hkrati (lahko) dovolj kompleksno, da opiše neke splošne lastnosti osebkov, kot se gibljejo v okolju ali naravi. Seveda pa je to poenostavitev, za katere ni nujno, da opiše gibanje osebkov dovolj dobro. Za bolj kompleksna gibanja raziskovalci uporabljajo realne podatke, ki jih naključno vzorčijo in simulirajo pseudopravilno gibanje osebkov (npr. ?). Tako generirani podatki nam res dajo bolj realen opis gibanja, vendar pa je možno, da dobimo zaradi vzorčenja nereprezentativne rezultate. Tak opis gibanja ni nujno prenosljiv med sezonomi ali območji.

V tem delu smo primerjali različne scenarije, kjer smo spremenjali velikost domačega okoliša, število odlovnih presledkov, število simuliranih osebkov (gostoto) in ulovljivost. Velikost in obliko vzorčenega območja smo ohranjali konstantno, ker nas v simulacijah bolj zanima relativen odnos med velikostjo domačega okoliša in velikostjo območja vzorčenja. Rezultate simulacij smo primerjali s pomočjo treh modelov. Dva sta Hugginsova modela,  $M_0$ , kjer predpostavljamo enako ulovljivost vseh osebkov in  $M_{sp}$ , kjer predpostavljamo, da se ulovljivost spreminja z individualno spremenljivko. Tretji model je TIRM, ki predpostavlja, da imajo osebki eno od dveh ulovljivosti.

#### 4.1 PRISTRANSKOST OCENE ULOVLJIVOSTI

Če bi območje vzorčenja zajemalo celotno območje na katerem se gibajo osebki, bi bili vsi ulovi znotraj območja vzorčenja. To v mnogih primerih ni realna možnost in se marsikater osebek giba v in iz območja vzorčenja. Pri lovru v simulaciji smo v primeru, da smo tak osebek ujeli zunaj območja vzorčenja, ta vzorec krnili. Brez krnjena bi bila ocenjena ulovljivost nepristranska, blizu simulirane vrednosti zaznavnosti. Ker pa smo točke, ki se niso nahajale znotraj območja vzorčenja krnili, smo to ulovljivost spremenili - jo zmanjšali. To je razvidno iz rezultatov (npr. slika 5), kjer se odmik ocenjene ulovljivosti odmakne od simulirane, predvsem na račun velikosti domačega okoliša v povezavi z velikostjo območja vzorčenja. Ko je domač okoliš relativno majhen, so te kršitve relativno majhne, a še vedno prisotne.

Ker gre za razmerje med ocenjeno in simulirano vrednostjo, so vrednosti bližje 1 bolj podobne simulirani vrednosti. Odstopanje od simulirane vrednosti je stalno glede na uporabljen model in je manjše za model  $M_0$  (bližje pravi simulirani vrednosti) kot za model  $M_{sp}$ .

Prvi model je preprost, za vse osebke predpostavi enako ulovljivost in "ne ve", da so bili podatki krnjeni. Drugi model "ve", da smo podatke krnili in do neke mere za to tudi nadoknadi s pomočjo individualne spremenljivke. Ta spremenljivka naj bi opisovala pojav zaradi česar ima posamezen osebek spremenjeno ulovljivost. V našem primeru opisuje do kakšne mere osebki niso na voljo za vzorčenje. V tem oziru kaže, da individualna spremenljivka izboljša model. Ocena se izboljša predvsem za simulacije z manj osebkami in manj odlovnimi presledki. Razlika med modeloma je komaj opazna, ko pa število osebkov in število odlovnih presledkov povečujemo, razlik med modeloma praktično ni več.

Odstopanje od simulirane vrednosti je različno. Odstopanje za 75 % ni nenavadno za tiste simulacije, kjer je bila površina domačega okoliša približno tako velika kot območje vzorčenja. S pomanjševanjem površine domačega okoliša relativno na površino območje vzorčenja (razmerje se manjša) pa se ta ocena izboljuje in v najboljšem primeru odstopa le približno 25 %. Zaradi krnenja je manj verjetno, da bi ocenjene vrednosti (v povprečju) kadarkoli dosegle teoretično vrednost 1.

Iz razmerja ocenjene in simulirane ulovljivosti lahko opazimo, da se variabilnost zmanjša, ko smo povečali število odlovnih presledkov, nekoliko pa tudi zaradi večjega števila simuliranih osebkov. To se zgodi, ko v model vključimo več podatkov. Takrat se ocene bolj natančne in je posledično oceni parametrov manj variabilna. Ta pojav nam je lahko v pomoč pri načrtovanju poskusa, kjer se odločamo, ali napor vložiti v večanje števila ulovljenih osebkov, v večanje števila odlovnih presledkov ali obojega.

Kot osnovo, s katero smo primerjali rezultate modela  $M_{sp}$ , smo vzeli rezultate modela  $M_0$ , ki o osebkih ne predpostavlja drugega, kot da so vsi enako ulovljivi. Tak model je primeren za raziskave, kjer se osebki gibljejo samo znotraj območja vzorčenja. V naravi si lahko tak primer predstavljamo na primeru ribnika, kjer ribe ne morejo zapustiti območja vzorčenja in ima vsak osebek teoretično enako verjetnost ulova. V primerjavi, model  $M_{sp}$ , ki preko individualne spremenljivke do neke mere upošteva kršenje predpostavke neprehajanja roba območja vzorčenja.

#### 4.2 PRIMERJAVA AICC ZA MODEL A $M_0$ IN $M_{sp}$

Primerjava modelov  $M_0$  in  $M_{sp}$  kaže, da ima slednji v povprečju konsistentno nižji AICc od prvega. V primerih, ko je boljši model  $M_{sp}$  je razlika v AICc med obema modeloma okoli 10. Ko je model  $M_0$  boljši, je razlika največ približno 2. Razlika v AICc 2 med dvema modeloma daje podporo tezi, da sta modela tako različna, da bi lahko tistega z nižjim AICc smatrali kot boljšega (Boulanger in McLellan, 2001). Modeli, ki so v AICc za 10 in več boljši od drugega (imajo nižji AICc) imajo močno podporo, da so boljši od alternativnega modela (? stran 71). Iz tega lahko sklepamo, da je model  $M_{sp}$  v tem pogledu v povprečju zares boljši od  $M_0$ .

Končni pokazatelj, če je ta model res pomembnejše boljši pa je izračun gostote osebkov v območju vzorčenja, kar je navadno cilj raziskav.

#### 4.3 PRIMERJAVA OCEN GOSTOT

Izračun gostote je za klasične raziskave ulova-ponovnega ulova po svoje problematičen, ker je del enačbe površina območja vzorčenja. Ocena velikosti populacije se nanaša na superpopulacijo, ki lahko prehaja rob vzorčenega območja, ni pa popolnoma znano kolikšno je to prehajanje roba. Taka ocena gostote je zato pristranska, če ne poznamo meja območja s katerega prihaja superpopulacija. Ta problem je reševal že (Dice, 1938) tako, da je območje vzorčenja razširil za polmer domačega okoliša. Na podoben način so poskušali reševati vsi raziskovalci po njem (Williams in sod., 2002).

Naš popravek spada med ti.i. ad hoc pristope, kot jih kritizira (Royle in sod., 2013)). Klasični pristopi po večini lovijo na mreži pasti in statisitke, ki jih je moč izluščiti iz takih podatkov so navadno na mnogo bolj grobi skali, ki jo določa razmak med pastmi. Za razliko od klasičnih pristopov, ki beležijo samo ulov oz. neulov, lahko iz podatkov

nabranih zvezno v prostoru izluščimo tudi točne lokacije ulovov. Ob zadostnem številu točk lahko sklepamo na velikost (in obliko) povprečnega domačega okoliša. To informacijo uporabimo, da izračunamo individualno spremenljivko, ki ponazarja verjetnost, da osebek ujamemo znotraj vzorčenega območja. Osebki, ki se nikoli ne približajo robu območja vzorčenja, imajo to vrednost 1, vsi ostali pa manj, odvisno od tega kako daleč roba se nahaja centroid domačega okoliša in kakšna je njegova oblika. Računanje gostote brez popravka Ko smo izračunali gostoto brez popravka velikosti območja vzorčenja, tako da smo oceno velikosti populacije delili s površino vzorčenega območja, je bila ocena vedno precenjena za vsa razmerja med velikostjo domačega okoliša in območja vzorčenja in vse modele ( $M_0$ ,  $M_{sp}$  in TIRM). Precenjena je na račun tega, ker metoda ulova-ponovnega ulova ocenjuje superpopulacijo (White in sod., 1982), katere osebki prehajajo območje vzorčenja.

To je tudi razlog, da se raziskovalci odločajo za popravek velikosti območja vzorčenja. Popravek je še posebej potreben v primerih, ko se s populacijo aktivno upravlja in bi jo prevelik odvzem lahko pahnil v spiralo izumiranja.

Na variabilnost ocen imata velik vpliv število odlovnih presledkov in število generiranih osebkov, ne glede na to s katerim modelom smo ocenili velikost populacije. To verjetno ni presenetljivo, saj z več odlovnimi intervali oz. osebki v model vključujemo več podatkov zaradi česar je ocena bolj natančna.

#### 4.3.1 Računanje gostote s pomočjo popravljene velikosti območja vzorčenja

Poleg naivne gostote smo izračunali tudi gostoto, kjer smo območja vzorčenja povečali za določeno razdaljo. Za izračun te razdalje smo histogramu parnih razdalj posameznih osebkov prilegli triparametrično Weibullovo porazdelitev oz. smo predpostavili normalno porazdelitev z vrednostjo standardnega odklona iz simulacije. Na podlagi teh porazdelitev smo določili razdaljo, ki sovpada s 50. do 99. percentilom prehodnih razdalj. Za dano razdaljo smo povečali premer, in s tem površino, vzorčenega območja. Območje smo razsirili še za dve statistiki, simuliran polmer domačega okoliša (hr) in za najdaljšo prehodeno razdaljo v simulaciji (effect).

Ocene gostote za posamezen model ( $M_0$ ,  $M_{sp}$  in TIRM) konsistentna glede na razmerje med površino domačega okoliša in območja vzorčenja. To pomeni, da kažejo podobna gibanja glede na spremenljivke (število odlovnih intervalov, število generiranih osebkov, razširitev območja vzorčenja) in v odnosu do drugih modelov v vseh scenarijih. Model

TIRM je najbolj precenjeval gostoto, vendar so te razlike za vse popravke glede na naivno gostoto relativno majhne.

Na sliki 6, kjer sta prikazana oba načina računanja individualne spremenljivke, lahko vidimo, da se natančnost ocene modelov spreminja glede na število generiranih osebkov (?). Ta učinek je še posebej viden na slikah 7 in 8, kjer podatke prikažemo še glede na odlovne presledke. Večje število odlovnih presledkov vpliva na natančnost ocene, ker smo v model preko odlovnih presledkov vključili več podatkov. Ta učinek je viden šele po tem, ko območje vzorčenja razširimo za katero koli mero.

Modela  $M_0$  in  $M_{sp}$  dajeta glede na simulirane spremenljivke (število odlovnih presledkov, število generiranih osebkov, razširitev območja vzorčenja) podobne rezultate in bi težko zagovarjali, da je v tem pogledu eden boljši od drugega. Razlike opazimo pri naivni gostoti, kjer TIRM model navadno ocenjuje nad Hugginsovima modeloma, le-ta ( $M_0$  in  $M_{sp}$ ) pa sta si relativno podobna. Ko območje vzorčenja razširimo, se razlike med Hugginsovima modeloma nekoliko zabrišejo.

Bolj kot povečamo površino območje vzorčenja, manjše so razlike v oceni gostote med modeloma  $M_0$  in  $M_{sp}$ . Naivno ocenjena gostota je najbolj pristranska in je v primerih, ko je razmerje med površino domačega okoliša in površino območja vzorčenja relativno veliko, precenjena tudi do 4-krat. To kaže na to, da je ocena gostote brez razmisleka o kršenju predpostavk lahko napačna, kar je pri upravljanju s prosto živečimi populacijami lahko problematično. Ker gre za precenjevanje, je potrebno odločevalcem pri podajanju rezultatov poudariti, da je to optimistična ocena.

Ko območje vzorčenja razširimo za velikost simulirane velikosti domačega okoliša (en standardni odklon od centroida glede na normalno porazdelitev) se površina s pomočjo katere računamo gostoto poveča in posledično se ocena gostote in pristranskost zmanjša. Najbolj točna ocena in relativno nepristranska za vse modele je okoli razširitve območja vzorčenja za 60. in 70. percentil. Če območje povečamo preveč, je ocena gostote lahko podcenjena. To velja za primere, ko za izračun individualne spremenljivke in posledično razdaljo za razširitev območja vzorčenja uporabimo znano, simulirano vrednost, ki opisuje gibanje osebkov in je t.i. „zlati standard“. Ko območje vzorčenja razširimo za razdaljo med 60. in 70. percentilom normalne porazdelitve, zaobjamemo velik delež območja, ki ga zasedajo osebki, ki imajo centroid gibanja blizu roba območja vzorčenja. To bi približno sovpadalo s polovico premora velikosti domačega okoliša, kot je to predlagal (Dice, 1938).

Za naš popravek smo uporabili še empirično porazdelitev, ki smo jo prilegli podatkom - histogramu razdalj med pari točk. Percentil iz posamezne porazdelitve smo uporabili

za izračun povečanja območja vzorčenja in izračun individualne spremenljivke.

Funkciji (empirična in normalna) se rahlo razlikujeta in njuni percentili niso identični. Empirična porazdelitev hitreje upade proti vrednosti 0 kot normalna. To je pričakovano, saj smo parametre za to porazdelitev ocenili iz vzorca lokacij osebka (za normalno porazdelitev smo uporabili pravo, simulirano, standardi odklon). Podoben učinek opazimo pri metodi zankanja, ki navadno dajejo manjše intervale zaupanja statistik v primerjavi s parametričnimi metodami (?). Empirična porazdelitev je tako le ocenjen približek iz podatkov. Ta ugotovitev nam služi kot varovalka, da se zavedamo, da so podatki le manjši izsek realnosti in za katere ni nujno, da odražajo pravo stanje.

Za normalno porazdelitev tako "vemo" pravo gibanje, stohastičen je samo centroid domačega okoliša. Razlike, ki jih vidimo med modeloma  $M_0$  in  $M_{sp}$  lahko pripisemo razlikam funkcij, ki jo uporabimo za računanje individualne spremenljivke in na račun odstopanj centroidov od simulirane vrednosti.

Razlike v AICc med modeloma  $M_0$  in  $M_{sp}$  so primerljive s tistimi za normalno porazdelitev. Vpliv števila odlovnih presledkov ( $K$ ) je podoben kot za normalno porazdelitev, kjer se varianca ocene gostote manjša z večanjem števila odlovnih presledkov.

Popravljanje gostote s širjenjem velikosti domačega okoliša sicer pomaga, a nikakor v taki meri kot smo to opazili za normalno porazdelitev (slika 8). Pristranskost je, tako kot za normalno porazdelitev, relativno majhna za primere, kjer je razmerje med velikostjo domačega okoliša in območja vzorčenja relativno majhno (relativno majhen učinek roba). Nepristranskost gostote se občutno zmanjša šele za primere, ko območje vzorčenja razširimo za približno 80.-90. percentil. Pri 90. percentilu in več že prihaja do podcenjevanja ocene gostote. Najbolj točni so rezultati za ekstreme simulacije - za največja in najmanjša razmerja med velikostjo domačega okoliša in območja vzorčenja. Podcenjevanje je z vidika upravljanja verjetno bolj zaželeno kot precenjevanje ker daje to zadržano oceno velikosti populacije.

## 5 SKLEPI

Ugotovili smo, da je model  $M_{sp}$  (model z dodano individualno spremenljivko) v povprečju boljši od modela, kjer te spremenljivke nismo uporabili ( $M_0$ ). Razlike v AICc so lahko relativno velike (do  $\sim 20$ ), v prid  $M_{sp}$ . V samo nekaterih primerih je bil boljši model  $M_0$ , a so relativne razlike med modeloma relativno majhne (do  $\sim 2$ ), do česar

lahko pride že zgolj po naključju.

Ucenjena ulovljivost po modelu  $M_{sp}$  je manj pristranska - nižja od simulirane, kar pomeni, da zaznamo krnenje zaradi prehajanje roba območja vzorčenja. Vendar pa so te razlike relativno majhne.

Med ocenama gostote modelov  $M_{sp}$ ,  $M_0$  in TIRM nismo zaznali praktično pomembnih razlik. Z vključitvijo individualne spremenljivke v model Msp nismo uspeli za praktične potrebe izboljšati ocene gostote.

S pomočjo simulacij smo pokazali, da je razmerje med velikostjo domačega okoliša in velikostjo območja vzorčenja pomemben dejavnik pri učinku roba. Z manjšanjem razmerja med velikostjo domačega okoliša in območja vzorčenja ( $\rightarrow 0$ ) se posledice učinka roba res zmanjšujejo. To se odraža na pristranskosti ocen parametrov ulovljivosti in velikosti populacije. Za večja razmerja ( $\rightarrow 1$ ) lahko do neke mere učinek roba popravimo s pomočjo podatkov, ki jih dobimo o velikosti domačega okoliša osebkov.

## 6 POVZETEK

## 7 VIRI IN REFERENCE

- Bivand R., Rundel C., 2017. rgeos: Interface to Geometry Engine - Open Source ('GEOS')
- Bivand R.S., Pebesma E.K., Gomez-Rubio V., 2008. Applied spatial data analysis with R. Springer, NY
- Bolker B.M., 2008. Ecological models and data in R. Princeton university press
- Boulanger J., McLellan B., 2001. Corrigendum: Closure violation in DNA-based mark-recapture estimation of grizzly bear populations. Canadian Journal of Zoology, 79(8): 1532. doi:10.1139/z01-121
- Boulanger J., Stenhouse G., Munro R., 2004. Sources of heterogeneity bias when DNA mark-recapture sampling methods are applied to grizzly bear (*Ursus arctos*) populations. Journal of Mammalogy, 85(4): 618–624. doi:10.1644/BRB-134
- Campbell T., Irvin P., Campbell K.R., Hoffmann K., Dykes M., Harding A., Johnson S., 2009. Evaluation of a new technique for marking anurans. Applied Herpetology, 6(3): 247–256. doi:10.1163/157075409X420042
- Chandler R.B., Hepinstall-Cymerman J., Merker S., Abernathy-Conners H., Cooper R.J., 2018. Characterizing spatio-temporal variation in survival and recruitment with integrated population models. The Auk, 135(3): 409–426. doi:10.1642/AUK-17-181.1
- Cooch E., White G., 2012. Program MARK: A gentle introduction (11th edition)
- Corporation M., Weston S., 2017. doParallel: Foreach Parallel Adaptor for the 'parallel' Package
- DeAngelis D.L., Grimm V., 2014. Individual-based models in ecology after four decades. F1000prime reports, 6(June): 39. doi:10.12703/P6-39
- Dice L.R., 1938. Some census methods for mammals. The Journal of Wildlife Management, 2(3): 119–130. doi:10.2307/3796432
- Dice L.R., 1941. Methods for estimating populations of animals. Journal of Wildlife Management, 5(4): 398–407
- Efford M., 2004. Density estimation in live-trapping studies. Oikos, 106(3): 598–610. doi:10.1111/j.0030-1299.2004.13043.x
- Faraway J.J., 2006. Extending the linear model with R: generalized linear, mixed effects and nonparametric regression. Taylor & Francis
- Foster R.J., Harmsen B.J., 2012. A critique of density estimation from camera-trap data. Journal of Wildlife Management, 76(2): 224–236. doi:10.1002/jwmg.275

- Hansson L., 1969. Home Range, Population Structure and Density Estimates at Removal Catches with Edge Effect. *Acta Theriologica*, 14(11): 153–160
- Hijmans R.J., Graham C.H., 2006. The ability of climate envelope models to predict the effect of climate change on species distributions. *Global Change Biology*, 12(12): 2272–2281. doi:10.1111/j.1365-2486.2006.01256.x
- Huggins R.M., 1989. On the statistical analysis of capture experiments. *Biometrika*, 76(1): 133–140. doi:10.1093/biomet/76.1.133
- Ivan J.S., White G.C., Shenk T.M., 2013. Using auxiliary telemetry information to estimate animal density from capture-recapture data. *Ecology*, 94(4): 809–816. doi:10.1890/12-0101.1
- Jugovic J., Crne M., Luznik M., 2017. Movement, demography and behaviour of a highly mobile species: A case study of the black-veined white, *Aporia crataegi* (Lepidoptera: Pieridae). *European Journal of Entomology*, 114: 113–122. doi:10.14411/eje.2017.016
- Kendall W.L., 1999. Robustness of closed capture-recapture methods to violations of the closure assumption. *Ecology*, 80(8): 2517–2525. doi:10.1890/0012-9658(1999)080[2517:ROCCRM]2.0.CO;2
- Krebs C.J., 2001. Ecology: the experimental analysis of distribution and abundance (5th ed). Benjamin Cummings (California, USA)
- Laake J., 2013. RMark: An r interface for analysis of capture-recapture data with MARK. AFSC Processed Rep. 2013-01, Alaska Fish. Sci. Cent., NOAA, Natl. Mar. Fish. Serv., Seattle, WA
- Maechler M., Rousseeuw P., Struyf A., Hubert M., Hornik K., 2018. cluster: Cluster Analysis Basics and Extensions
- Microsoft, Weston S., 2017. foreach: Provides Foreach Looping Construct for R
- Miller C.R., Joyce P., Waits L.P., 2005. A new method for estimating the size of small populations from genetic mark-recapture data. *Molecular Ecology*, 14(7): 1991–2005. doi:10.1111/j.1365-294X.2005.02577.x
- Miller S.D., White G.C., Sellers R.A., Reynolds H.V., Schoen J.W., Titus K., Barnes V.G.J., Smith R.B., Nelson R.R., Ballard W.B., Schwartz C.C., 1997. Brown and Black Bear Density Estimation in Alaska Using Radiotelemetry and Replicated Mark-Resight Techniques. *Wildlife Monographs*, No. 133(January): 1–55
- Otis D.L., Burnham K.P., White G.C., Anderson D.R., 1978. Statistical inference from capture data on closed animal populations. *Wildlife monographs*, 62: 3–135
- Parmenter R.R., Yates T.L., Anderson D.R., Burnham K.P., Dunnum J.L., Franklin A.B., Friggens M.T., Lubow B.C., Miller M., Olson G.S., Parmenter C.A.,

- Pollard J., Rexstad E., Shenk T.M., Stanley T.R., White G.C., 2003. Small-mammal density estimation: A field comparison of grid-based vs. web-based density estimators. *Ecological Monographs*, 73(1): 1–26. doi:10.1890/0012-9615(2003)073[0001:SMDEAF]2.0.CO;2
- Pennell M.W., Miller C.R., 2012. capwire: Estimates population size from non-invasive sampling
- Pike D., Crabill T., Smith R., Seigel R., Dinsmore A., 2005. Short-term effects of handling and permanently marking gopher tortoises (*Gopherus polyphemus*) on recapture rates and behavior. *Applied Herpetology*, 2(2): 139–147
- Pimm S.L., Dollar L., Bass O.L., 2006. The genetic rescue of the florida panther. *Animal Conservation*, 9(2): 115–122. doi:10.1111/j.1469-1795.2005.00010.x
- Pollock K.H., 2000. Capture-Recapture Models. *Journal of the American Statistical Association*, 95(449): 293–296. doi:10.2307/2669550
- R Core Team, 2018. R: A Language and Environment for Statistical Computing
- Rowlingson B., Diggle P., 2017. splancs: Spatial and Space-Time Point Pattern Analysis
- Royle J.A., Chandler R.B., Sollmann R., Gardner B., 2013. Spatial mark-recapture. Academic press, 1 izdaja
- Seber G., 1986. A review of estimating animal abundance. *Biometrics*, 42(2): 267–292. doi:10.2307/2531049
- Stanley T.R., Burnham K.P., 1999. A closure test for time-specific capture-recapture data. *Environmental and Ecological Statistics*, 6(2): 197–209. doi:10.1023/A:1009674322348
- Waits L.P., Paetkau D., 2005. Noninvasive genetic sampling tools for wildlife biologists: a review of applications and recommendations for accurate data collection. *Journal of Wildlife Management*, 69(4): 1419–1433. doi:10.2193/0022-541X(2005)69[1419:NGSTFW]2.0.CO;2
- White G.C., Anderson D.R., Burnham K.P., Otis D.L., 1982. Capture-recapture and removal methods for sampling closed populations. Los Alamos national laboratory
- White G.C., Burnham K.P., 1999. Program MARK: survival estimation from populations of marked animals. *Bird Study*, 46(sup001): S120–S139. doi:10.1080/00063659909477239
- White G.C., Shenk T.M., 2001. Chapter 13 - Population Estimation with Radio-Marked Animals. V: J.J. Millspaugh, J.M. Marzluff, Uredniki, Radio Tracking and Animal Populations, Academic Press, San Diego. str. 329–350. doi:10.1016/B978-012497781-5/50014-1

- Wickham H., 2009. *ggplot2: elegant graphics for data analysis*. Springer New York
- Wiewel A.S., Clark W.R., Sovada M.A., 2007. Assessing Small Mammal Abundance with Track-Tube Indices and Mark–Recapture Population Estimates. *Journal of Mammalogy*, 88(1): 250–260. doi:10.1644/06-MAMM-A-098R1.1
- Williams B.K., Nichols J.D., Conroy M.J., 2002. *Analysis and management of animal populations*. Academic Press
- Wilson K.R., Anderson D.R., 1985. Evaluation of two density estimators of small mammal population size. *Journal of Mammalogy*, 66(1): 13–21. doi:10.2307/1380951