

UNIVERZA V LJUBLJANI
BIOTEHNIŠKA FAKULTETA

Roman LUŠTRIK

**MODELI S PROSTORSKO OMEJITVIJO
ZA OCENJEVANJE GOSTOT
SUPERPOPULACIJ**

DOKTORSKA DISERTACIJA

Ljubljana, 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

Ljubljana, 2018

In god we trust, all others bring data.

– W. Edward Deming

Na podlagi Statuta Univerze v Ljubljani ter po sklepu Senata Biotehniške fakultete in sklepa 44. seje Komisije za doktorski študij Univerze v Ljubljani z dne 13. 11. 2013 (po pooblastilu Senata Univerze z dne 20. 1. 2009) je bilo potrjeno, da kandidat izpolnjuje pogoje za opravljanje doktorata znanosti na Interdisciplinarnem doktorskem študijskem programu Statistika. Za mentorja je bil imenovan prof. dr. Andrej Blejec. S sklepom Senata UL z dne 13. 1. 2018 je bil za novega mentorja imenovan doc. dr. Tomaž Skrbinšek.

Komisija za oceno in zagovor:

Predsednik: prof. dr. Janez STARE
Univerza v Ljubljani, Medicinska fakulteta, Inštitut za
biostatistiko in medicinsko informatiko
doc. dr. Martina LUŽNIK
Članica: Univerza na Primorskem, Fakulteta za matematiko, na-
ravoslovje in informacijske tehnologije, Oddelek za bio-
diverziteto
prof. dr. Andrej BLEJEC
Član: Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za
biologijo

Datum zagovora:

Roman LUŠTRIK

KLJUČNA DOKUMENTACIJSKA INFORMACIJA (KDI)

ŠD	Dd
DK	UDK 519.22:57(043.3)
KG	populacije, gostota, ocena velikosti populacije, učinek roba, metoda lova-ponovnega ulova, program MARK, program R
AV	LUŠTRIK, Roman, univ. dipl. biol.
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, 56 str., 1 pregl., 11 sl., 66 vir.
IJ	sl
JI	sl/en
AI	Eden od načinov ocenjevanja velikosti populacij je z uporabo metod lova-ponovnega ulova. Metoda predpostavlja, da je populacija zaprta (ni rojstev, smrti, priseljevanja in odseljevanja) in da imajo vsi osebki enako verjetnost ulovljivosti. Ker populacij pogosto ne moremo vzorčiti v celoti, zaradi prehoda roba območja vzorčenja prihaja do kršenja teh dveh predpostavk, kar imenujemo učinek roba. Klasični Hugginssov model za zaprte populacije za oceno parametrov sam po sebi ne omogoča uporabe prostorskih statistik, omogoča pa vključevanje individualne spremenljivke. V tem delu s pomočjo simulacij testiramo učinkovitost vključevanja individualne spremenljivke v model z namenom odpravljanja posledic učinka roba. Ugotovili smo, da je model, ki vključuje prostorsko informacijo, boljši od modela, ki te informacije ne nosi. Razlika v ocenjenem parametru verjetnosti ulovljivosti (\hat{p}) je s praktičnega vidika zelo majhna. Pristranskost ocene parametra \hat{p} je najmanjša za tiste simulacije, kjer je velikost domačega okoliša znatno manjša od velikosti območja vzorčenja, za ostale pa je močno pristranska. Pristranskost ocene parametra \hat{p} se pozna tudi pri oceni gostote, ki je zelo pristranska za primere, kjer je domač okoliš velik v primerjavi z velikostjo območja vzorčenja. Na podlagi porazdelitve za izračun individualne spremenljivke smo povečali območje vzorčenja in uspeli do neke mere popraviti gostoto, a le ob predpostavki, da imamo na voljo reprezentativno obliko in velikost domačega okoliša.

KEY WORDS DOCUMENTATION (KWD)

DN	Dd
DC	UDC 519.22:57(043.3)
CX	populations, density, population size estimate, edge effect, mark-recapture, program MARK, program R
AU	LUŠTRIK, Roman
AA	SKRBINŠEK, Tomaž (mentor)
PP	SI-1000 Ljubljana, Jamnikarjeva 101
PB	University of Ljubljana, Biotechnical Faculty, Interdisciplinary Doctoral Programme in Statistics
PY	2018
TI	SPATIALLY EXPLICIT MODELING OF SUPERPOPULATION DENSITY
DT	Doctoral dissertation
NO	X, 56 p., 1 tab., 11 fig., 66 ref.
LA	sl
AL	sl/en <p>Among methods for estimating population sizes, mark-recapture is a popular choice. It assumes population closure (void of deaths, births, immigration and emigration) and equal probability of capture. Since populations often cannot be encompassed entirely, some individuals cross in and out of the sampling area in violation of aforementioned assumptions, which is termed edge effect. The time-tested Huggins model does not in itself use spatial information to estimate parameters; however, it does enable use of an individual covariate. In this thesis, we use simulations to test whether including spatial information through an individual covariate helps alleviate edge effect. Our findings suggest that including spatial information does improve the model. For practical purposes, the difference in estimates of probability of capture (\hat{p}) between models is negligible. Bias of \hat{p} is smallest in cases where home range size is small relative to sampling area size and large for cases where home range is comparatively large. This is also evident in density estimates, which are highly biased in cases where home range is relatively large compared to sampling area. We increased the sampling area radius based on distributions used to calculate the individual covariate and managed to somewhat alleviate the bias, provided that the calculated home range shape and size are representative.</p>
AB	

KAZALO VSEBINE

KLJUČNA DOKUMENTACIJSKA INFORMACIJA (KDI)	III
KEY WORDS DOCUMENTATION (KWD)	IV
KAZALO VSEBINE	V
KAZALO PREGLEDNIC	VII
KAZALO SLIK	VIII
OKRAJŠAVE IN SIMBOLI	IX
1 UVOD S PREDSTAVITVIJO PROBLEMATIKE, CILJEV IN ZNAN-STVENIH VPRAŠANJ	1
1.1 CILJI DISERTACIJE	2
2 PREGLED OBJAV	4
2.1 UČINEK ROBA	4
2.1.1 Poskusi reševanja posledic učinka roba	4
2.2 NAŠ PRISTOP K REŠEVANJU POSLEDIC UČINKA ROBA	5
2.3 METODA LOVA-PONOVNEGA ULOVA	6
2.3.1 Osnovne značilnosti metode lova-ponovnega ulova	8
2.3.1.1 Prepoznavanje osebkov	8
2.3.1.2 Območje vzorčenja	8
2.3.1.3 Domač okoliš	9
2.3.1.4 Superpopulacija	9
2.4 PREDPOSTAVKE METOD LOVA-PONOVNEGA ULOVA	9
2.4.1 Zanesljivost označevanja	9
2.4.2 Enaka ulovljivost	10
2.4.3 Zaprtost populacije	10
2.5 CENILKA VELIKOSTI POPULACIJE ZA DVA ODLOVNA INTER-VALA	11
2.5.1 Hugginsov model za zaprte populacije	13
2.6 MODEL CAPWIRE	14
2.7 SIMULACIJE KOT ORODJE ZA RAZISKOVANJE	14
3 MATERIAL IN METODE	15
3.1 SIMULIRANJE GIBANJA OSEBKOV	15
3.2 VZORČENJE OSEBKOV	15
3.3 IZRAČUN INDIVIDUALNE SPREMENLJIVKE	16
3.3.1 Dvorazsežnostna normalna porazdelitev	16
3.3.2 Empirična porazdelitev	16
3.4 OCENJEVANJE VELIKOSTI POPULACIJ	17
3.4.1 Hugginsov model	17

3.4.2 Model CAPWIRE	17
3.5 STATISTIKE ZA PRIMERJAVO UČINKOVITOSTI DELOVANJA PO- PRAVKA	17
4 REZULTATI	20
4.1 SIMULACIJE	20
4.2 PARAMETER P MODELOV M_0 IN M_{sp} PO HUGGINSU	21
4.2.1 Primerjava simulirane in ocnjene ulovljivosti glede na porazdelitev	21
4.2.1.1 Empirična porazdelitev	21
4.2.1.2 Normalna porazdelitev	22
4.2.2 Primerjava verjetnosti ulovljivosti glede na število simu- liranih osebkov	23
4.2.3 Primerjava ulovljivosti glede na razmerje velikosti do- mačega okoliša in območja vzorčenja	25
4.3 GOSTOTA	30
4.3.1 Razlike v gostoti glede na število simuliranih osebkov in funkcij za izračun individualne spremenljivke	30
4.3.2 Razlike v gostoti glede na funkcijo za izračun individu- alne spremenljivke, število odlovnih intervalov ter število simuliranih osebkov	33
4.4 AICc	37
5 RAZPRAVA	40
5.1 PRISTRANSKOST OCENE ULOVLJIVOSTI	40
5.2 PRIMERJAVA AICC ZA MODELIA M_0 IN M_{sp}	42
5.3 PRIMERJAVA OCEN GOSTOT	42
5.3.1 Računanje gostote brez popravka	42
5.3.2 Računanje gostote s pomočjo popravljene velikosti ob- močja vzorčenja	43
6 SKLEPI	46
7 POVZETEK (SUMMARY)	47
7.1 POVZETEK	47
7.2 SUMMARY	49
8 VIRI	52
ZAHVALA	

KAZALO PREGLEDNIC

Preglednica 1: Preglednica števila simulacij, ki so dale smiseln rezultat 20

KAZALO SLIK

Slika 1: Vzorčene točke in območje vzorčenja	6
Slika 2: Funkcija simuliranja vzorčnih točk in realizirane točke.	7
Slika 3: Primerjava ocenjene ulovljivosti s simulirano	23
Slika 4: Primerjava ocenjene ulovljivosti s simulirano glede na število simuliranih osebkov	25
Slika 5: Vpliv razmerja med ocenjeno in pravo ulovljivostjo v povezavi z razmerjem velikosti domačega okoliša in vzorčenega območja.	27
Slika 6: Prikaz pristranskosti ocene ulovljivosti glede na razmerje površin domačega okoliša in območja vzorčenja v povezavi s številom simuliranih osebkov in odlovnih intervalov.	29
Slika 7: Prikaz kazalnika gostot	32
Slika 8: Prikaz ocenjene gostote (empirična porazdelitev)	35
Slika 9: Prikaz ocenjene gostote (normalna porazdelitev)	37
Slika 10: Prikaz razlike v AICc za modela M_0 in M_{sp}	38
Slika 11: Podrobnejši prikaz porazdelitev razlik v AICc za modela M_0 in M_{sp}	39

OKRAJŠAVE IN SIMBOLI

p (\hat{p}), c	Verjetnost ulovljivosti (in njena cenilka) v kontekstu Hugginsovega modela za zaprte populacije
M_0	Model, kjer smo predpostavili, da je ulovljivost enaka za prvo in vse nadaljnje ulove
M_{sp}	Model, kjer smo predpostavili heterogenost ulovljivosti. To heterogenost opišemo s pomočjo individualne spremenljivke
CAPWIRE	Model za ocenjevanje velikost populacije, ki predpostavlja različno število skupin osebkov z enakim p
TIRM	Ena oblika modela CAPWIRE
D (\hat{D})	Ocenjena gostota populacije in njena cenilka
AICc	Akaikov informacijski kriterij s popravkom za male vzorce

1 UVOD S PREDSTAVITVIJO PROBLEMATIKE, CILJEV IN ZNANSTVENIH VPRAŠANJ

Populacija je skupina osebkov iste vrste, ki naseljujejo skupni prostor v danem času. Osnovna enota populacije je osebek, ki jih združujemo v deme. Osebki znotraj dema si delijo nekatere genetske značilnosti. Meje populacij navadno niso točno določene, ampak jih uokvirimo raziskovalci iz praktičnih razlogov. Populacijo lahko opišemo s parametri, kot so rodnost, smrtnost, imigracija, emigracija in gostota oz. velikost populacije (Krebs, 2001).

Populacije proučujemo s pomočjo parametrov, na katere vpliva spekter dejavnikov, ki jih lahko delimo na biotske in abiotske. Med abiotske dejavnike razvrščamo temperaturo, vlažnost, pH ipd. Merjenje teh dejavnikov je kdaj enostavno, drugič pa mikrohabitati ekstremno spremenijo razmere in zakrijejo prave vrednosti (Cockburn in Lidicker, 1983). Zelo pomembni, a včasih spregledani, pa so tudi biotski dejavniki, h katerim štejemo druge organizme, ki lahko pripadajo isti ali drugi vrsti (Krebs, 2001). Druge vrste lahko očitno vplivajo na populacijske parametre prek npr. kompeticije ali plenjenja, interakcije med osebki iste vrste pa so včasih zabrisane in jih zaznamo šele, ko sistem pogledamo bolj podrobno (Luštrik in sod., 2011).

Za namen proučevanja lahko razdelimo populacijo na posamezne strukturne enote. Te enote lahko sestavlja osebki, ki pripadajo istemu spolu, spadajo v isti starostni razred, so v istem razvojnem stadiju, so reproduktivno aktivni ipd. Te dejavnike proučuje populacijska ekologija z namenom boljšega razumevanja vpliva dejavnikov na parameter populacije (Krebs, 2001). V tem delu bi še posebno izpostavili parameter velikosti populacije.

Velikost populacije predstavlja število osebkov oz. število potencialno reproduktivnih enot. Od velikosti populacije in drugih dejavnikov je odvisen njen dolgoročen obstoj. Če je populacija premajhna, lahko zaradi (kombinacije) parjenja v sorodstvu in stohastičnih dogodkov pride do sesutja populacije (Amos in Balmford, 2001). V drugem skrajnjem primeru, ko je velikost populacije večja od velikosti, ki jo okolje še lahko prenaša, pa pride do sesutja zaradi npr. pomanjkanja hrane v ekstremnih obdobjih (Klein, 1968). V praksi so slednji primeri prostoživečih populacij verjetno relativno redki, saj se je skozi čas vzpostavil mehanizem, ki zagotavlja, da so populacije v dinamičnem ravnovesju z drugimi vrstami in okoljem. Da lahko te pojave proučujemo ali s populacijo upravljamo, je ključnega pomena, da poznamo velikost populacije.

V literaturi je velikost populacije pogosto izpostavljena kot pomemben parameter za proučevanje (Kindberg in sod., 2011; Chandler in Andrew Royle, 2013; Jiménez in sod., 2017; Moqanaki in sod., 2018). Velikost populacije lahko določimo ali ocenimo na različne načine. Omenimo metode lova-ponovnega ulova, ki so med bolj primernimi metodami za oceno številčnosti prostoživečih vrst. Pri teh metodah vzorčimo osebke v časovnem obdobju na določenem območju in v odlovnih intervalih. Osebke zaznamo ali odlovimo, jih označimo in jih nato v naslednjem odlovnem intervalu ponovno za-

znamo ali odlovimo ter beležimo, ali so bili že ujeti ali ne. Vsi osebki, ki pridejo v stik z območjem vzorčenja in imajo takrat nezanemarljivo možnost, da jih ujamemo, predstavljajo superpopulacijo.

Področje metod lova-ponovnega ulova je široko. Za ocenjevanje velikosti populacij so zanimive tiste metode, ki zaradi matematičnih lastnosti delujejo na t. i. zaprtih populacijah. To so populacije, ki se v času bistveno ne spreminja, torej ni odseljevanja, doseljevanja, rojstev ali smrte. Ta predpostavka je lahko omejujoča, zato se glede na biološke značilnosti vrste vzorčenje prilagodi z denimo relativno kratkim obdobjem vzorčenja. V primeru vzorčenja na območju, ki ne zajame populacije v celoti, navadno prihaja do prehajanja roba območja vzorčenja v in iz območja. S tem je kršena predpostavka o enaki verjetnosti ulovljivosti osebkov in ocena parametrov je lahko pristranska. To prehajanje roba ima za posledico t. i. učinek roba in je že dolgo znan pojav brez zadovoljivih rešitev.

Dosedanje metode rešujejo posledice učinka roba tako, da analizirajo podatke iz zmanjšanega območja vzorčenja, kar naj bi dalo nepristransko oceno gostote. Drugi modeli omogočajo vključevanje individualne spremenljivke. Na podlagi poznavanja neke statistike, povezane z domačim okolišem, lahko raziskovalci povečajo območje vzorčenja v upanju, da bo ocena gostote manj pristranska. Te mere so povprečni najdaljši premik (Wilson in Anderson, 1985), razdalja od centroida vzorčenih točk do roba (Whittington in Sawaya, 2015) in druge. Kot imata pomisleke že Whittington in Sawaya (2015), so te mere pomanjkljive, ker verjetno ne opišejo v zadostni meri odnosa med domačim okolišem osebka in njegovo prisotnostjo oz. odsotnostjo v času vzorčenja.

V tem delu predlagamo novo mero, za katero menimo, da bo bolje opisala domač okoliš osebka in pripomogla k boljši oceni ulovljivosti in velikosti populacije, razširitvi območja vzorčenja in posledično boljši oceni gostote. Verjamemo, da lahko iz porazdelitve prehodenih razdalj ocenimo dvorazsežno porazdelitev, ki nam bo omogočila izračun prostorske individualne spremenljivke in posledično boljšo oceno populacijske gostote.

1.1 CILJI DISERTACIJE

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

Na podlagi razdalj med vzorčenimi točkami osebkov v območju vzorčenja bomo ocenili delež prostora/časa, 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 mo-

del 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 spremenjeno ulovljivost zaradi prehoda čez rob območja vzorčenja.

Ocena gostote bo izboljšana na račun boljšega poznavanja velikosti dejanskega prispevka območja osebkov v območje vzorčenja, saj bomo iz gibanja osebkov lahko razbrali, kako so razporejeni v prostoru.

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

2 PREGLED OBJAV

2.1 UČINEK ROBA

Prehajanje roba vzorčenega območja povzroča kršitev predpostavke o zaprtosti populacije oz. enaki verjetnosti ulovljivosti. Sicer sta to dve ločeni predpostavki, a ima kršenje katere koli za posledico različno, heterogeno, ulovljivost osebkov. Heterogenost ulovljivosti med osebki povzroči, da je ocena velikosti populacije vzorčenega območja lahko pristranska. Vpliv tega pojava smo opazili pri naknadni analizi podatkov, ki smo jih uporabili za oceno številčnosti medvedov v Sloveniji (Skrbinšek in sod., 2018), saj nekateri osebki prehajajo robove območja vzorčenja čez državno mejo Slovenija-Hrvaška. Kendall (1999) je pokazal, da je verjetnost ulovljivosti osebkov za populacije, ki prehajajo rob vzorčenega območja, $E(\hat{p}_i) = \tau_i p_i$. τ_i je verjetnost, da se osebek nahaja znotraj vzorčenega območja, pi pa ulovljivost. Ocena velikosti populacije bo nepristranska, če je $\tau_{au_i} = 1$. V primeru, da je τ_{au_i} manjši od 1, pa bo ocena pristranska na način, da bo ulovljivost nižja. To imenujemo učinek roba (?White in sod., 1982; Wilson in Anderson, 1985) in je v studijah lova-ponovnega ulova nezaželen.

2.1.1 Poskusi reševanja posledic učinka roba

Učinek roba je že dolgo proučevan problem (Efford in sod., 2004). Prvič se v literaturi s tem ukvarja Dice (1938, 1941), ki predlaga, 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 razširitev območja vzorčenja za povprečni najdaljši premik (MMDM, angl. mean maximum distance moved, Wilson in Anderson (1985)), razdaljo med pastmi, oddaljenost ulova od roba območja vzorčenja (Boulanger in sod., 2004), povprečno razdaljo premikanja med pastmi in drugi (Miller in sod., 1997; Boulanger in McLellan, 2001; Royle in sod., 2013). Ti popravki po poročanju nekaterih dajejo dobre obete za izboljšanje ocene parametrov (Whittington in Sawaya, 2015).

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. (2013b) 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 (Ψ) 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 ocenujemo Ψ , opremljeni s telemetrijskimi ovratnicami reprezentativno glede na njihovo dejansko ulovljivost.

Našteti pristopi k reševanju posledic učinka roba so ad hoc. Gre za rešitve ali poskuse

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 predstavlja vmesno pot med modeli, ki prostora ne upoštevajo, in pristopi, ki vgradijo prostorsko komponento neposredno v model.

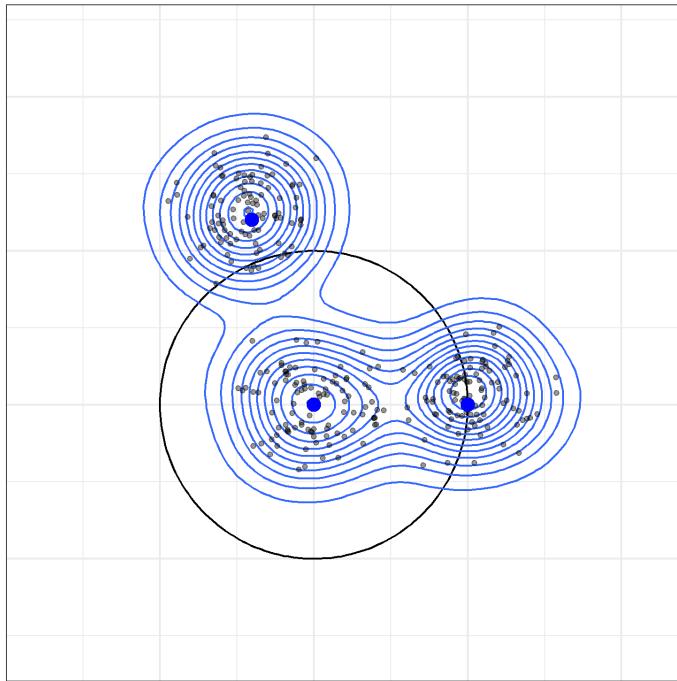
2.2 NAŠ PRISTOP K REŠEVANJU POSLEDIC UČINKA ROBA

Na podlagi preliminarnih analiz podatkov za oceno številčnosti medvedov v Sloveniji leta 2007 (Skrbinšek s sod., v recenziji) bomo v tem delu raziskali učinkovitost možnega popravka posledic učinka roba. Boulanger in McLellan (2001) v modelu za izračun velikosti populacije uporabita 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 prispevala k popravku učinka roba.

Predpostavimo, da imamo območje, kjer se nahajajo osebki in ki je praviloma večje od 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 njega pa pada (slika 2). Gibanje bi tako lahko opisali s simetrično dvorazsežnostno normalno porazdelitvijo ali polnormalno porazdelitvijo $p_{ij} = p_0 \exp(-\frac{1}{2\sigma^2} \cdot d)$ (Royle in sod., 2013, stran 127), kjer je d razdalja od centroma domačega okoliša osebka do točke ulova, σ^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 centroma.

Osebkom, ki imajo pretežni delež domačega okoliša znotraj vzorčenega območja, lahko pripisemo delež njihovega domačega okoliša znotraj območja vzorčenja 1. Če 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 ter 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



Slika 1: Vzorčene točke in območje vzorčenja. Primer pojavljanja treh osebkov in območja vzorčenja (črn krog). Modre pike predstavljajo centroide domačega okoliša, sive točke pa točke, ki so bile generirane s pomočjo dvorazsežne normalne porazdelitve. Modre črte povezujejo točke enake verjetnosti pojavljanja (gostoto) osebka.

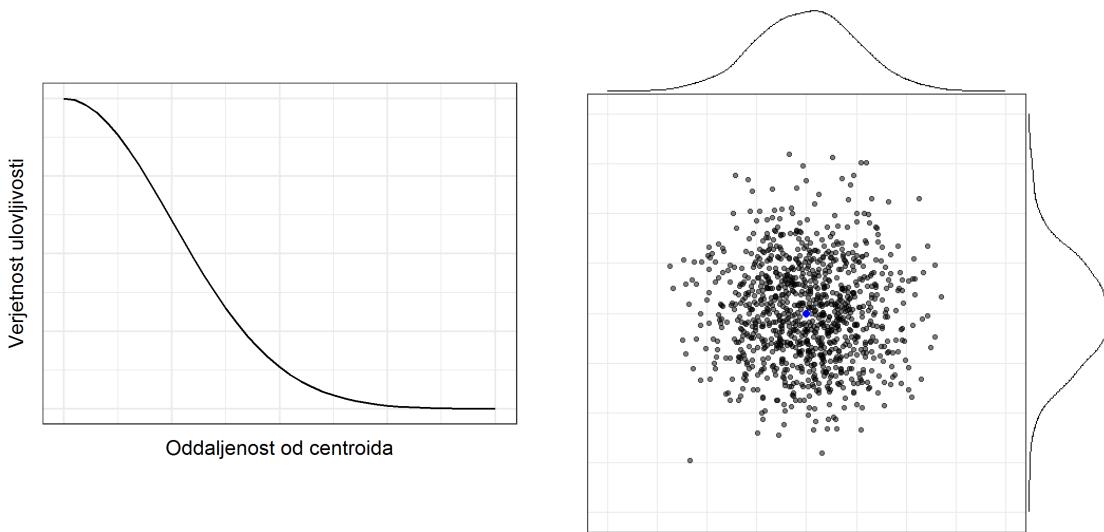
Figure 1: Sampled points and sampling area. Distribution of samples from three individuals within sampling area (black circle). Blue points represent centroid of home range, grey points are realized samples according to a two-dimensional normal distribution. Blue lines connect equal probabilities of occurrence (density) of an individual.

pripada okolišu osebka znotraj območja vzorčenja. Delež porazdelitve znotraj območja vzorčenja uporabimo v modelu kot individualno spremenljivko.

2.3 METODA LOVA-PONOVNEGA ULOVA

Velikost populacije lahko podamo na dva načina - kot število osebkov na območju vzorčenja ali neposredno kot gostoto (število osebkov na enoto). Zaradi bioloških značilnosti oziroma 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 zaznamim 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 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 cenzus, kjer predpostavimo, da vse osebke zagotovo zaznamo ($p = 1$).



Slika 2: Funkcija simuliranja vzorčnih točk in realizirane točke. Levo je ponazorjena verjetnost pojavljanja osebka glede na polnormalno porazdelitev, ki jo uporabimo za simulirane domačega okoliša. Osebek se bo najpogosteje nahajal okoli izhodišča (levo), z oddaljenostjo od njega pa bo ta verjetnost padala. Prikaz desno ponazarja "ulov" osebka (sive pike) v prostoru in realizacijo dvorazsežne normalne porazdelitve. Krivulji ob desni in zgornji stranici sta robni porazdelitvi.

Figure 2: Function used for simulation and its realization. Left depicts a half-normal distribution which we use to model home range. Highest probability of occurrence is in the center and decreases with distance from it. Right image shows realization of two-dimensional normal distribution. Density curves above and to the right of the figure margins are marginal distributions.

Cenzusi v ekologiji niso pogosti, ker določenega deleža osebkov običajno 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 ter lova-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 sklepamo o številčnosti na večjem območju. Pri tem moramo upoštevati več predpostavk, saj ima sklepanje z vzorca na večje območje več pasti (Williams in sod., 2002).

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 podatek za računanje pa je razdalja osebka do transekta. Predpostavlja se, da je na transektilih, ki jih popisovalec pregleduje, zaznavnost 1, z oddaljenostjo od transekta pa zaznavnost pada po krivulji. Tip krivulje lahko izberemo na podlagi teorije ali prakse (Williams in sod., 2002).

Do sedaj opisani načini za oceno številčnosti ne upoštevajo identitete zaznanega osebka, kar je domena metod lova-ponovnega ulova.

2.3.1 Osnovne značilnosti metode lova-ponovnega ulova

Metode lova-ponovnega ulova se od zgoraj naštetih metod razlikujejo po tem, da kot 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-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 prednosti in slabosti. V eni od delitev lahko modele razvrstimo v dve večji skupini, ki se razlikujeta po tem, da so primerni 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).

2.3.1.1 Prepoznavanje osebkov

Da lahko osebek prepoznamo kot edinstvenega, ga moramo nekako ujeti oz. zaznati. V preteklosti je to predstavljal fizičen ulov, nakar so žival nedvoumno označili. Osebke se označuje na načine, da je oznaka trajna ali začasna (npr. označevanje robinih 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...).

Z napredovanjem molekularnih tehnik in večanjem računalniške moči je danes možen "ulov" brez fizičnega stika z osebki. Tako je manj verjetno verjetno, da bo osebek spremenil vedenje in s tem svojo ulovljivost. V praksi to pomeni, da iz neinvazivno nabranih vzorcev (brez stika s preiskovanim osebkom, npr. iz sluzi, sline, dlake, iztrebka, perja) preko genotipizacije DNK z visoko zanesljivostjo določimo identiteto osebka (Waits in Paetkau, 2005). Tak genotipiziran vzorec štejemo za "ulov" osebka. Drug pristop lahko uporabimo pri živalih, ki imajo značilno obarvano kožo ali dlako (npr. trebuh nekaterih dvoživk, obarvanost kožuha pri risu), kjer lahko s pomočjo fotografij in računalniških programov razločimo med osebki.

2.3.1.2 Območje vzorčenja

Ulov in označevanje osebkov potekata na geografsko določenem območju, ki za živali z velikim območjem gibanja pogosto le deloma zaobjame proučevano populacijo. Velikost območja vzorčenja je navadno omejena zaradi velikih razdalj potovanj, s fizičnimi

mejami (otoki) ali pa s političnimi mejami (parkov, držav ali drugih administrativnih enot). Iz tega sledi, da ko osebke zaznavamo (lovimo), njihovo gibanje ni nujno omejeno na naše arbitrarno določeno območje vzorčenja.

2.3.1.3 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, bioloških značilnosti vrste, topografije, znotrajvrstnih in medvrstnih interakcij ter razpoložljivosti hrane in drugih virov, potrebnih za življenje.

2.3.1.4 Superpopulacija

Populacijo opredeljujemo kot skupino 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 arbitrarnim 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.

2.4 PREDPOSTAVKE METOD LOVA-PONOVNEGA ULOVA

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

2.4.1 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 vsaj za nekatere skupine živali (npr. večji sesalci) uporablja pristope, temelječe na molekularni genetiki. Ker za nabiranje nekaterih tipov vzorcev (iztrebkov, dlak, peres) ne potrebujemo neposrednega stika z osebkom in pri nabiranju ne motimo vedenja, to imenujemo neinvazivno vzorčenje. Pri genetskem vzorcu je zaradi napak genotipizacije ali identičnih dvojčkov mogoče, da genotip določimo narobe, vendar so metode, če so pravilno uporabljene, kljub temu zelo zanesljive

v smislu označevanja osebkov (Waits in Paetkau, 2005). Za živali, ki imajo značilne vzorce kožuha, se lahko uporablja posnetke s samodejnih foto pasti, ki so doživele razmah uporabe v zadnjih desetih letih (O'Connell in sod., 2010).

2.4.2 Enaka ulovljivost

Predpostavlja se, da imajo vse živali znotraj območja vzorčenja enako verjetnost, da jih bomo zaznali oz. ulovili v danem odlovnem intervalu. Do neenake ulovljivosti lahko pride zaradi več vzrokov. Če imajo živali različne ulovljivosti, govorimo o heterogenosti ulovljivosti. Zvišanje (t. i. *trap-happy*) ali znižanje ulovljivosti (t. i. *trap-shy*) je posledica pristranskega vzorčenja, bodisi zaradi opazovalca ali proučevanega subjekta (Roche in sod., 2013). Do heterogenosti ulovljivosti lahko pride tudi v primerih, ko območje vzorčenja geografsko ni zaprto, osebki pa prehajajo iz in v območje vzorčenja. 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).

Heterogenost ulovljivosti lahko testiramo s pomočjo informacijsko-teoretičnega pristopa (Burnham in Anderson, 2002). Podatke analiziramo s pomočjo različnih parametrizacij modela. V osnovnem modelu predpostavimo enako ulovljivost, v drugih pa upoštevamo heterogenost ulovljivosti, ki jo sumimo (npr. na račun privabljanja ali odbijanja od pasti, prehajanja roba območja vzorčenja). Modele sortiramo po velikosti statistike AICc. Če je zaznati heterogenost ulovljivosti, bo imel osnovni model, ki predpostavlja enako ulovljivost, znatno višjo vrednost AICc od ostalih. Ista avtorja na strani 70 podajata okvirne vrednosti razlik med modeli. Razlike v razponu med 0 in 2 so v zmerno podporo modelu z nižjo vrednostjo AICc, od 4 do 7 v znatno podporo, razlike nad 10 pa lahko smatramo kot ključno podporo v prid modela z nižjim AICc.

2.4.3 Zaprtost populacije

Demografsko zaprta populacija je tista, pri kateri v času vzorčenja ne prihaja do sprememb v smislu emigriranja, imigriranja, rojstev ali smrti. Populacija se torej strukturno ne spreminja. Z nekaterimi metodološkimi pristopi lahko vsaj za nekatere dolgožive vrste zagotovimo, da je kršenje predpostavke v smislu rojstev ali smrti minimalno. V prispevku, kjer smo ocenili velikost populacije rjavega medveda v Sloveniji, smo izbrali obdobje vzorčenja jesen, ko smo prepričani, da ne prihaja do poleganja mladičev (Skrbinšek s sod. (v recenziji)). Osebke, ki so dokumentirano izgubljeni iz populacije v času vzorčenja, lahko naknadno prištejemo končni oceni. Ker imajo nekatere vrste relativno velika območja gibanja, je zagotavljanje vzorčenja na dovolj velikem območju težavno. V takih primerih lahko upoštevamo heterogenost ulovljivosti in uporabimo, če je možno, več virov informacij (Boulanger in sod., 2008).

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 prehajanje lahko zmanjša ulovljivost do te mere, da je parametre modela težko oceniti oz. so nesmiselni (npr. interval zaupanja od $-\infty$ do ∞). Pri organizmih, ki imajo že zaradi svojih bioloških značilnosti nizko zaznavnost (npr. ris), je to lahko za raziskavo pogubno.

2.5 CENILKA VELIKOSTI POPULACIJE ZA DVA ODLOVNA INTERVALA

Cenilko za velikost populacije, ko vzorčimo v dveh odlovnih intervalih, imenujemo Lincoln-Petersenova cenilka. Ime je dobila po Lincolnu in Petersenu, ki sta jo uporabila pri svojem delu, vendar nista avtorja (Williams in sod., 2002). Pred njima jo je uporabil že vsaj Laplace, ko je v 18. stoletju ocenil število prebivalcev Pariza. Lincoln-Petersonovo cenilko je mogoče izpeljati na več načinov, spodaj pa bomo predstavili le enega, 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, stolpci pa odlovne intervale. V primeru ulova in prepoznavanje osebka v določenem odlovnem intervalu je vrednost matrike na tem mestu 1, sicer pa 0. Spodnja matrika prikazuje tri osebke, ki so bili ujeti v dveh odlovnih intervalih. Prvi osebek je bil ujet (in označen) samo v prvem intervalu, drugi osebek samo v zadnjem, tretji osebek pa je bil zabeležen v obeh odlovnih intervalih.

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

Lincoln-Petersenova cenilka (Williams in sod., 2002) je prizjena 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 odlovnna intervala, je verjetnost ulovljivosti p^* produkt verjetnosti ulovljivosti v prvem (p_1) in drugem (p_2) odlovnem intervalu z naslednjo 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}

$$\begin{aligned} \hat{N} &= \frac{r}{\hat{p}^*} = \frac{(x_{11} + x_{10})(x_{11} + x_{01})}{x_{11}} = \frac{n_1 n_2}{m_2} \\ \hat{N} &= \frac{(n_1 + 1)(n_2 + 1)}{m_2 + 1} - 1. \end{aligned}$$

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).

Za modeliranje heterogenosti 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 $\pi_j = \int_0^1 \frac{K!}{(K-j)!j!} p^j (1-p)^{K-j} dF(p)$.

2.5.1 Hugginsov model za zaprte populacije

Primer za dva odlovna intervala zgoraj lahko posplošimo na poljubno število. Eno takih posplošitev najdemo v modelu Huggins-Alho (znan tudi kot Hugginsov model, kot ga imenujemo tudi v tem delu, Huggins (1989)). Ta model omogoča, da ocenimo velikost populacije, pri tem pa lahko upoštevamo tudi razlike v verjetnosti ulovljivosti (M_h) s pomočjo individualne spremenljivke. Pogojna verjetnostna funkcija je

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

Ključni dodatek je ocena ulovljivosti z uporabo spremenljivk na nivoju posameznika za vsak odlovni interval s pomočjo že prej omenjene logistične funkcije

$$p_{ij} = \frac{e^{b_0 + \beta_1 z_i \beta_2 x_j \alpha z_{ij}}}{1 + e^{b_0 + \beta_1 z_i \beta_2 x_j \alpha z_{ij}}},$$

kjer je z_i vektor individualnih spremenljivk (npr. delež časa, ki ga osebek prezivi znotraj območja vzorčenja), x_j vektor okoljskih spremenljivk, kot je denimo količina padavin, z_{ij} pa predstavlja spremenljivo, ki spremeni ulovljivost v času. Tako lahko s pomočjo individualne spremenljivke upoštevamo, da imajo osebki različno ulovljivost. Akanda in Alpizar-Jara (2014) sta za izračun p_{ij} uporabila še GLM, GEE in GLMM.

Hugginsov model, za razliko od npr. Pollock in sod. (1984), ne vsebuje parametra za velikost populacije \hat{N} , zato lahko p_i ocenimo s pomočjo zveznih spremenljivk. Cenilka za velikost populacije je

$$\hat{N}(\theta) = \sum_{i=1}^N \frac{1}{p_i(\theta)},$$

kjer predpostavimo, da je θ že znana. Parameter p_i definiramo kot verjetnost, da smo osebek ulovili vsaj v enem odlovnem intervalu

$$p_i(\theta) = P(C_i) = 1 - \prod_{j=1}^K (1 - \hat{p}_{ij}).$$

Nepristranska cenilka variance ($\hat{\nu}(\theta)$) je

$$s^2 = \sum_{i=1}^N p_i^{-2}(\theta)(1 - p_i(\theta)).$$

2.6 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 intervalov ali koliko vzorcev je bilo nabranih v posameznem intervalu. 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 pa predpostavlja, da obstajata dve skupini osebkov, ki imata dve različni ulovljivosti. Skupino modelov, ki predpostavljajo več skupin, imenujemo modeli mešane ulovljivosti.

2.7 SIMULACIJE KOT ORODJE ZA RAZISKOVANJE

V matematičnih modelih opišemo pojav s funkcijami, ki so lahko preproste ali kompleksne. S kompleksnostjo pojava 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 poznavanja pojava. Podenota ni samostojna, ampak je odvisna od drugih podenot in parametrov, ki jih določi raziskovalec. V ekologiji lahko na tak način simuliramo gibanje in obnašanje živali ter proučujemo interakcijo med osebki oz. višjimi enotami in okoljem (Bolker, 2008).

3 MATERIAL 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, kar je še posebej uporabno pri proučevanju vpliva variabilnosti na raziskovani pojav.

3.1 SIMULIRANJE GIBANJA OSEBKOV

Simulacijo smo postavili v svet A, ki je tako velik, da noben simuliran osebek blizu območja vzorčenja, ne bo imel praktične možnosti, da bi dosegel rob sveta. 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 (krog s polmerom 1000) razporedili naključno s pomočjo enakomerne porazdelitve. Gibanje osebkov smo opisali z dvorazsežnostno normalno porazdelitvijo. V simulacijah smo generirali 500, 800, 1000, 1300 in 1500 osebkov. Gostota je znašala 0.00016, 0.00025, 0.00032, 0.00041 in 0.00048 za pripadajoče število simuliranih osebkov. V območju vzorčenja to v povprečju znese 20, 30, 40, 50 in 60 osebkov na območju vzorčenja. To je tudi število, ki ga prikazujemo kot simulirano število osebkov v rezultatih. Zaradi računske zahtevnosti večjega števila osebkov nismo simulirali, ker bi po naših ocenah izračun trajal nesorazmerno dolgo (približno 30 dni).

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

3.2 VZORČENJE OSEBKOV

Osebke smo vzorčili s ponavljanjem v K odlovnih intervalih (5, 10, 15), kjer smo v posameznem intervalu 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 lova-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.

3.3 IZRAČUN INDIVIDUALNE SPREMENLJIVKE

Iz praktičnih razlogov smo izračune 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 se šteli. Ta vrednost je povezana s časom, ki ga osebek porabi znotraj območja vzorčenja, oziroma s časom, v katerem je na voljo za odlov.

3.3.1 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.

3.3.2 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, je predstavljal utež razreda. Uteženemu histogramu smo prilegли in ocenili funkcijo s tremi parametri, ki spominja na kumulativno porazdelitveno funkcijo 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 opisuje 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 dolžin prehojenih razdalj.

3.4 OCENJEVANJE VELIKOSTI POPULACIJ

Velikost populacije smo ocenili s pomočjo dveh modelov, modela po Hugginsu in modela CAPWIRE.

Parametre smo ocenili s pomočjo programa MARK (Cooch in White, 2012), ki smo ga krmili preko R paketa RMark (Laake, 2013). Nato smo izluščili parametra ulovljivost in velikost ocenjene populacije ter pripadajoče intervale zaupanja.

3.4.1 Hugginsov model

Z vzorčenjem smo za vsak osebek ustvarili odlovnno zgodovino z individualno spremenljivko v K odlovnih intervalih. 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. To smo izračunali na dva različna načina (glej poglavje Izračun individualne spremenljivke). Ta model bo predvidoma upošteval, da obstaja učinek roba, in ga označujemo kot M_{sp} .

3.4.2 Model CAPWIRE

Pri modeliranju s paketom CAPWIRE (Pennell in Miller, 2012) smo uporabili model TIRM, ki predpostavlja, da gre za mešanico dveh skupin osebkov z različnima verjetnostima ulovljivosti. Odlovne zgodovine smo sešteli po osebku in število ulovov uporabili v modelu. Iz rezultata smo izluščili velikost populacije.

3.5 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 intervalov, 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 za modela M_0 in M_{sp} , ocenjeno velikost populacije, AICc ter razliko v AICc 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 (σ), za empirično porazdelitev pa trije parametri (λ, k, a).

Za vsako simulacijo smo izračunali naslednje statistike:

- Gostota osebkov brez popravka, t. i. "naivna gostota". Izračunali smo jo z deljenjem ocene velikosti populacije (\hat{N}) z velikostjo območja vzorčenja.
- Gostota 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 standardnim odklonom 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 oziroma je relativno zapletena. S pomočjo simulacijskega pristopa in z 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 smo povečali 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 AICc razlikuje po tem, da je bolj primeren za manjše vzorce. Model z nižjo vrednostjo AICc, z razliko vsaj 2, bomo smatrali kot boljšega 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 ($I = \hat{D}/D$).

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

Simulacije smo spisali v programskem okolju R (R Core Team, 2018). Pri tem smo 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, 2017b) in **doParallel** (Microsoft in Weston, 2017a). Za vse grafične prikaze smo uporabili paket **ggplot2** (Wickham, 2009). Teza je spisana z orodjem **L^AT_EX**.

Simulacije smo poganjali na delovni postaji na Oddelku za biologijo Katedre 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, izračuni pa so potekali v virtualnem sistemu Ubuntu. Vsa koda simulacij in analize, skupaj s slikami, je dostopna na GitHub odložišču <https://github.com/romunov/duhturat>.

4 REZULTATI

4.1 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). S pomočjo programa MARK smo uspešno ocenili parametre 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 (Preglednica 1).

Da parametrov nekaterih simulacij 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 lova-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 funkciji. Število uspešnih simulacij glede na število generiranih osebkov in število odlovnih intervalov se med skupinama razlikuje. Večina neuspelih simulacij pri uporabi empirične funkcije je za najmanjše število odlovnih intervalov ($K = 5$).

Preglednica 1: Preglednica števila simulacij, ki so dale smiseln rezultat. Prikaz števila simulacij, kjer smo uspešno ocenili parametre, glede na funkcijo, ki je bila uporabljena za izračun individualne spremenljivke. V vrsticah je povprečno število generiranih osebkov znotraj območja vzorčenja (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 preglednicama tako, da smo empirični (levo) odsteli normalno (desno) preglednico. Pretežni delež neuspelih simulacij je predvsem na račun simulacij, ki imajo 5 odlovnih intervalov in/ali manjše število simuliranih osebkov (nižja gostota).

Table 1: Number of simulations for which model parameters have been produced based on function used to calculate individual covariate. Rows represent average number of individuals generated within the sampling polygon (N), columns represent number of sampling sessions (K). For the empirical function (left), number of successfully estimated simulations is slightly lower. Simulations with lower numbers of sampling sessions and simulated individuals were less likely to produce model estimates. Table on the right shows difference between left and center table.

Empirična funkcija			Normalna funkcija			Razlike med matrikama		
	$K = 5$	$k = 10$	$K = 15$		$K = 5$	$k = 10$	$K = 15$	
$N = 20$	142	187	190	$N = 20$	188	209	197	$N = 20$
$N = 30$	141	192	178	$N = 30$	179	205	183	$N = 30$
$N = 40$	135	182	185	$N = 40$	175	190	185	$N = 40$
$N = 50$	168	212	194	$N = 50$	191	212	195	$N = 50$
$N = 60$	171	186	218	$N = 60$	199	190	218	$N = 60$

4.2 PARAMETER P MODELOV M_0 IN M_{sp} PO HUGGINSU

Za modela M_0 in M_{sp} smo predpostavili, da sta parametra p in c enaka. Ulovljivost v prvem in vseh ostalih odlovnih intervalih 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 ni neposredno primerljivo s parametrom p po Hugginsovem modelu, zato 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).

4.2.1 Primerjava simulirane in ocnjene ulovljivosti glede na porazdelitev

4.2.1.1 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.varp.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
```

4.2.1.2 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

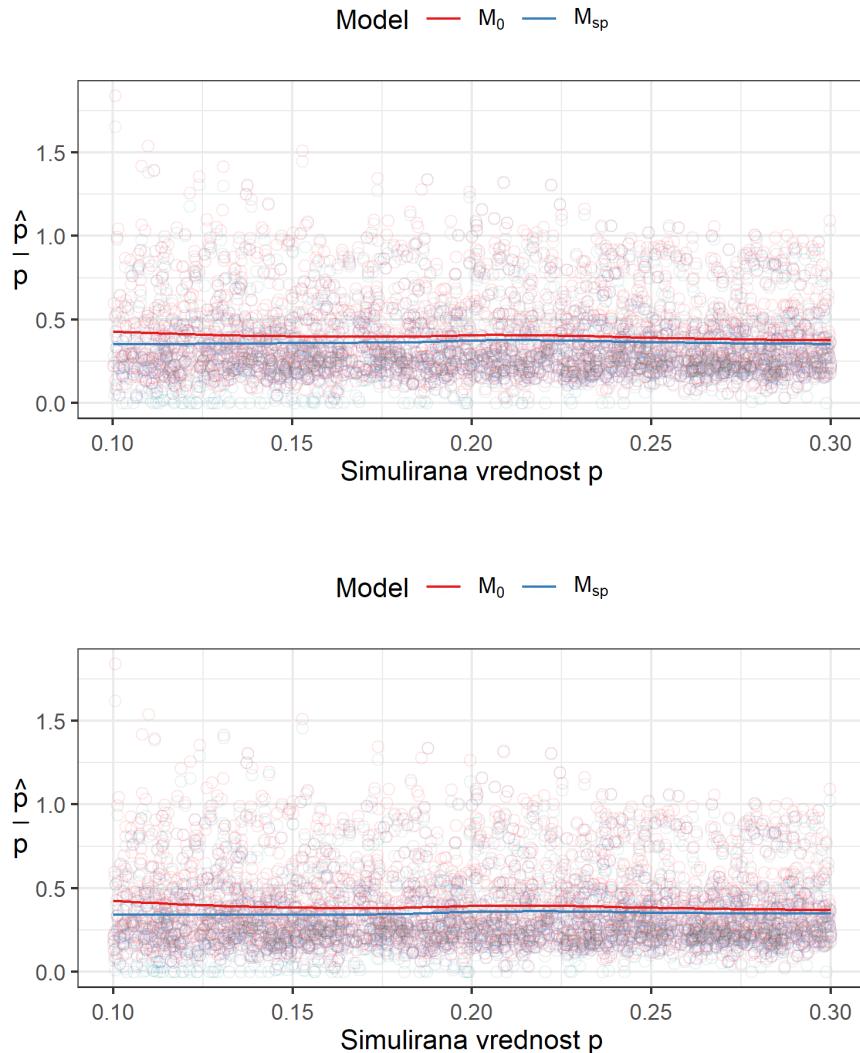
Coefficients:

	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: Primerjava ocenjene ulovljivosti s simulirano. Povprečna ocena ulovljivosti (\hat{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 verjetnosti ulovljivosti. 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 .

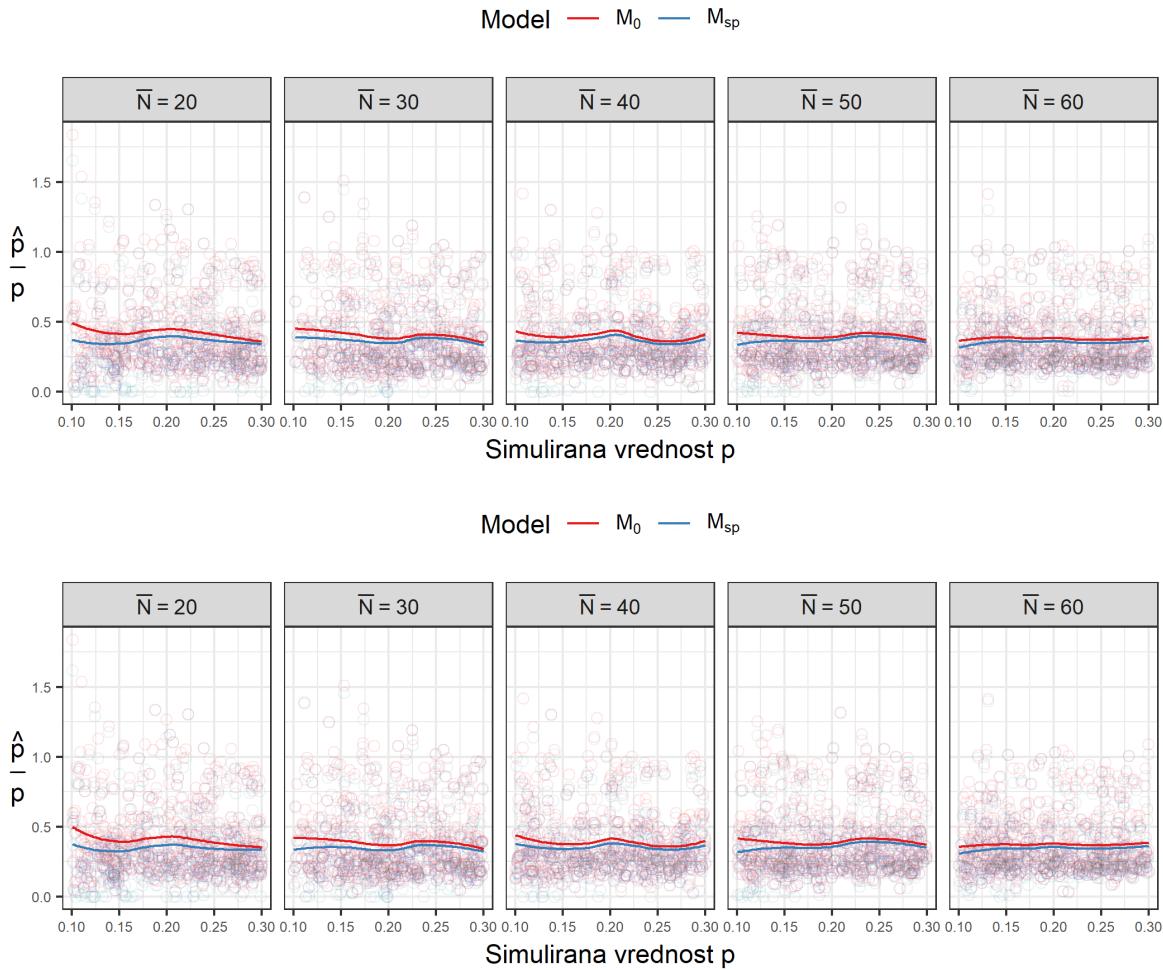
Figure 3: Comparing estimated probability of capture against simulated. Average capture probability estimate from Huggins models M_0 and M_{sp} . Axis x shows simulated p and y shows ratio between estimated (\hat{p}) and simulated value. Upper and lower figures show results where individual covariate has been calculated using empirical and two-dimensional normal functions, respectively. Type of function used does not appear to directly influence the result in a meaningful way.

4.2.2 Primerjava verjetnosti ulovljivosti glede na število simuliranih osebkov

Ocena verjetnosti ulovljivosti (p) glede na število simuliranih osebkov kaže podobno sliko (slika 4). V povprečju model, kjer smo uporabili za izračun še individualno spre-

menljivko (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: Primerjava ocenjene ulovljivosti s simulirano glede na število simuliranih osebkov. Točke prikazujejo razmerje med oceno ulovljivosti (\hat{p}) in simulirano vrednostjo glede na povprečno število generiranih osebkov znotraj območja vzorčenja. Na x osi so prikazane simulirane vrednosti ulovljivosti, na y osi pa razmerje med ocenjeno in simulirano vrednostjo. Okanca prikazujejo podatke glede na povprečno število simuliranih osebkov znotraj območja vzorčenja. Zgornja slika prikazuje podatke, kjer smo individualno spremenljivko ocenili s pomočjo empirične porazdelitve, spodnja pa podatke za simulacije, kjer smo individualno spremenljivko izračunali s pomočjo normalne porazdelitve.

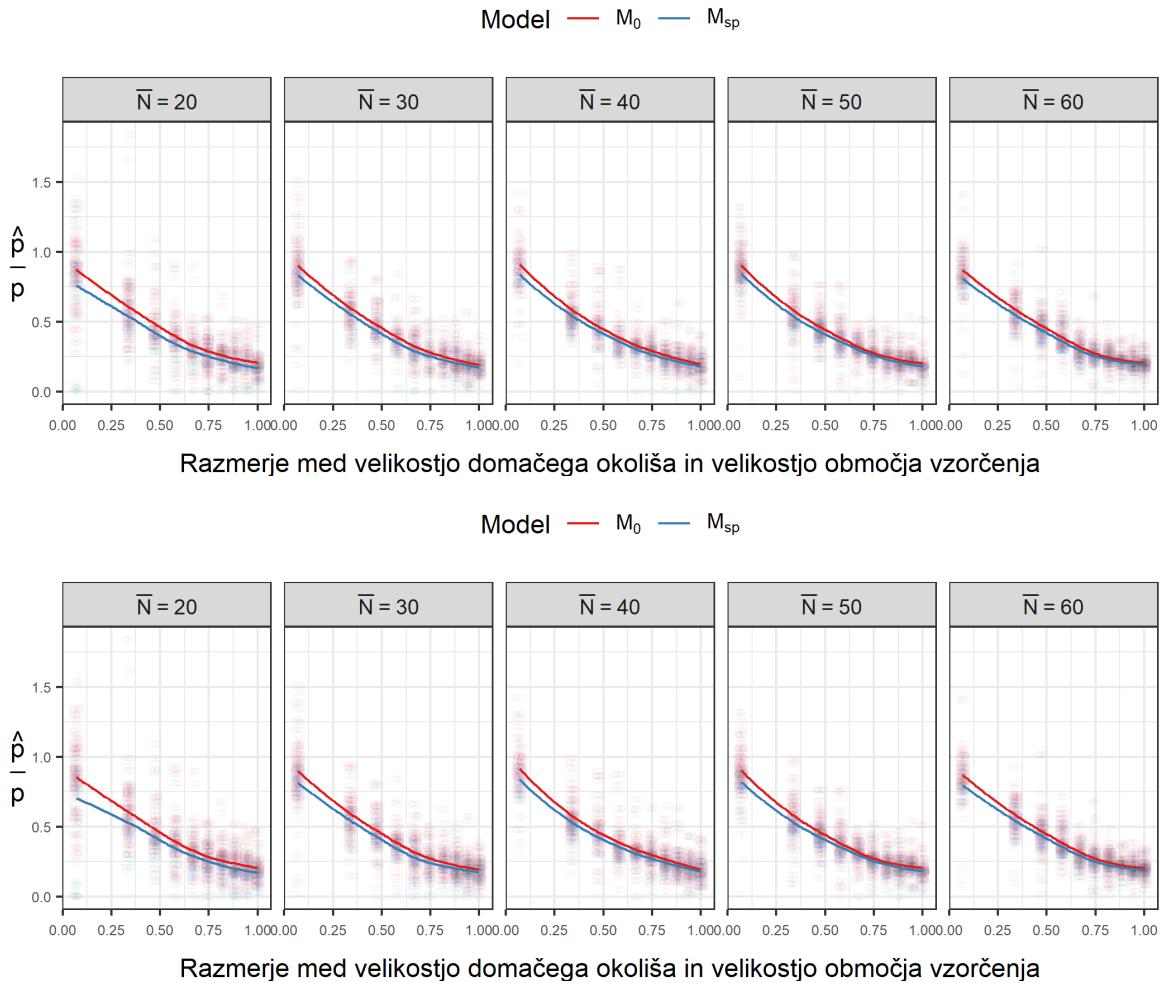
Figure 4: Comparing estimate probability of capture to simulated by number of simulated individuals. Figure shows ratio between estimated (\hat{p}) and simulated (p) probability of capture on the y axis given the simulated values on x axis. Number of individuals generated within the sampling polygon per simulation is divided by facets. Upper and lower figures represent results where individual covariate has been calculated using empirical and two-dimensional normal distribution, respectively.

4.2.3 Primerjava ulovljivosti glede na razmerje velikosti domačega okoliša in območja vzorčenja

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

vzorčenja), 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 rezultatih s slike 5, ne moremo sklepati, da ima število generiranih osebkov v povezavi z razmerjem velikosti domačega okoliša in območja vzorčenja bistven vpliv na oceno ulovljivosti, oziroma je vpliv neznaten. Razlike med simulacijami, kjer smo spreminjali funkcijo za izračun individualne spremenljivke, so majhne oz. jih v praktičnem smislu ni.

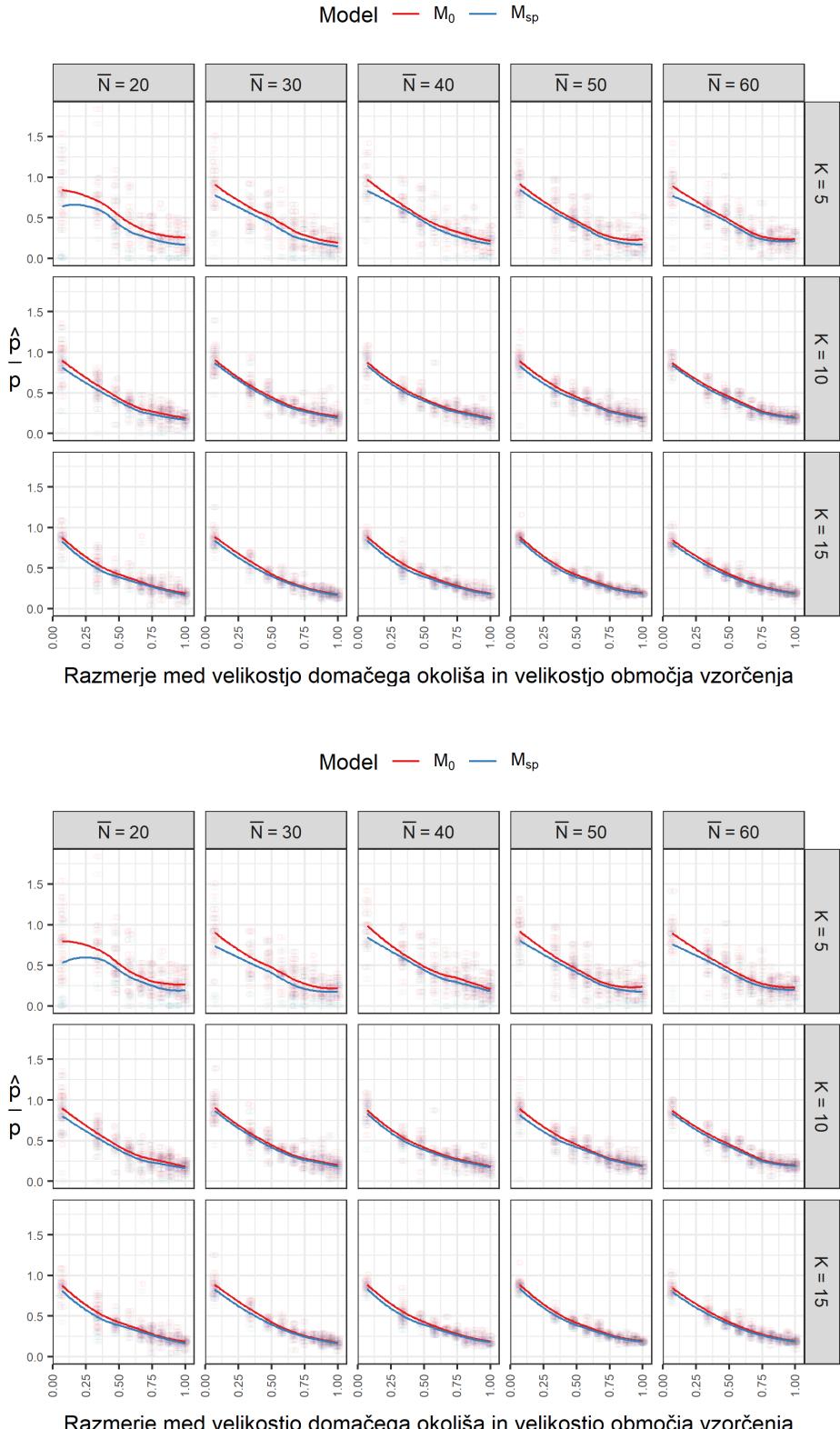


Slika 5: Vpliv razmerja med ocenjeno in pravo ulovljivostjo v povezavi z razmerjem velikosti domačega okoliša in vzorčenega območja. Zgornja slika prikazuje rezultate simulacij, kjer smo individualno spremenljivko izračunali s pomočjo empirične porazdelitve, spodnja pa za izračune s pomočjo normalne porazdelitve. Okenca predstavljajo povprečno število osebkov, ki smo jih generirali znotraj območja vzorčenja.

Figure 5: Influence of ratio between estimated and simulated probability of capture by ratio of home range size and sampling polygon size. Influence of ratio between home range size and sampling polygon size (on x) on estimate of probability of capture depicted as ratio between estimated and simulated probabilities of capture (y axis). Facets depict simulations based on number of simulated individuals within the sampling polygon. Upper and lower images represent simulations where individual covariate was calculated using empirical and two-dimensional normal distribution, respectively.

Razmerje med površinama domačega okoliša in območja vzorčenja ima na pristranskost ocene parametra p velik vpliv. Na sliki 6 je prikazano, kako se pristranskost ocene parametra p spreminja z razmerjem površin domačega okoliša in vzorčenega območja glede na število generiranih osebkov in število odlovnih intervalov. 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 intervalov, ki poleg tega vpliva tudi na variabilnost ocen. Več kot je odlovnih intervalov, manjša je variabilnost v oceni. Slike 6 ne moremo trditi, da smo zaznali bistvene razlike med simulacijami, kjer smo uporabili za izračun normalno ali empirično funkcijo.



Slika 6: Prikaz pristranskosti ocene ulovljivosti (\hat{p}) glede na razmerje površin domačega okoliša in območja vzorčenja v povezavi s številom simuliranih osebkov (stolpci) in odlovnih intervalov (vrstice). Na zgornji sliki so prikazani rezultati simulacij, kjer smo za izračun individualne spremenljivke uporabili empirično, na spodnji pa normalno funkcijo. Z večanjem števila simuliranih osebkov se veča pristranskost in manjša razlika v oceni parametra. Podobno sliko vidimo z večanjem števila odlovnih intervalov (vrstice), kjer poleg višanja pristranskosti in konvergenco povprečij modelov opazimo tudi zmanjševanje variance ocene.

Se nadaljuje

nadaljevanje Figure 6: Depiction of bias as ratio of estimated probability of capture and simulated value against ratio of home range area and sampling polygon area by number of simulated individuals and sampling sessions. Data has been additionally sliced according to number of sampling occasions (K , facets in rows) and number of simulated individuals within sampling polygon (\bar{N} , facets in columns). Bias increases with increased ratio between home range size and sampling polygon size. On average, means of models converge and are highly biased for large ratios of home range size and sampling polygon size. By increasing number of sampling sessions, decrease in variance of estimates is apparent.

4.3 GOSTOTA

Simulirano 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 osebki niso bili omejeni na območje vzorčenja in so prehajali njegove meje, smo dejansko ocenili velikost superpopulacije, katere prostorska omejenost ni nedvoumno določena. Na podlagi ocjenjenega 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.

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

Na sliki 6 so prikazani rezultati izračuna gostot. Na sliki uporabimo kazalnik, ki je z gostoto povezan kot razmerje med gostoto, izračunano iz ocenjene velikosti populacije, in gostoto, izračunano iz 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 povečali 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.

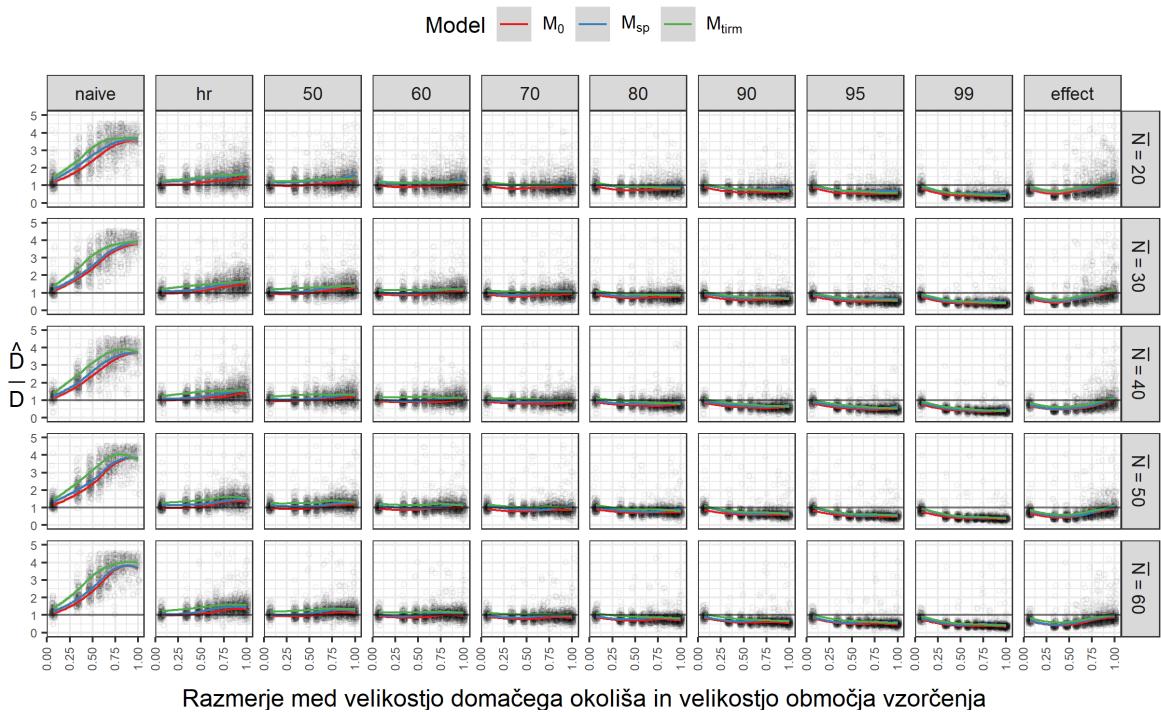
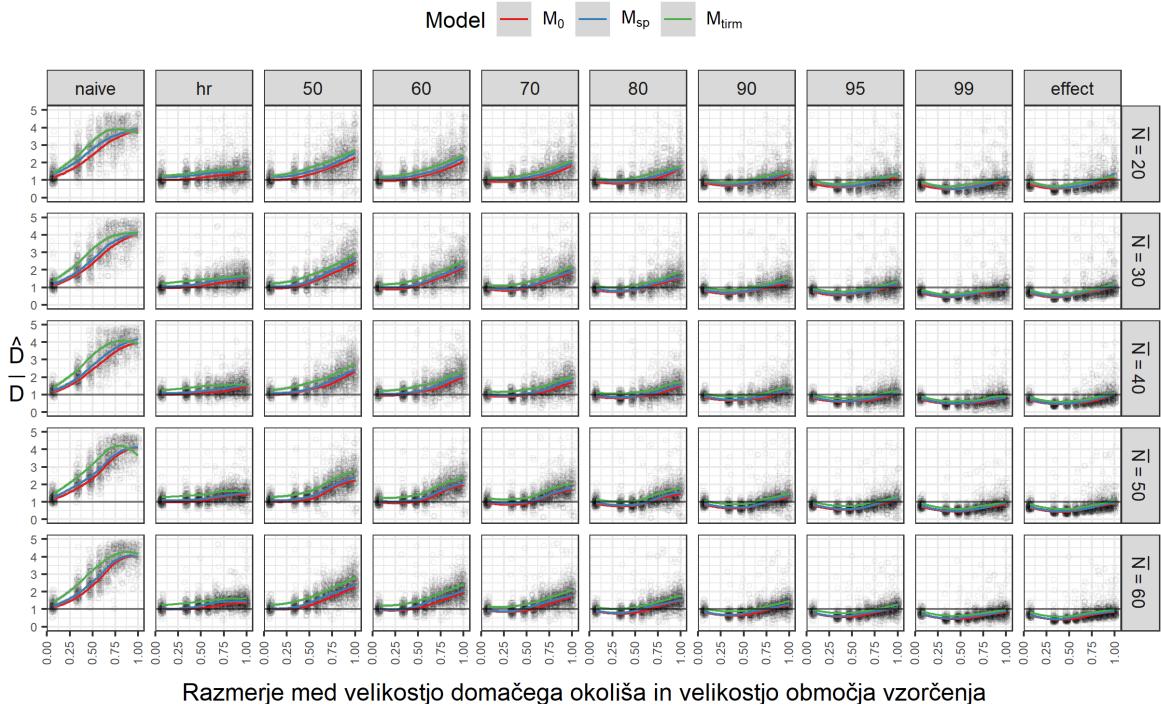
Pristranskost kazalnika v primeru nepovečanega območja hitro narašča z večanjem

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 pristranskost 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 konkavnost gladilke 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 primerljive. 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.



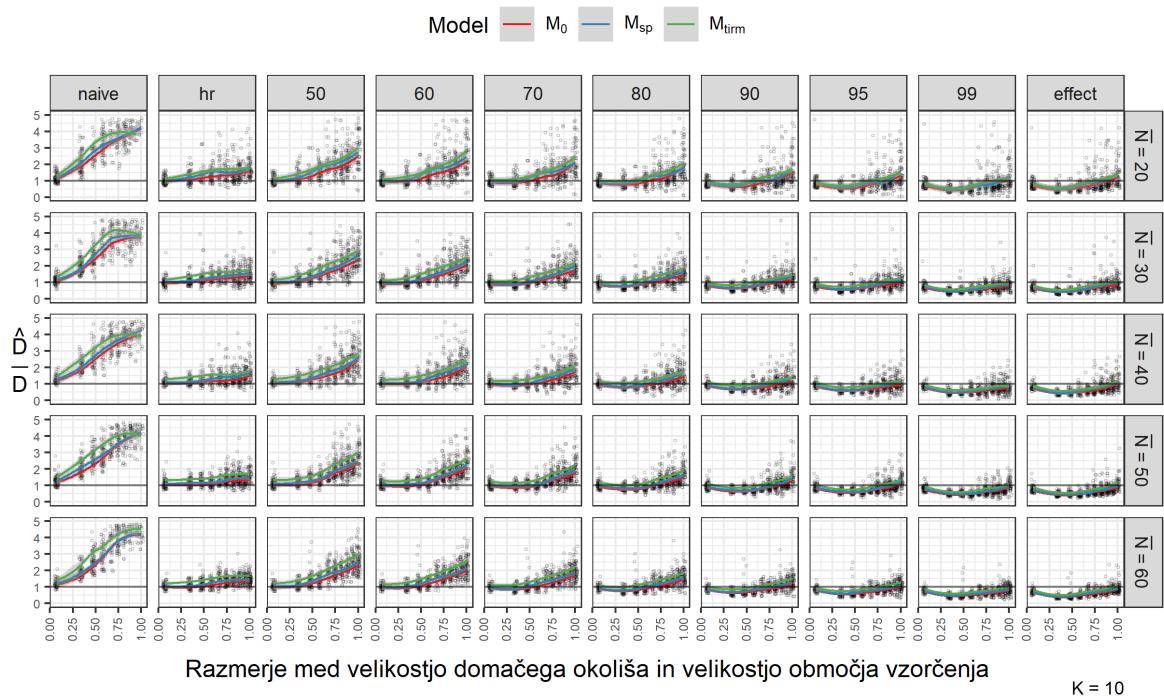
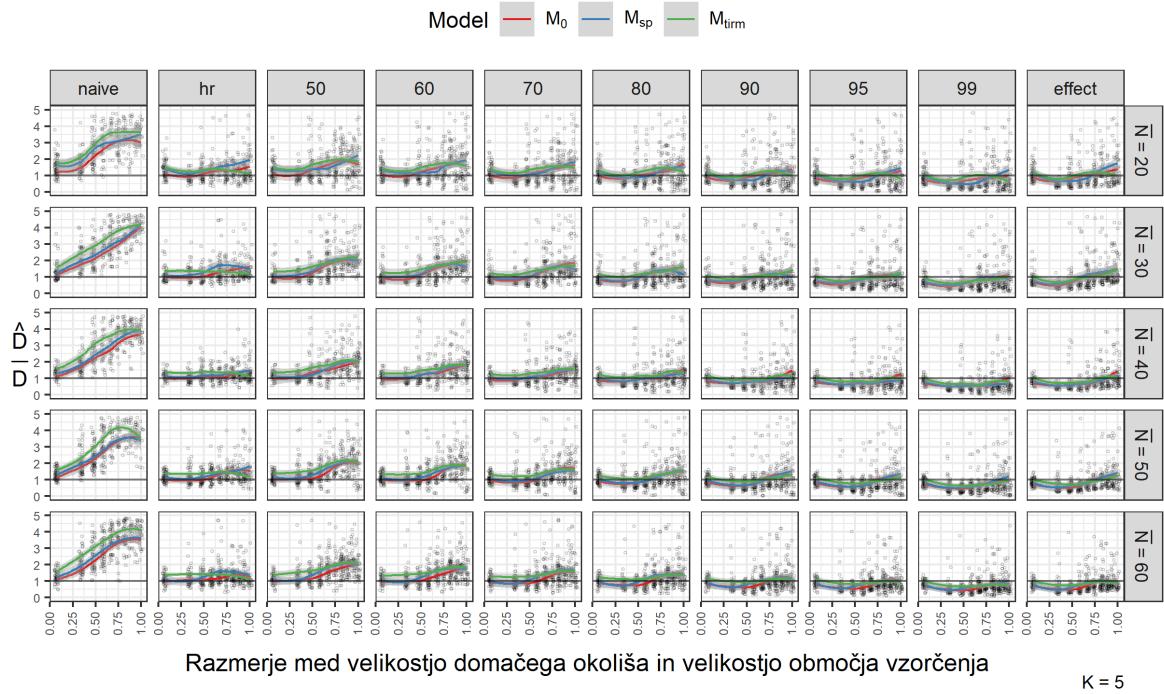
Slika 7: Prikaz kazalnika gostot (\hat{D}/D) glede na razmerje velikosti domačega okoliša in velikosti območja vzorčenja. Na zgornji sliki so prikazani rezultati za simulacije, kjer smo za izračun individualne spremenljivke uporabili empirično porazdelitev, na spodnji pa za simulacije, kjer smo uporabili dvo-razsežno normalno porazdelitev. Stolpci oken ponazarjajo razširitev vzorčenega območja, vrstice pa povprečno število simuliranih osebkov znotraj območja vzorčenja, na katerih smo izvedli vzorčenje ter posledično izračun velikosti populacije in gostote. Pri izračunu smo uporabili tri modela, dve različici Hugginsovega modela (H_0 , H_{sp}), kjer smo ulovljivost modelirali glede na individualno spremenljivko, in TIRM, ki predpostavlja dve skupini z različno ulovljivostjo.

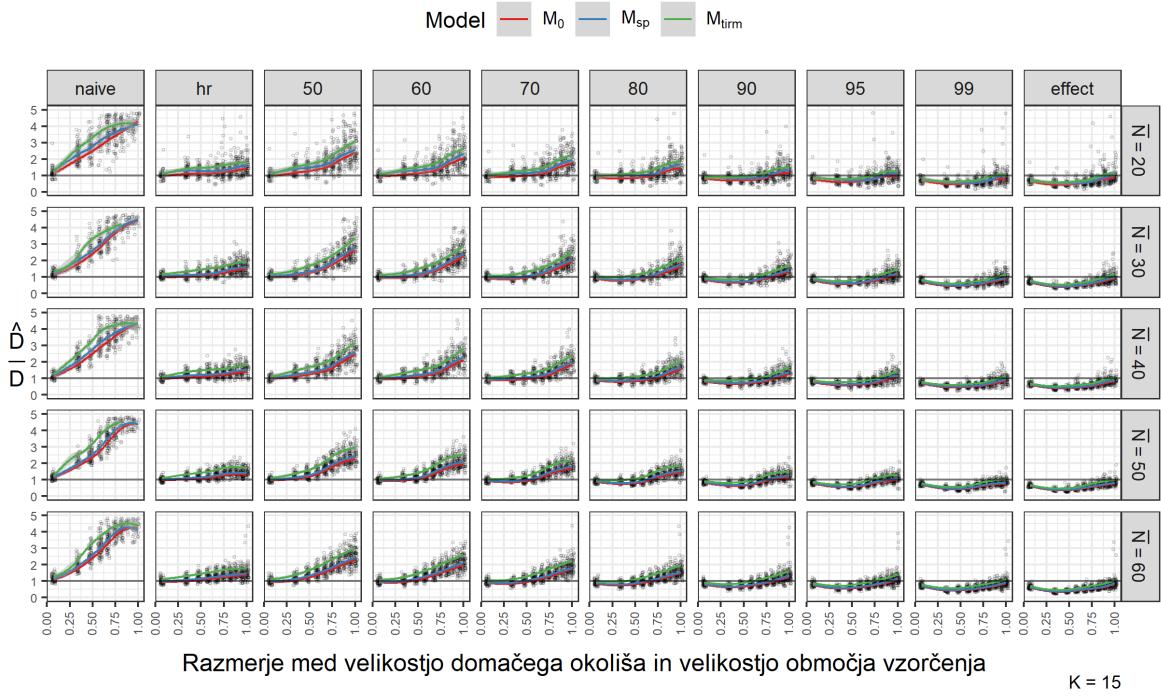
Se nadaljuje

nadaljevanje Figure 7: Index of density estimates. Index of density estimates (\hat{D}/D) on y axis by ratio of home range size and sampling polygon area (x axis). Upper and lower images represent simulations where individual covariate was calculated using empirical and two-dimensional normal distributions, respectively. Facets in columns represent the sampling polygon expanded by a given measure and rows represent the mean number of simulated individuals within the sampling polygon. Three models are used to calculate population size, two Huggins models M_0 and M_{sp} (equal capture probability and probability that depends on individual covariate) and a TIRM model that assumes two groups with each having its own capture probability.

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

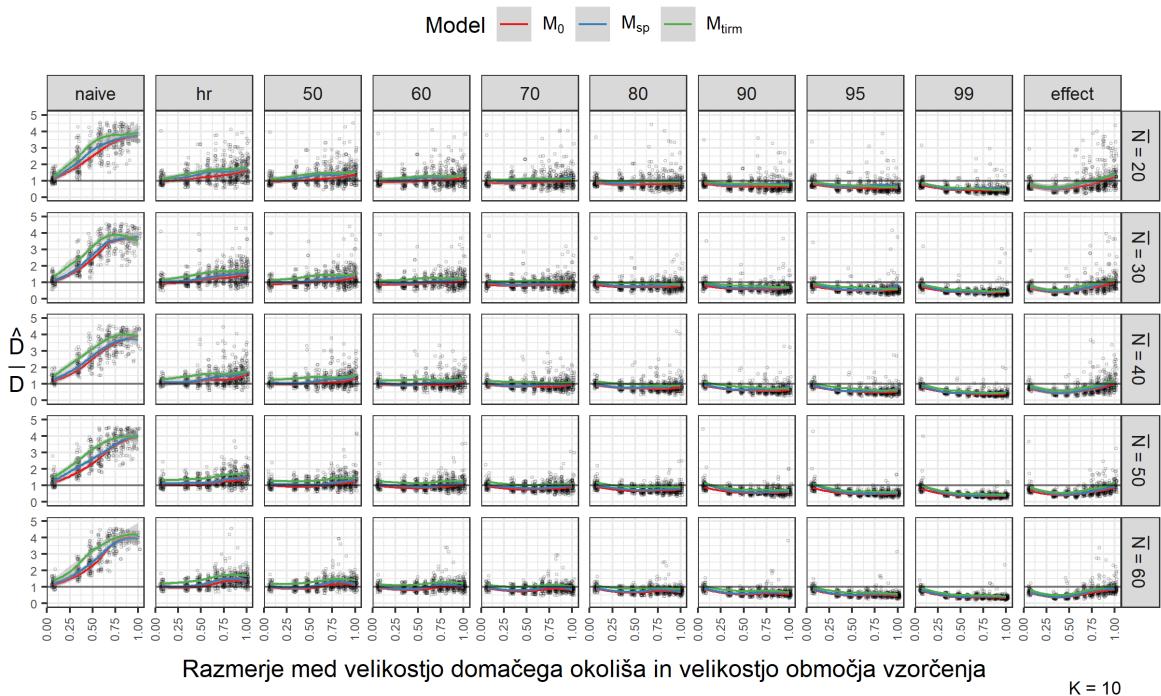
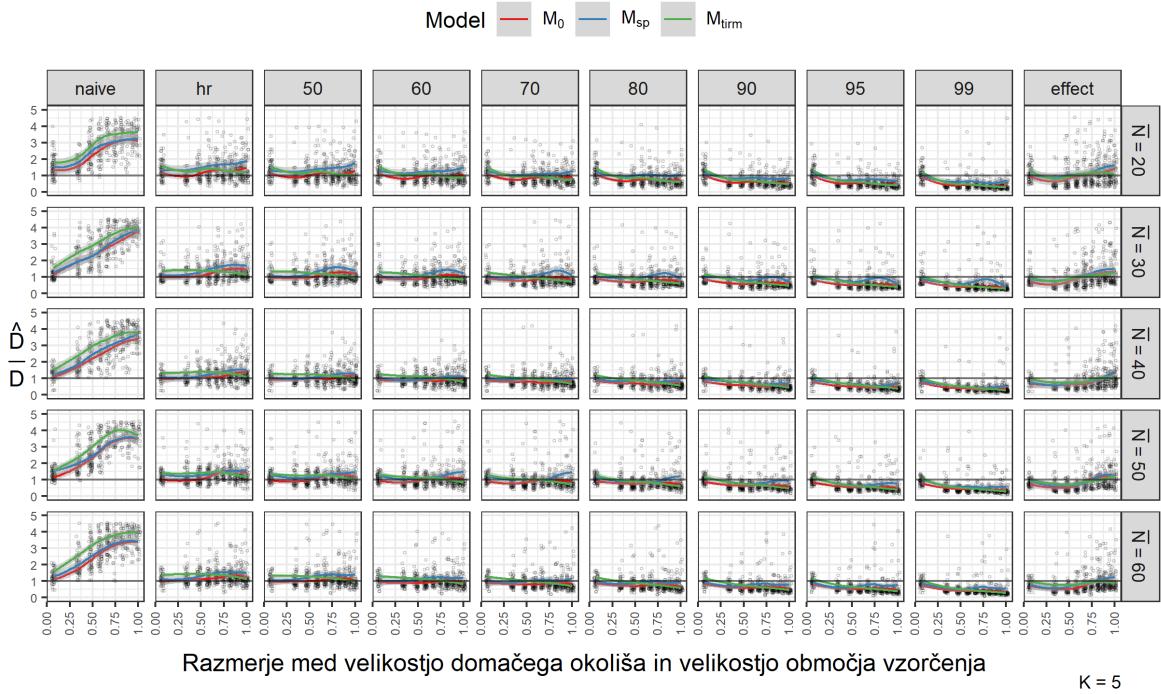
Na slikah 7 in 8 smo podatke s slike 6 prikazali tako, da so podatki razdeljeni na odlovne intervale ($K = [5, 10, 15]$). Prikaz glede na število odlovnih intervalov nam je dal predvsem vpogled v spremembo variabilnosti. Pri simulacijah z najmanj odlovnimi intervali ($K = 5$) je varianca večja kot pri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

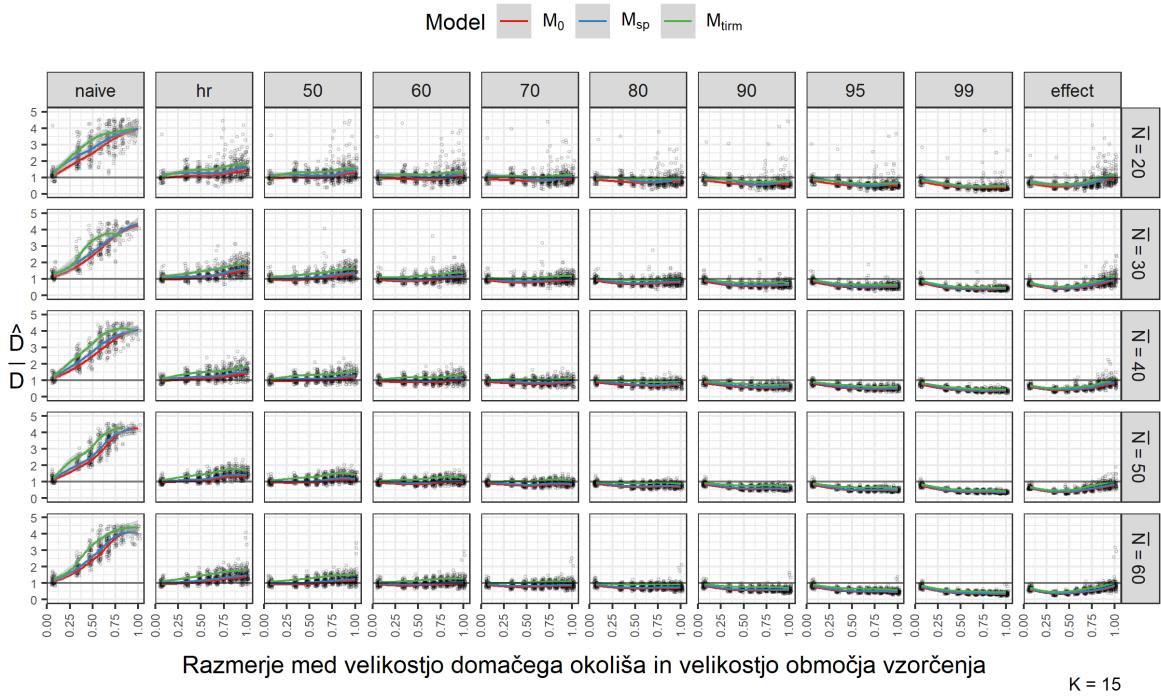




Slika 8: Prikaz ocenjene gostote (empirična porazdelitev). Bolj podroben prikaz podatkov s slike 7 (zgoraj, individualna spremenljivka izračunana s pomočjo empirične porazdelitve), kjer smo podatke dodatno prikazali glede na število odlovnih intervalov (K). Zgoraj $K = 5$, v sredini $K = 10$, spodaj $K = 15$.

Figure 8: Depiction of estimated densities (empirical distribution). Same data as depicted in figure 7 (upper, individual covariate calculated using empirical distribution), but also sliced according to number of sampling sessions ($K = 5$ above, $K = 10$ in the middle and $K = 15$ at the bottom).





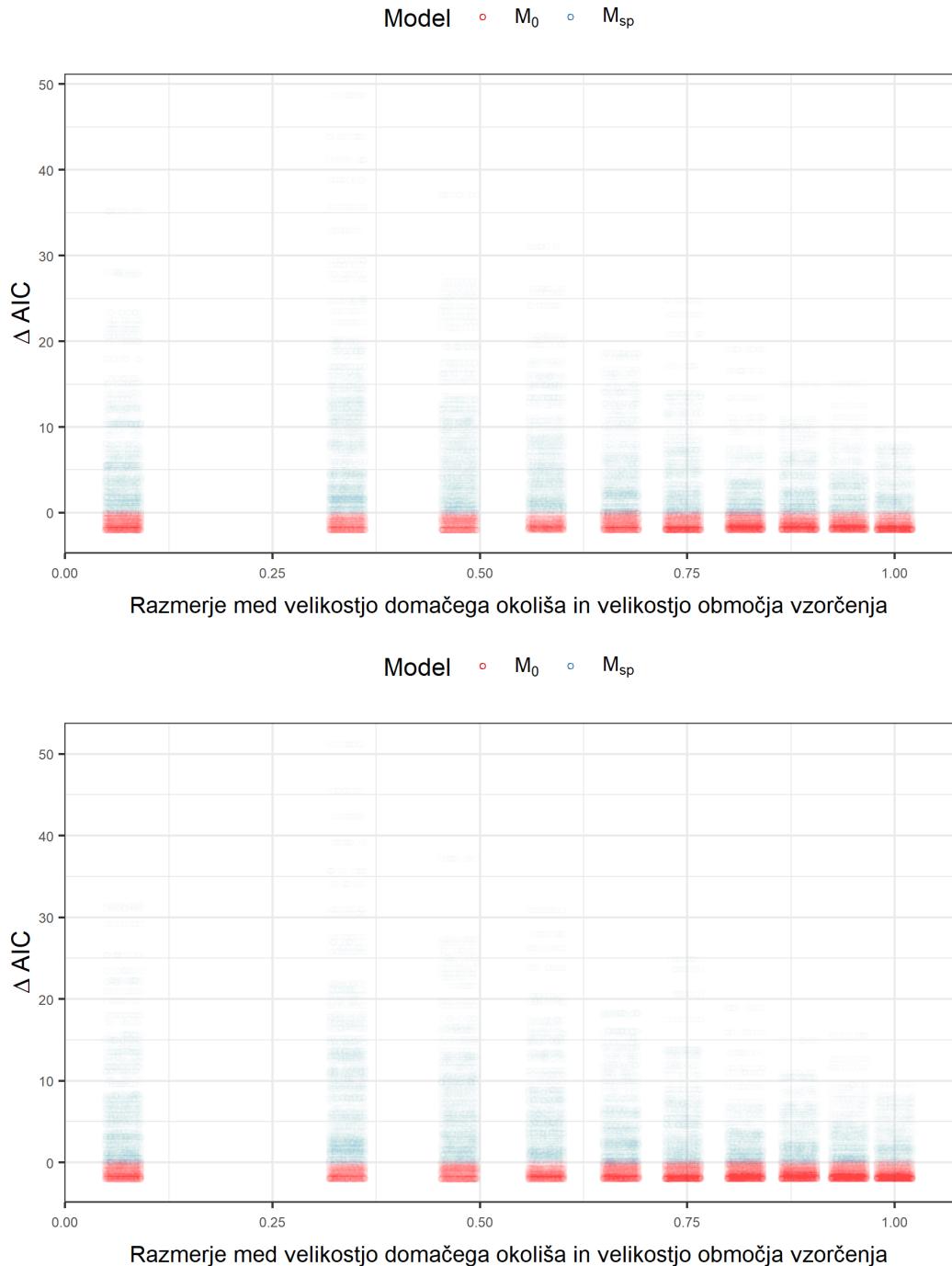
Slika 9: Prikaz ocenjene gostote (normalna porazdelitev). Bolj podroben prikaz podatkov s slike 7 (spodaj, individualna spremenljivka izračunana s pomočjo dvorazsežne normalne porazdelitve), kjer smo podatke dodatno prikazali glede na število odlovnih intervalov (K). Zgoraj $K = 5$, v sredini $K = 10$, spodaj $K = 15$.

Figure 9: Depiction of estimated densities (normal distribution). Same data as depicted in figure 7 (lower, individual covariate calculated using two-dimensional normal distribution), but also sliced according to number of sampling sessions ($K = 5$ above, $K = 10$ in the middle and $K = 15$ at the bottom).

4.4 AICc

Metriko AICc smo primerjali samo za Hugginsova modela M_0 in M_{sp} . Verjetje modela TIRM ni primerljivo s Hugginsovim modelom. Poleg druge funkcije največjega verjetja uporabi za izračun velikosti populacije drugačen set podatkov. Primerjava vrednosti AICc z ostalima dvema modeloma zato ne bi bila pravilna.

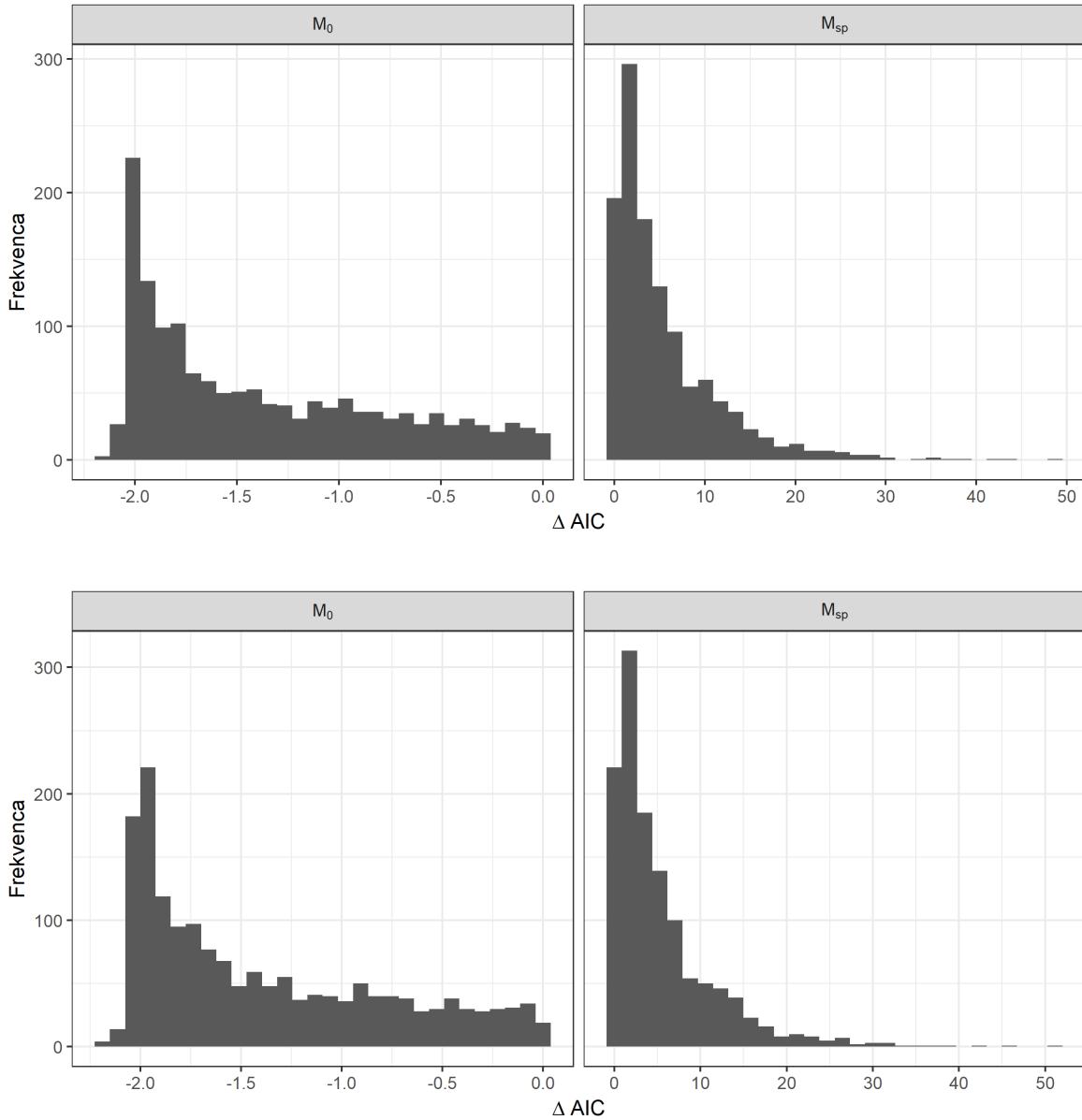
Za vsako simulacijo smo določili, kateri model ima manjši AICc, nato pa razliko med modeloma prikazali na osi y . Vedno smo odšteli AICc modela M_0 od AICc modela M_{sp} . Negativne vrednosti so povezane s tistimi simulacijami, kjer je bil boljši model M_0 , pozitivne pa s simulacijami, kjer je bil boljši model M_{sp} . Če je bil boljši model M_{sp} , je bila razlika pogosto mnogo večja (modra barva) kot v primerih, ko je bil boljši model M_0 (rdeča barva). Število simuliranih osebkov ne vpliva na razliko v AICc, zato smo podatke prikazali brez te spremenljivke na sliki 10.



Slika 10: Prikaz razlike v $AICc$ za modela M_0 in M_{sp} . Na zgornji sliki so prikazani podatki, kjer smo za izračun individualne spremenljivke uporabili empirično porazdelitev, na spodnji pa za normalno porazdelitev. Rdeče obarvane točke predstavljajo simulacije, kjer je bil boljši model M_0 , v modrih pa je bil boljši M_{sp} .

Figure 10: Distribution of $AICc$ for models M_0 and M_{sp} given ratio of home range size and sampling polygon size. Upper and lower figures show data for simulations where individual covariate has been calculated using empirical and two-dimensional normal distribution, respectively. Red color depicts values where model M_0 was superior to model M_{sp} and blue represents cases where model M_{sp} was superior.

Porazdelitve 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 11: Podrobnejši prikaz porazdelitev razlik v ΔAIC_C za modela M_0 in M_{sp} s slike 10. Zgornja dva histograma predstavljata vrednosti, kjer smo za izračun individualne spremenljivke uporabili empirično porazdelitev, spodnja pa vrednosti, kjer smo uporabili normalno porazdelitev. Previdni moramo biti pri razlaganju osi x , saj med modeloma nista neposredno primerljivi.

Figure 11: Detailed view of distribution of δAIC_C for models M_0 and M_{sp} from figure 10. Two histograms at the top and bottom represent simulations where individual covariate has been calculated using empirical and two-dimensional normal distributions, respectively. Note that x axes are not directly comparable between graphs.

5 RAZPRAVA

Pomembno orodje za raziskovanje nekaterih osnovnih parametrov populacij, kot je njena velikost, so metode lova-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 predpostavk zaprtosti populacije in enake verjetnosti ulovljivosti zaradi prihajanja in odhajanja osebkov v in iz območja vzorčenja. To prehajanje imenujemo učinek roba. Posledice učinka roba smo poskusili omiliti s popravkom, ki naj bi po našem mnenju opisoval ulovljivost osebkov v območju vzorčenja.

S pomočjo simulacij in klasičnih modelov za zaprte populacije (Hugginsov model) smo preverili delovanje predlaganega popravka. Vsakemu osebku smo izračunali individualno spremenljivko, ki je odvisna od lege vzorcev (centroida) v prostoru in predpostavljene funkcije gibanja. Gibanje okoli centroida smo opisali s pomočjo dveh porazdelitev. Prva je dvorazsežna normalna porazdelitev, za katero smo uporabili simuliran standardni odklon. Ta 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 (Ivan in sod., 2013a)).

Opisovanje gibanja po (pol)normalni porazdelitvi je v simulacijskih študijah relativno pogosto (Bolker, 2008; Ivan in sod., 2013b), saj omogoča razumevanje gibanja in je hkrati (lahko) dovolj kompleksno, da opiše neke splošne lastnosti osebkov, kot se gibljejo v prostoru. Seveda pa je to poenostavitev, za katero ni nujno, da opiše gibanje osebkov dovolj dobro. Za bolj kompleksna gibanja raziskovalci uporabljajo realne podatke, ki jih naključno vzorčijo, in tako simulirajo gibanje osebkov (npr. Manning in Goldberg, 2010). Tako generirani podatki nam res dajo bolj realen opis gibanja, vendar pa je možno, da dobimo zaradi nerepresentativnega vzorčenja pristranske 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 odklovnih intervalov, število simuliranih osebkov (gostoto) in ulovljivost. Velikost in obliko vzorčenega območja smo ohranjali konstantni, 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.

5.1 PRISTRANSKOST OCENE ULOVLJIVOSTI

Ker območje vzorčenja ne zaobjame populacije v celoti, nekateri osebki prehajajo rob vzorčenja. Pri lovu v simulaciji smo v primeru, da smo tak osebek ujeli zunaj območja

vzorčenja, ta vzorec krnili. Brez krnjenja bi bila ocenjena ulovljivost nepristranska, blizu simulirane vrednosti zaznavnosti. Ker pa smo točke, ki se niso nahajale znotraj območja vzorčenja krnili, smo to ulovljivost za robne osebke 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 glede na velikost 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 krenjeni. Drugi model "ve", da smo podatke krnili, in do neke mere to poskuša nadoknaditi s pomočjo individualne spremenljivke. Ta spremenljivka naj bi opisovala pojav, ki spreminja ulovljivost osebka - v našem primeru opisuje, do kakšne mere osebki niso na voljo za vzorčenje zaradi prehajanja roba območja vzorčenja. V tem oziru kaže, da individualna spremenljivka izboljša model. Ocena se izboljša predvsem za simulacije z manj osebkami in manj odlovnimi intervali. Razlika med modeloma je komaj opazna, ko pa število osebkov in število odlovnih intervalov 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 območja vzorčenja (razmerje se manjša) pa se ta ocena izboljšuje in v najboljšem primeru odstopa le približno 25 %. Zaradi krnjenja je manj verjetno, da bi ocenjene vrednosti v povprečju kadarkoli dosegle teoretično vrednost razmerja 1.

Iz razmerja ocenjene in simulirane ulovljivosti lahko opazimo, da se je variabilnost zmanjšala, ko smo povečali število odlovnih intervalov, nekoliko pa tudi zaradi večjega števila simuliranih osebkov. Manjša variabilnost je na račun tega, da smo v model vključili več podatkov. Takrat so ocene bolj natančne in je posledično ocena parametrov manj variabilna. Ta pojav nam je lahko v pomoč pri načrtovanju poskusa, kjer se odločamo, ali napor vložiti v večanje ulovljivosti osebkov, v večanje števila odlovnih intervalov 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. Za primerjavo, model M_{sp} preko individualne spremenljivke do neke mere upošteva kršenje predpostavke neprehajanja roba območja vzorčenja.

5.2 PRIMERJAVA AICC ZA MODELA 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 (Burnham in Anderson, 2002). Iz tega lahko sklepamo, da je model M_{sp} v tem pogledu v povprečju zares boljši od M_0 . Vendar je razlika v oceni parametra verjetnosti ulovljivosti p za praktične potrebe majhna. Končni pokazatelj, ali je ta model res pomembnejše boljši, je izračun gostote osebkov v območju vzorčenja, kar je navadno cilj raziskav.

5.3 PRIMERJAVA OCEN GOSTOT

Izračun gostote je za klasične raziskave lova-ponovnega ulova po svoje problematičen, ker je ključen 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 mej 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 problem poskušali reševati vsi kasnejši raziskovalci (Williams in sod., 2002).

Naš popravek spada med t. i. ad hoc pristope, kot jih kritizirajo Royle in sod. (2013). Klasični pristopi po večini lovijo na mreži pasti in statistike, 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 od roba se nahaja centroid domačega okoliša in kakšna je njegova oblika.

5.3.1 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 za vse modele (M_0 , M_{sp} in TIRM). Precenjena je na račun tega, da metoda lova-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 intervalov 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.

5.3.2 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 najdaljšo prehodeno razdaljo v simulaciji (effect).

Ocenjena gostota je za posamezen model (M_0 , M_{sp} in TIRM) konsistentna glede na razmerje med površino domačega okoliša in območja vzorčenja. To pomeni, da kažejo rezultati 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 v primerjavi z 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 (Eberhardt, 1990). Ta učinek je še posebej viden na slikah 7 in 8, kjer podatke prikažemo še glede na odlovne intervale. Večje število odlovnih intervalov vpliva na natančnost ocene, ker smo v model prek več odlovnih intervalov vključili več podatkov. Ta učinek je viden šele po tem, ko območje vzorčenja razsirimo za katero koli mero.

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

Bolj kot povečamo površino območja vzorčenja glede na domač okoliš osebka, 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, 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 treba odločevalcem pri podajanju rezultatov poudariti, da gre za optimistično oceno.

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 zmanjšata. 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 prilegли 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 razdalje pripadajočim percentilom niso identične. 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 simuliran standardni odklon). Podoben učinek opazimo pri metodi zankanja, ki navadno daje manjše intervale zaupanja statistik v primerjavi s parametričnimi metodami (Hesterberg, 2015). 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 ni nujno, da odražajo pravo stanje.

Za normalno porazdelitev tako poznamo pravo gibanje, stohastičen je samo položaj centroida domačega okoliša. Razlike, ki jih vidimo med modeloma M_0 in M_{sp} , lahko pripisemo razlikam funkcij, ki jih uporabimo za računanje individualne spremenljivke. Odstopanje centroida, izračunanega na podlagi vzorčenih točk od simuliranega gre pripisati vzorčenju majhnega števila točk.

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

Popravljanje gostote s širjenjem velikosti domačega okoliša sicer pomaga pri bolj pravilni oceni gostote, a šele ko območje vzorčenja povečamo za 80. percentil empirične

porazdelitve (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). Pri povečanju za 90. percentil in več že prihaja do podcenjevanja ocene gostote. Najbolj točni so rezultati za simulacije z robnimi parametri - za največja in najmanjša razmerja med velikostjo domačega okoliša in območja vzorčenja. Podcenjevanje je z vidika upravljanja verjetno bolj zaželeno kot precenjevanje, ker daje zadržano oceno velikosti populacije.

Podatke smo simulirali s pomočjo normalne porazdelitve z znano varianco oz. standardnim odklonom. Ko smo računali individualno spremenljivko za posamezen osebek s pomočjo normalne porazdelitve, smo za vrednost standardnega odklona uporabili simulirano vrednost. Za popravek gostote smo območje vzorčenja razširili za razdaljo 60. percentila (približno en standardni odklon), kjer je bila gostota še najbližja simulirani. Z empirično porazdelitvijo smo to dosegli šele pri razširitvi območja vzorčenja za razdaljo 80. percentila. Iz tega bi lahko sklepali, da moramo območje povečati za določeno razdaljo, za katero, pa iz samih podatkov verjetno težko ocenimo. Morebitna rešitev bi bila z uporabo telemetrije (ali pridobimo podatek iz literature), kjer bi lahko območje vzorčenja povečali za toliko, da zajame večji delež domačih okolišev osebkov, ki so ob robu območja vzorčenja in še nezanemarljivo prispevajo k oceni številčnosti superpopulacije in gostote. Ta pristop ne bi potreboval izračuna individualne spremenljivke in so se ga poslužili že drugod (Ivan in sod., 2013a). Nekateri sicer napovedujejo težve pri ocenjevanju parametra velikosti populacije v primeru heterogenosti ulovljivosti (Link, 2003), a glede na naše rezultate obstajajo primeri, ko napoved ni nujno tako slaba.

Popravek velikosti domačega okoliša bo najbolj koristil študijam, ki imajo opravka z vrstami z relativno velikim domačim okolišem (denimo medved (Miller in sod., 1997; Whittington in Sawaya, 2015)), območje vzorčenja pa je relativno majhno. Če je območje vzorčenja znatno večje od velikosti domačega okoliša (npr. v primeru voluharjev Erlinge in sod., 1990) in je torej razmerje blizu 0, popravek velikosti območja vzorčenja na gostoto ne bo imel bistvenega vpliva.

6 SKLEPI

Ugotovili smo, da je model M_{sp} z dodano individualno spremenljivko v povprečju boljši od modela, kjer spremenljivke nismo uporabili (M_0). Razlike v AICc so relativno velike (do približno 20) v prid M_{sp} . Le v nekaterih primerih je bil boljši model M_0 , a so relativne razlike med modeloma razmeroma majhne (približno dve), do česar lahko pride že zgolj po naključju.

Ocenjeni ulovljivosti po modelu M_{sp} in M_0 sta nižji od simulirane, kar pomeni, da zaznamo krnjeno zaradi prehajanja roba območja vzorčenja, s prvim modelom morda nekoliko bolje. Vendar pa so te razlike s praktičnega vidika zanemarljivo majhne.

Med ocenami gostote modelov M_{sp} , M_0 in TIRM nismo zaznali praktično pomembnih razlik. Z vključitvijo individualne spremenljivke v model M_{sp} 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 proučevanju učinka 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 posledično na 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. Da bomo lahko zanesljivo popravili posledice učinka roba, bo potrebno oceniti velikost oziroma parametre funkcije domačega okoliša s pomočjo več podatkov (npr. telemetrija ali iz literature).

7 POVZETEK (SUMMARY)

7.1 POVZETEK

Populacija je skupina osebkov iste vrste, ki naseljujejo skupni prostor v danem času. Meje populacij navadno niso ostro določene, jih pa raziskovalci uokvirimo iz praktičnih razlogov. Populacijo lahko opišemo s pomočjo več parametrov, kot so rodnost, smrtnost, imigracija, emigracija in nenazadnje številčnost osebkov oziroma gostota. Velikost populacije lahko ocenimo na različne načine, npr. s pomočjo štetja, z metodo štetja na ploskvah, z vzorčenjem na točkah ali transektilih ali s pomočjo metod lova-ponovnega ulova. Te metode imajo predpostavke o zaprtosti populacij, enaki ulovljivosti v danem odlovnem intervalu ter o točnosti oznak, s katerimi nedvoumno označimo osebke. Zaradi bioloških značilnosti vrste ali finančnih, časovnih in pravnih ovir včasih populacije ne moremo vzorčiti v celoti. V tem primeru nekateri osebki prehajajo rob območja vzorčenja in kršijo sicer povezani predpostavki zaprtosti populacije in enake verjetnosti ulovljivosti. Posledice prehajanja roba imenujemo učinek roba in je za oceno velikosti populacije nezaželen. Naš pristop k reševanju tega problema je uporaba Hugginsovega modela, ki sam po sebi ne deluje v prostorskem kontekstu, omogoča pa vključevanje individualnih spremenljivk. Ocena parametrov je tako lahko neposredno pod vplivom prostorskih statistik, ki jih vključimo v model. Cilj disertacije je izboljšati računanje gostote populacij, kjer med vzorčenjem prihaja do kršitve temeljnih predpostavk metode lova-ponovnega ulova. Na podlagi razdalj med vzorčenimi točkami smo izračunali individualno spremenljivko, ki ponazarja delež v prostoru in času, ko je osebek na voljo za odlov. Pričakovali smo, da bo imel model z vključeno individualno spremenljivko nižji AICc in bo tako boljši od modela, kjer te informacije nismo upoštevali. Z vključitvijo informacije o prehajanju roba smo pričakovali, da bomo izboljšali oceno verjetnosti ulovljivosti. Sklepali smo, da bomo oceno gostote izboljšali, ker bomo bolje poznali velikost dejanskega prispevnega območja osebkov v območju vzorčenja. Manjše kršitve zaradi učinka roba ne bi smele bistveno vplivati na pristranskost ocene gostote.

Učinkovitost popravka smo preverili s pomočjo simulacij na nivoju osebka. Za vsako simulacijo smo izračunali parametre modela po Hugginsu, in sicer verjetnost ulovljivosti in velikost ocenjene populacije s pripadajočimi intervali zaupanja. Pri prvem modelu (M_0) smo predpostavili, da je verjetnost ulovljivosti v prvem in vsakem nadaljnjem odlovnem intervalu enaka. Pri drugem modelu (M_{sp}) smo predpostavili, da je verjetnost ulovljivosti za posamezen osebek odvisna od individualne spremenljivke in sovpada z deležem, ki ga je osebek domnevno preživel znotraj območja vzorčenja. Izračunali smo še velikost populacije s pomočjo modela CAPWIRE, tako da smo predpostavili, da obstajata dve skupini osebkov, ki imata različni verjetnosti ulovljivosti. Individualno spremenljivko smo izračunali tako, da smo izračunali histogram prehojenih razdalj med vsemi možnimi točkami za posamezen osebek. Celice histograma smo dodatno utežili s pomočjo metrike, ki uteži razdalje glede na delež daljic, ki padejo v območje vzorčenja, če jih naključno postavimo in orientiramo v območju vzorčenja. Krajše razdalje

so tako utežene bolj navzgor kot daljše razdalje. Uteženemu histogramu smo prilegli parametrično funkcijo, ki spominja na kumulativno Weibullovo porazdelitev s tremi parametri. Porazdelitev smo v izhodišču postavili v centroid točk, ki smo jih vzorčili za posamezen osebek, in izračunali delež, ki se nahaja znotraj območja vzorčenja. Na ta način smo v model vpeljali prostorsko komponento, ki je prej manjkala. Da bi imeli orientacijsko točko, s katero lahko primerjamo popravek, smo za računanje individualne spremenljivke uporabili dvorazsežno normalno porazdelitev, za parameter μ smo uporabili centroid vzorčenih točk, za parameter σ pa simulirano vrednost. Tako smo dobili ‐zlati standard‐, ki nam je služil kot najboljši približek simuliranim vrednostim. Mere, s katerimi smo preverjali učinkovitost popravka, so ocenjena ulovljivost, AICc in gostota. Izvedli smo 6000 simulacij, kar je skupaj z MARK analizo trajalo več kot dva dni na zmogljivem računalniku s 24 sredicami. Ocenjena verjetnost ulovljivosti p ni bila odvisna od števila simuliranih osebkov niti od načina izračuna individualne spremenljivke in je bila nekoliko nižja za model M_{sp} . Za M_0 in M_{sp} je bila ocena \hat{p} pod simulirano. Odstopanje od simulirane vrednosti je bilo največje za simulacije, kjer je bilo razmerje med velikostjo domačega okoliša in velikostjo območja vzorčenja največje. Natančnost ocene parametra \hat{p} je bila v največji meri odvisna od števila odlovnih intervalov in števila generiranih osebkov. Gostoto smo izračunali tako, da smo ocenjeno velikost populacije delili z velikostjo območja vzorčenja, povečanega za percentil (50., 60., 70., 80., 90., 95., 99.) prehujene razdalje obeh funkcij, ki smo jih uporabili za izračun individualne spremenljivke, za standardni odklon simulirane velikosti domačega okoliša in za najdaljšo zaznano prehujeno razdaljo. Huginsova modela M_0 in M_{sp} sta dala nekoliko manj pristranske rezultate kot model CAPWIRE (TIRM). V primerih, ko je bilo razmerje med velikostjo domačega okoliša in velikostjo območja vzorčenja majhno, je bila pristransko ocene gostote najmanjša, z večanjem tega razmerja pa je strmo naraščala. Za simulacije, kjer je bil model M_0 boljši od M_{sp} , je bil δ AICc navadno majhen, do ~ 2 . V primerih, ko je bil M_{sp} boljši od modela M_0 , pa je bila razlika v AICc večja (> 4).

Ocena parametra verjetnosti ulovljivosti (p) je bila v splošnem pristranska za vse simulacije, še najbolj pa za tiste simulacije, kjer je razmerje med velikostjo domačega okoliša in velikostjo območja vzorčenja relativno veliko. Pojav je podoben za oba modela (M_0 , M_{sp}), razlika pa se zmanjša za simulacije, kjer smo generirali več osebkov in vzorčili v več odlovnih intervalih. Razliko od simulirane vrednosti parametra p pojasnjujemo s tem, da smo točke, ki smo jih vzorčili zunaj območja vzorčenja, krnili in posledično znižali simulirano vrednost. Iz podatkov, ki so nam na voljo, ne moremo na neodvisen način izluščiti, kakšna je bila dejanska simulirana vrednost, ampak verjetno ni daleč od tega, kar sta ocenila modela M_0 in M_{sp} za simulacije z majhnim razmerjem med velikostjo domačega okoliša in velikostjo območja vzorčenja. Nismo zaznali razlik v rezultatih, kjer smo za računanje individualne spremenljivke uporabili normalno in empirično porazdelitev. Metrika AICc sicer v povprečju kaže, da je model M_{sp} boljši, a so razlike v ocenah med tema modeloma v povprečju za praktične potrebe zanesljivo majhne. Teoretično gostoto osebkov smo izračunali tako, da smo simulirano število osebkov delili s površino območja, kjer smo generirali njihove centroide. To smo uporabili kot referenco, s katero smo primerjali ocenjeno gostoto. Naivna gostota je

v splošnem zelo pristranska v simulacijah, ki so imele relativno veliko razmerje med velikostjo domačega okoliša in velikostjo območja vzorčenja. Pokazali smo, na kar v literaturi opozarjajo, da so nepopravljene ocene gostote lahko zelo pristranske navzgor. Z večanjem območja vzorčenja se pristranskost zmanjšuje in je najmanjša, ko ga povečamo za razdaljo 60. percentila v primeru, ko smo individualno spremenljivko izračunali s pomočjo normalne porazdelitve, in za razdaljo 80. percentila v primeru uporabe empirične porazdelitve. To namiguje, da je potrebno območje vzorčenja povečati za tako mero, da je vanj vključen večji delež domačega okoliša glede na pravo porazdelitev. Razlike v oceni gostote med modeli niso bile velike. Vsi modeli precejnujejo oceno gostote, najbolj odstopajo ocene po modelu TIRM, po Hugginsu pa sta si modela M_0 in M_{sp} podobna. Uporaba modela M_{sp} , kjer upoštevamo individualno spremenljivko, sama po sebi bistveno ne izboljša ocene. Še vedno moramo biti pazljivi pri tem, kako povečamo območje vzorčenja.

7.2 SUMMARY

Population is defined as individuals belonging to the same species that cohabit particular space in time. Though population borders are generally diffuse, they are delineated by researchers for practical purposes. A population can be described using different parameters, incl. fecundity, mortality, immigration, emigration and population size. Population size can be estimated using different approaches, either based on counts, distance-based or using mark-recapture models. Mark-recapture methods assume closure, equal probability of capture and no tag loss. Due to biological, financial or political reasons some populations can not be sampled in their entirety. As a result, some individuals traverse the sampling area border, thus violating the assumptions of closure and equal probability of capture. This causes the edge effect and has been a known issue for decades. In this study, we used the Huggins model to try and improve model parameters using a spatially explicit individual covariate. We expected to improve density estimates by expanding the sampling area based on the proportion of area/time where or when an individual was observable within the sampling polygon. A model that includes this individual covariate was expected to perform better in terms of AICc as well as be superior in estimating probability of capture p . Slight violations of model assumptions should have no discernable effect on model parameter estimates.

To assess the feasibility of this correction, we used a simulation-based approach. For each simulation we calculated model parameters (\hat{p} , \hat{N}) for Huggins and \hat{N} for the CAPWIRE (TIRM) model, including 95 % confidence intervals. For Huggins we constructed model M_0 where initial and all subsequent captures were assumed to be equal, and model M_{sp} where initial capture and all subsequent captures were governed by an introduced individual covariate which described the proportion of area/time spent within the sampling polygon. TIRM formulation of the CAPWIRE model assumed two groups of individuals with each group having a distinct probability of capture. Individual covariate was calculated by calculating the histogram of pooled individual pairwise walked distances. A weight was used for each interval, depending on the pro-

portion of randomly placed and oriented distances from that interval that fell entirely within the sampling polygon. A function which resembles a three-parameter cumulative Weibull distribution was fitted to this histogram. Origin point of the distribution was placed into the centroid calculated for each individual. Proportion of area/time within the sampling polygon was calculated and stored as the individual's covariate governed by space use. In order to obtain a reference point, a similar procedure was performed using a two-dimensional normal distribution that was parameterized using simulated σ and calculated centroid for μ . To discern any effect the correction would have, we compared AICc values of models M_0 and M_{sp} as well as bias in density between all three models (M_0 , M_{sp} , TIRM). In total 6000 simulations were performed, which took a little over 48 hours of computation time on a powerful 24-core computer. Estimated probability of capture (\hat{p}) did not depend on the number of simulated individuals nor on how individual covariate was computed (based on a two-dimensional normal or empirical distribution) and was lower for model M_{sp} . Parameter \hat{p} never reached simulated value. Discrepancy between simulated and estimated probability of capture was greatest for simulations where ratio between home range size and sampling polygon size was highest. Precision of the estimate was most influenced by number of sampling sessions and number of simulated individuals. Density was compared against simulated density by calculating naïve density where estimated population size was divided by area of the sampling polygon. In addition, the sampling area was expanded by a percentile (50th, 60th, 70th, 80th, 90th, 95th, 99th) of the functions used to calculate the individual covariate, by the standard deviation of simulated home range size, and by maximal distance walked. Density estimates calculated from Huggins models (M_0 , M_{sp}) yielded slightly less biased results compared to TIRM. Bias in density was smallest in cases where ratio between home range size and sampling polygon size was lowest and drastically increased with increased ratio. In cases where AICc was smaller for model M_0 , δ AICc was on average small and increased for cases where M_{sp} was superior.

Probability of capture estimate \hat{p} was biased in reference to the simulated value for all simulations. This was especially the case where ratio between home range size and sampling polygon size was high ($\rightarrow 1$). The difference between models M_0 and M_{sp} is small for practical purposes and even decreases when large numbers of individuals and many sampling sessions are used. This holds true for all simulations irrespective of the way the individual covariate has been calculated. The discrepancy between estimated and simulated value could be due to kerning of sampling points that have fallen outside of the sampling area. From the data at hand it is impossible to discern the true probability of capture; however, it is probably close to what has been estimated with models M_0 and M_{sp} for cases where ratio between home range size and sampling polygon area was low. While AICc indicates that model M_{sp} performed better, the effect size (difference between M_0 and M_{sp} estimates) is quite small to have much practical effect. To obtain theoretical density, we divided the number of simulated individuals with the area used to simulate their centroids. This was used as a "gold standard" against which we measured calculated densities. Naïve density (population size divided by sampling polygon size) proved to be very biased for all cases other than those

where ratio between home range size and sampling polygon size was low. This is in accord with what is warned against in the literature. If density is calculated without accounting for sampling extent, it can be biased upward. By increasing sampling polygon size, the overestimated density is approaching estimated and is close to unbiased for cases where the sampling polygon has been expanded by the 60th percentile of the simulated two-dimensional normal distribution and the 80th percentile of the empirical distribution. This hints that the sampling polygon used to calculate density should include the majority of home ranges. Density estimates were similar between different models, but on average higher for the TIRM model. Using an individual covariate in a model does not by itself significantly improve the density estimate and special care still needs to be taken when increasing the sampling polygon area to compensate for estimating density.

8 VIRI

- Akanda M.A.S., Alpizar-Jara R. 2014. Estimation of capture probabilities using generalized estimating equations and mixed effects approaches. *Ecology and Evolution*, 4, 7: 1158–1165
- Amos W., Balmford A. 2001. When does conservation genetics matter? *Heredity*, 87, 3: 257–265
- Bivand R., Rundel C. 2017. rgeos: Interface to Geometry Engine - Open Source (“GEOS”) (dodatni paket za program R)
- Bivand R.S., Pebesma E.K., Gomez-Rubio V. 2008. Applied spatial data analysis with R. 1st ed. New York, Springer: 374 str.
- Bolker B.M. 2008. Ecological models and data in R. 1st ed. Princeton, Princeton University Press: 408 str.
- Boulanger J., Kendall K.C., Stetz J.B., Roon D.A., Waits L.P., Paetkau D. 2008. Multiple data sources improve DNA-based mark-recapture population estimates of grizzly bears. *Ecological Applications*, 18, 3: 577–589
- 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
- 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
- Burnham K., Anderson D. 2002. Model selection and multimodel inference: A practical information-theoretic approach. 2nd ed. New York, Springer-Verlag: 488 str.
- 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
- Chandler R.B., Andrew Royle J. 2013. Spatially explicit models for inference about density in unmarked or partially marked populations. *Annals of Applied Statistics*, 7, 2: 936–954
- Chandler R.B., Hepinstall-Cyberman 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
- Cockburn A., Lidicker W.Z. 1983. Microhabitat heterogeneity and population ecology of an herbivorous rodent, *Microtus californicus*. *Oecologia*, 59, 2: 167–177

- Cooch E., White G. 2012. Program MARK: A gentle introduction. 1116 str.
<http://www.phidot.org/software/mark/docs/book/> (4. maja 2017)
- DeAngelis D.L., Grimm V. 2014. Individual-based models in ecology after four decades. F1000prime reports, 6: 39, doi: 0.12703/P6-39, 6 str.
- Dice L.R. 1938. Some census methods for mammals. The Journal of Wildlife Management, 2, 3: 119–130
- Dice L.R. 1941. Methods for estimating populations of animals. The Journal of Wildlife Management, 5, 4: 398–407
- Eberhardt L.L. 1990. Using radio-telemetry for mark-recapture studies with edge effects. The Journal of Applied Ecology, 27, 1: 259–271
- Efford M.G., Dawson D.K., Robbins C.S. 2004. DENSITY: Software for analysing capture-recapture data from passive detector arrays. Animal Biodiversity and Conservation, 27, 1: 217–228
- Erlinge S., Hoogenboom I., Agrell J., Nelson J., Sandell M. 1990. Density-related home-range size and overlap in adult field voles (*Microtus agrestis*) in southern Sweden. Journal of Mammalogy, 71, 4: 597–603
- Faraway J.J. 2006. Extending the linear model with R: generalized linear, mixed effects and nonparametric regression. 1st ed. Boca Raton, Chapman & Hall/CRC: 301 str
- Foster R.J., Harmsen B.J. 2012. A critique of density estimation from camera-trap data. Journal of Wildlife Management, 76, 2: 224–236
- Hesterberg T.C. 2015. What teachers should know about the bootstrap: resampling in the undergraduate statistics curriculum. American Statistician, 69, 4: 371–386
- 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
- Huggins R.M. 1989. On the statistical analysis of capture experiments. Biometrika, 76, 1: 133–140
- Ivan J.S., White G.C., Shenk T.M. 2013a. Using auxiliary telemetry information to estimate animal density from capture-recapture data. Ecology, 94, 4: 809–816
- Ivan J.S., White G.C., Shenk T.M. 2013b. Using simulation to compare methods for estimating density from capture-recapture data. Ecology, 94, 4: 817–826
- Jiménez J., Higuero R., Charre-Medellin J.F., Acevedo P. 2017. Spatial mark-resight models to estimate feral pig population density. *Hystrix*, 26: 1–6

- Jugovic J., Črne M., Lužnik 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
- Kendall W.L. 1999. Robustness of closed capture-recapture methods to violations of the closure assumption. Ecology, 80, 8: 2517–2525
- Kindberg J., Swenson J.E., Ericsson G.G., Bellemain E., Miquel C., Taberlet P. 2011. Estimating population size and trends of the Swedish brown bear *Ursus arctos* population. Wildlife Biology, 17, 2: 114–123
- Klein D.R. 1968. The introduction, increase, and crash of reindeer on St. Matthew Island. The Journal of Wildlife Management, 32, 2: 350
- Krebs C.J. 2001. Ecology: The experimental analysis of distribution and abundance. 5th ed. San Francisco, Benjamin Cummings: 695 str.
- Laake J. 2013. RMark: An R interface for analysis of capture-recapture data with MARK (dodatni paket za program R)
- Link W.A. 2003. Nonidentifiability of population size from capture-recapture data with heterogeneous detection probabilities. Biometrics, 59, 4: 1123–1130
- Luštrik R., Turjak M., Kralj-Fišer S., Fišer C. 2011. Coexistence of surface and cave amphipods in an ecotone environment. Contributions to Zoology, 80, 2: 133–141
- Maechler M., Rousseeuw P., Struyf A., Hubert M., Hornik K. 2018. cluster: cluster analysis basics and extensions (dodatni paket za program R)
- Manning J.A., Goldberg C.S. 2010. Estimating population size using capture-recapture encounter histories created from point-coordinate locations of animals. Methods in Ecology and Evolution, 1, 4: 389–397
- Microsoft, Weston S. 2017a. doParallel: Foreach parallel adaptor for the “parallel” package (dodatni paket za program R)
- Microsoft, Weston S. 2017b. foreach: Provides foreach looping construct for R (dodatni paket za program 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
- 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, 133: 1–55
- Moqanaki E.M., Jiménez J., Bensch S., López-Bao J.V. 2018. Counting bears in the Iranian Caucasus: Remarkable mismatch between scientifically-sound population estimates and perceptions. Biological Conservation, 220: 182–191

- O'Connell A.F., Nichols J.D., Karanth K.U. 2010. Camera traps in animal ecology: Methods and analyses. 1st ed. Tokyo, Springer Science & Business Media: 271 str.
- 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
- Pennell M.W., Miller C.R. 2012. capwire: Estimates population size from non-invasive sampling (dodatni paket za program R)
- 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
- Pollock K.H. 2000. Capture-recapture models. *Journal of the American Statistical Association*, 95, 449: 293–296
- Pollock K.H., Hines J.E., Nichols J.D. 1984. The use of auxiliary variables in capture-recapture and removal experiments. *Biometrics*, 40, 2: 329–340
- R Core Team 2018. R: A language and environment for statistical computing (program R)
- Roche E.A., Brown C.R., Brown M.B., Lear K.M. 2013. Recapture heterogeneity in cliff swallows: Increased exposure to mist nets leads to net avoidance. *PLoS ONE*, 8, 3: 1–15: e58092, doi: 10.1371/journal.pone.0058092, 15 str.
- Rowlingson B., Diggle P. 2017. splancs: Spatial and space-time point pattern analysis (dodatni paket za program R)
- Royle J.A., Chandler R.B., Sollmann R., Gardner B. 2013. Spatial capture-recapture. 1st ed. Waltham, Academic Press: 576 str.
- Seber G. 1986. A review of estimating animal abundance. *Biometrics*, 42, 2: 267–292
- Skrbinšek T., Luštrik R., Majić-Skrbinšek A., Potočnik H., Kljun F., Jelenčič M., Kos I., Trontelj P. 2018. From science to practice: genetic estimate of brown bear population size in Slovenia and how it influenced bear management. *European Journal of Wildlife Research*. (v recenziji)
- Stanley T.R., Burnham K.P. 1999. A closure test for time-specific capture-recapture data. *Environmental and Ecological Statistics*, 6, 2: 197–209

- 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
- White G.C., Anderson D.R., Burnham K.P., Otis D.L. 1982. Capture-recapture and removal methods for sampling closed populations. 1st ed. Los Alamos, Los Alamos National Laboratory: 235 str.
- White G.C., Burnham K.P. 1999. Program MARK: survival estimation from populations of marked animals. *Bird Study*, 46: S120–S139
- White G.C., Shenk T.M. 2001. Population estimation with radio-marked animals. V: Radio tracking and animal populations. Millspaugh J. J., Marzluff J. M. (ur.). Waltham, Academic Press: 329–350
- Whittington J., Sawaya M.A. 2015. A comparison of grizzly bear demographic parameters estimated from non-spatial and spatial open population capture-recapture models. *PLoS ONE*, 10, 7: e0137940, doi: 10.1371/journal.pone.0137940, 17 str.
- Wickham H. 2009. *ggplot2: elegant graphics for data analysis*. 1st ed. New York, Springer: 213 str.
- 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
- Williams B.K., Nichols J.D., Conroy M.J. 2002. Analysis and management of animal populations. 1st ed. Waltham, Academic Press: 817 str.
- Wilson K.R., Anderson D.R. 1985. Evaluation of two density estimators of small mammal population size. *Journal of Mammalogy*, 66, 1: 13–21

ZAHVALA

Doktorski študij je delno sofinancirala Evropska unija, in sicer iz Evropskega socialnega sklada. Sofinanciranje se izvaja v okviru Operativnega programa razvoja človeških virov za obdobje 2007-2013, 1. razvojne prioritete Spodbujanje podjetništva in prilagodljivosti; prednostne usmeritve_1. 3: Štipendijske sheme.

Omenjeno sofinanciranje je bilo v obliki neobdavčenih 29.350,00 EUR. Odšteti gre še zneske, ki sem jih zapravil za priporočeno pošiljanje dokazil in bančnih potrdil na Univerzo.

Davnega leta 2008 me je za analizo podatkov v okviru diplomske naloge Jernej 'yerpo' Polajnar usmeril k programu R. Program in skriptni jezik sta se mi dopadla in programiranje pozno v noč na domačiji Štelcar ni bilo nič nenavadnega. Pri analizi mi je bil v veliko pomoč Joris Meys z gentske univerze v Belgiji, ki je poskrbel, da nisem obupal že na začetku. Na srečo sem na Oddelku za biologijo našel še nekaj zanesenjakov in skupaj z Maartenom de Grootom in Majo Zagmajster smo ustanovili skupino navdušencev nad R-jem, ki sem jo poimenoval Rkoholiki. Na enem od srečanj je do mene pristopil Tomaž Skrbinšek in mi, še preden sem diplomiral, ponudil delo na projektu, ki ga je vodil Peter Trontelj. Z zamahom peresa sta mi dala edinstveno priložnost, da sem hobti zamenjal za poklic, kjer sem lahko programiral in si s tem še rezal kruh. Preko Rkoholikov sem spoznal Ano Kolar, ki mi je namignila, da "bi pa lahko vpisal doktorat iz statistike". To je bilo verjetno prvič, da sem se sploh poigral z mislijo statistike. Leta 2011 sem vpisal doktorat. Hip hip strašen trik in projekta je končno konec. Pa hvala za vse ribe.

Veliko prijateljev mi je stalo ob strani in mi pomagalo v času doktorata. Težko bi poimensko omenil vse in jim naklonil pravično mesto med zvezdami digitalne univerzitetne knjižnice. Izpostavil pa bi svojo družino, ki mi je v dobrih in slabih časih brezkomпромисno stala ob strani, za kar ji bom hvaležen do britofa. Hvala Mateji M za hitro in temeljito lektoriranje.

Brez vseh zgoraj naštetih in nenaštetih bi bil danes morebiti dosegljiv na telefonsko številko, ki jo imam v telefonu shranjeno pod "A1 zaprti" (01 587 22 67, Studenec 48, 1260 Ljubljana). Obljubim, da obstaja popolnoma racionalna razlaga, zakaj imam to številko shranjeno v telefonskem imeniku. Hvala.