

UNIVERZA V LJUBLJANI  
BIOTEHNIŠKA FAKULTETA

Jana Obšteter

**UPORABA GENETSKEGA ALGORITMA ZA OPTIMIZACIJO  
REFERENČNE POPULACIJE V GENOMSKI SELEKCIJI**

Seminarska naloga pri predmetu RAČUNSKA BIOLOGIJA  
Doktorski študij Bioznanosti

Ljubljana, 2018

## Kazalo vsebine

|   |           |
|---|-----------|
| <b>1 PREDSTAVITEV RAZISKOVALNEGA PROBLEMA IN CILJEV.....</b>  | <b>1</b>  |
| 1.1 SELEKCIJA PRI GOVEDU.....                                 | 1         |
| 1.2 NAPOVED GENOMSKIH PLEMENSKIH VREDNOSTI.....               | 1         |
| 1.3 RAZISKOVALNI PROBLEM.....                                 | 2         |
| 1.4 RAZISKOVALNI CILJI.....                                   | 2         |
| <b>2 MATERIALI IN METODE DELA.....</b>                        | <b>3</b>  |
| 2. 1 PRIPRAVA ORODJA IN PARAMETROV ZA SIMULACIJO.....         | 3         |
| 2.2 OPTIMIZACIJA REFERENČNE POPULACIJE.....                   | 3         |
| 2.2.1 <i>Genetski algoritem</i> .....                         | 3         |
| 2.2.2 <i>Optimizacija sestave referenčne populacije</i> ..... | 4         |
| 2.2.3 Napoved genomskih plemenskih vrednosti.....             | 7         |
| <b>3 REZULTATI.....</b>                                       | <b>7</b>  |
| <b>4 ZAKLJUČKI.....</b>                                       | <b>10</b> |
| <b>5 VIRI IN LITERATURA.....</b>                              | <b>10</b> |

## 1 PREDSTAVITEV RAZISKOVALNEGA PROBLEMA IN CILJEV

### 1.1 SELEKCIJA PRI GOVEDU

Selekcija stremi h genetskemu izboljšanju populacije. V živinoreji izvajamo umetno selekcijo na gospodarsko pomembne lastnosti. V govedoreji so to predvsem proizvodne lastnosti, npr. prireja mleka ali prirast mesa, lastnosti zunanosti živali in lastnosti zdravja. Ključen korak selekcije je izbira najboljših živali za starše naslednje generacije. Zaradi dolgih generacijskih intervalov, velikega vpliva nekaj izbranih moških živalih ter visokih stroškov testiranja in vzreje živali je izbira genetsko najboljših živali izrednega pomena. Hkrati pa je identifikacija teh živali težavna zaradi več virov variabilnosti. Med temi so genetska variabilnost za lastnost, velik okoljski vpliv, segregacija in rekombinacija pri potomcih že identificiranih elitnih živali ter vključenost različnih organizacij v selekcijsko delo. Živali so rangirane in odbrane na podlagi plemenskih vrednosti (PV), ki so dvakratno odstopanje potomcev od povprečja populacije (Falconer in MacKay, 1996).

Uspešnost selekcije merimo z doseženim genetskim napredkom ( $\Delta G$ ), za izračun katerega velja:

$$\Delta G = \frac{i \times r \times \sigma_A}{L} \quad (1)$$

; kjer je  $\Delta G$  = letni genetski napredek, merjen v enotah standardnega odklona na leto,  $i$  = intenzivnost selekcije,  $r$  = točnost selekcije,  $\sigma$  = genetski standardni odklon,  $L$  = generacijski interval (Lush, 1945).

V selekciji živali lahko PV napovemo na dva načina: klasičnega ali genomskega. Oboje PV napovemo z metodo mešanega modela, pri čemer je vpliv živali naključen vpliv (enačba 2). Za napoved potrebujemo vektor fenotipskih vrednosti (rešitev), matriko sistematskih vplivov in matriko sorodstva. Model za napoved je sledeč:

$$y = Xb + Za + e \quad (2)$$

; kjer je  $X$  matrika dogodkov,  $b$  je vektor ocen za sistematske vplive,  $Z$  je matrika sorodstva,  $a$  je vektor iskanih plemenskih vrednosti,  $e$  pa je vektor ostankov. Za napoved plemenskih vrednosti potrebujemo matriko sorodstva živali, ki nam omogoča izkoristek vseh razpoložljivih informacij o živalih in sorodnikih. Matriko sorodstva lahko zgradimo na podlagi rodovniških podatkov (klasična selekcija) ali na podlagi genomskih podatkov (genomska selekcija).

### 1.2 NAPOVED GENOMSKIH PLEMENSKIH VREDNOSTI

Razširitev genotipizacije domačih živali je omogočila uvedbo genomske selekcije (genotipizacija = pridobitev podatkov o genotipu na velikem številu genetskih označevalcev). V slednji selekcijo izvajamo na podlagi genomskih plemenskih vrednosti (gPV), ki jih pridobimo z regresijo fenotipskih podatkov na dejansko sorodstvo ocenjeno iz genomskih podatkov (Meuwissen in sod., 2001). V genovski napovedi ocenjujemo vplive posameznih označevalcev, na podlagi katerih napovemo gPV. Za napoved gPV zaradi večjega števila spremenljivk kot podatkov uporabimo

metodo L2 regularizacije (Ridge regresija) ali pa uporabimo Bayesovski pristop, kjer za vplive posameznih označevalcev uporabimo apriorij distribucijo (Meuwissen in sod., 2001). Genomske PV lahko napovemo vsem genotipiziranim živalim, tudi tistim brez lastnih fenotipskih vrednosti. Pri slednjih je tako točnost ocene plemenskih vrednosti znatno večja kot pri klasični selekciji. Večja točnost omogoča odbiro staršev naslednje generacije takoj ob rojstvu, kar bistveno skrajša generacijski interval in poveča genetski napredek (Schaeffer, 2006).

Napoved gPV zahteva veliko referenčno populacijo živali z genotipskimi podatki in fenotipskimi vrednostmi. Točnost napovedi gPV je v največji meri odvisna od velikosti referenčne populacije in njene sorodnosti s selekcijskimi kandidati. Večjo točnost napovedi gPV lahko dosežemo z večjo referenčno populacijo, večjo sorodnostjo napovedne in referenčne populacije ter manjšo sorodnostjo živali v referenčni populaciji (Pszczola in sod., 2012). Zaradi majhnega števila genotipiziranih in progeno testiranih bikov je v majhnih populacijah točnost napovedanih gPV selekcijskih kandidatov manjša. Prvo možno rešitev tega problema predstavlja uporaba mednarodne oz. tuje referenčne populacije (Schöpke in Swalve, 2016), katere učinkovitost je odvisna od sorodnosti živali med populacijami. Bolj kot so si živali sorodne, večja bo točnost napovedi plemenskih vrednosti. Drugi način izboljšanja točnosti napovedi gPV predstavlja genotipizacija krav. Zaradi manjše količine fenotipskih podatkov na kravo v primerjavi s progeno testiranimi biki je relativen prispevek ene krave k referenčni populaciji manjši kot enega progeno testiranega bika (de Roos, 2011). Posledično ta način prinaša manjši prispevek k točnosti ob enakem vložku sredstev. Rejske organizacije se tako srečujejo s problemom, kako z razpoložljivimi sredstvi sestaviti referenčno populacijo za maksimiranje točnosti napovedi gPV.

### 1.3 RAZISKOVALNI PROBLEM

Že prehodne študije so se ukvarjale s problemom optimizacije referenčne populacije, pri čemer so uporabile različne metode in optimizirale različne spremenljivke. Kot primer - študija Isidro in sod., 2015, je uporabile različne načine vzorčenja za maksimiranje fenotipske variance, študija Akdemir in sod., 2015, pa je uporabila genetski algoritem za rešitev kombinatoričnega problema vključitve posamezne živali v referenčno populacijo. Dobra referenčna populacija lahko bistveno izboljša točnost genomske napovedi. Optimizacija referenčne populacije je še posebej pomembna v majhnih populacijah, kje je zaradi majhnega števila živali doseganje zelenih točnosti oteženo. Majhna velikost populacije z majhnim številom progeno testiranih bikov je značilnost vseh pasem goved, ki jih redimo v Sloveniji. Zaradi malega števila moških živali s točnimi PV moramo v referenčno populacijo vključiti tudi krave. Ker pa so sredstva omejena, moramo zagotoviti, da bomo za razpoložljiva sredstva genotipizirali krave oz. živali, ki bomo zagotovile najboljšo možno točnost v dani populaciji. Trenutno pri nobeni populaciji ne izvajamo napovedi PV z lastno referenčno populacijo. Zato nas je zanimalo, ali lahko v slovenski populaciji z optimizacijo referenčne populacije izboljšamo točnost genomske napovedi in dosežemo zadostne točnosti za lastno genomsko napoved.

## 1.4 RAZISKOVALNI CILJI

Cilj naloge je bil:

- razviti orodje za optimizacijo referenčne populacije na podlagi sorodstva med živalmi v referenčni in napovedni populaciji z genetskim algoritmom,
- primerjati točnost napovedi genetskih plemenskih vrednosti z optimizirano in naključno izbrano referenčno populacijo.

## 2 MATERIALI IN METODE DELA

Delo je vključevalo tri korake, in sicer simulacijo slovenske populacije govedi, optimizacijo sestave referenčne populacije in napoved gPV z optimizirano in naključno izbrano referenčno populacijo. Shema dela je prikazana na sliki 1.

### 2.1 PRIPRAVA ORODJA IN PARAMETROV ZA SIMULACIJO

Najprej smo razvili simulator populacije govedi v selekciji. Uporabnik lahko nastavi vse selekcijske parametre, kar mu omogoča selekcijo specifične populacije govedi. Simulator je zgrajen kot Python ogrodje, ki povezuje programe za izvedbo vseh korakov enega kroga selekcije:

1. AlphaSim (Faux in sod., 2016) za stohastično simulacijo rodovnika in genoma živali,
2. blupf90 (Misztal in sod., 2002) za ocenitev genetskih plemenskih vrednosti,
3. Python program za določitev staršev nove generacije potomcev.

V naši študiji so bili vsi parametri povzeti po slovenski populaciji rjave pasme govedi, ki šteje ~30.000 aktivnih živali, od tega 10.000 krav. Simulirani genetski podatki vključujejo 10 kromosomov, iz katerih je bilo 10.000 mest izbranih za vzročna mesta za lastnost s heritabiliteto 0,25. Da smo dosegli primerno začetno strukturo populacije, smo najprej smo simulirali 20 generacij selekcijo z naključno odbiro staršev, nato pa še 20 generacij klasične selekcije (selektivna odbira).

### 2.2 OPTIMIZACIJA REFERENČNE POPULACIJE

V naslednjem koraku smo razvili orodje za optimizacijo referenčne populacije. Izbran optimizacijski algoritem je bil bil hevrističen genetski algoritem.

#### 2.2.1 Genetski algoritem

Glavna lastnosti hevrističnih optimizacijskih algoritmov je, da začnejo s poljubno rešitvijo, potem pa preko iteracij po nekem pravilu proizvajajo nove rešitve, ki jih ovrednotijo in sčasoma podajo najboljšo rešitev, ki so jo našli tekom iskanja. Iterativni proces je ponavadi ustavljen, ko: i) dano število iteracij ne uspemo izboljšati rešitve (algoritem konvergira); ii) ko je najdena rešitev dovolj dobra; iii) ko proces doseže dovoljen računalniški čas; ali iv) ko nek interni parameter konča izvajanje procesa (Maringer, 2005).

Skupina hevrističnih algoritmov je raznolika v tipu uporabljenih metod. Prav tako obstajajo različne razvrstitve hevrističnih algoritmov, ki delijo algoritme glede na različne parametre. Ena izmed bolj uporabnih je delitev glede na število sočasno uporabljenih rešitev. Tako jih delimo na trajektorne metode, ki uporabljajo eno rešitev, in populacijske rešitve, ki delujejo nad populacijo rešitev. Za trajektorne metode je značilno, da pri iskanju opišejo tirnico v iskalnem prostoru, iskalni proces populacijskih algoritmov pa opisuje evolucijo množice točk v iskalnem prostoru. Ker bomo v nalogi uporabili algoritem, ki spada med populacijske metode, se bomo podrobneje osredotočili na to skupino. Med najbolj raziskanimi populacijskimi algoritmi sta optimizacija s kolonijami mravelj (angl. *Ant-Colony Optimisation*) in evolucijsko računanje (angl. *Evolutionary computation*). Oba posnemata fenomena iz narave: medtem ko prvi posnema orientacijo mravelj na podlagi feromonov, drugi posnema proces naravne selekcije (Korošec, 2004).

Evolucijsko računanje posnema koncept evolucije oz. naravne selekcije – imamo populacijo osebkov, v kateri okoljski pritiski ustvarijo naravno selekcijo (preživetje najmočnejšega), kar povzroči naraščanje fitnesa populacije. Najprej naključno ustvarimo populacijo osebkov ter definiramo t.i. fitnes funkcijo, ki je ciljna funkcija, s katero bomo ocenili fitnes osebkov. Glede na fitnes najboljše osebkke odberemo kot starše prihodnje generacije, ki jo ustvarimo z mutacijo in / ali rekombinacijo dveh staršev. V naslednjem koraku tako starše kot tudi potomce ovrednotimo s fitnes funkcijo in le najboljše prenesemo v naslednjo generacijo. Vsaka iteracija algoritma ustvari novo generacijo osebkov, vsak osebek pa predstavlja eno možno rešitev. Potek evolucijskega algoritma lahko predstavimo z naslednjo shemo:

```

inicializiraj populacijo (naključen proces)
    osebki (kandidatne rešitve)
ovrednoti vse osebkke (s fitnes funkcijo)
DOKLER ne STOP delaj:
    izberi starše iz populacije
    ustvari potomce z mutacijo in rekombinacijo staršev
    ovrednoti novorojene potomce
    zamenjaj nekatere izmed staršev s potomci
  
```

Uspeh evolucijskih algoritmov je tako močno odvisen od upravljanja populacije. V splošnem sta v evolucijskem algoritmu dve gonilni sili: selekcija, ki si prizadeva za kakovost in zmanjšuje genetsko variabilnost populacije, in variacija – implementirana kot mutacija in rekombinacija - ki si prizadeva za novosti in povečuje genetsko variabilnost. Zato je uravnoteženje teh dveh sil ključno za uspeh evolucijskega algoritma (Eiben in Schoenauer, 2005).

### 2.2.2 Optimizacija sestave referenčne populacije

V tej raziskavi smo za implementacijo genetskega algoritma izbrali Python program Pyevolve (Perone, 2009). Ciljna funkcija je optimizirala sorodstvo med živalmi. Med kandidatne

živali za referenčno oz. trening populacijo smo vključili krave zadnjih sedem generacij, t.j. 31 – 37 (26.131 živali) in vse progeno testirane bike (96 živali), kar je skupaj štelo 26.227 živali. Izbrali smo tudi napovedno oz. testno populacijo, v kateri smo preverjali točnost napovedi genomskih plemenskih vrednosti. Slednja je štela 90 novorojenih osebkov (generacija 40) brez lastnih meritev. Za vseh 26.317 živali smo s programom AlphaRelate ustvarili matriko sorodstva na podlagi rodovniških podatkov, ki je služila za optimizacijo sorodstva.

V koraku optimizacije smo optimizirali le odbiro krav za referenčno populacijo, vse moške živali pa smo nato izbranim kravam avtomatsko dodali. Za izvedljivost komputacije smo ~10,000 krav razvrstili v 100 čred, za kar smo uporabili metodo voditeljev (angl. *k-means clustering*) na podlagi podatkov o materi ter podatkov o fenotipski in pravi genetski vrednosti. Uporabljena metoda voditeljev je implementirana v R-ovem paketu 'stats'. Črede tako niso štele enako število krav, temveč so štele od 45 do 205 krav, s povprečno velikostjo črede 107 krav. Za skrajšanje optimizacijskega časa smo pred optimizacijo ustvarili tabelo sorodstva med vsemi možnimi kombinaciji čred ter med vsemi čredami krav in moškimi živalmi / napovedno populacijo.

Ena možna rešitev optimizacije / en osebek / en kromosom je tako predstavljal vektor s sto elementi 0 ali 1, ki je podal informacijo o vključitvi vsake izmed čred krav (1 – 100) v referenčno populacijo. Izbranim čredam krav smo nato v referenčno populacijo dodali še moške živali. Ciljna funkcija je najprej za vsako rešitev oz. kromosom izračunala sorodstvo med živalmi v referenčni populaciji (izbrane črede + moške živali) ter sorodstvo živali med živalmi v referenčni in napovedni populaciji. Pri tem smo želeli doseči čim večje sorodstvo med živalmi v referenčni in napovedni populaciji (max) ter čim manjše sorodstvo samih živalih v referenčni populaciji (min) – zato je funkcija maksimirala razliko med slednjima. Za ta namen smo na razliki uporabili t.i. reLu funkcijo, ki vrednosti funkcije pripiše 0, če je vrednost negativna, če pa je vrednost nenegativna, pa vrednosti funkcije ne spreminja. Dobljeno razliko smo **tudi kvadrirali, saj je bilo pokazano, da eksponentne funkcije dosežejo boljši rezultat kot linearne.**

Prav tako je program omogočal nastavitve omejitve za število živali, kar smo v tej raziskavi omejili na 5000, pri čemer so bile rešitve oz. kromosomi, ki so za več kot 15 % odstopale od tarčnega števila izbranih krav, penalizirane tako, da je bila vrednost teh rešitev 0.

Pseudo koda za fitnes funkcijo:

```
def fitnes_fukncija(kromosom):
    velikostReference = uporabniško-določena vrednost za želeno število
    živali v referenci
    steviloZivali = sum(število živali v ?redah izbranih v
    optimizaciji)

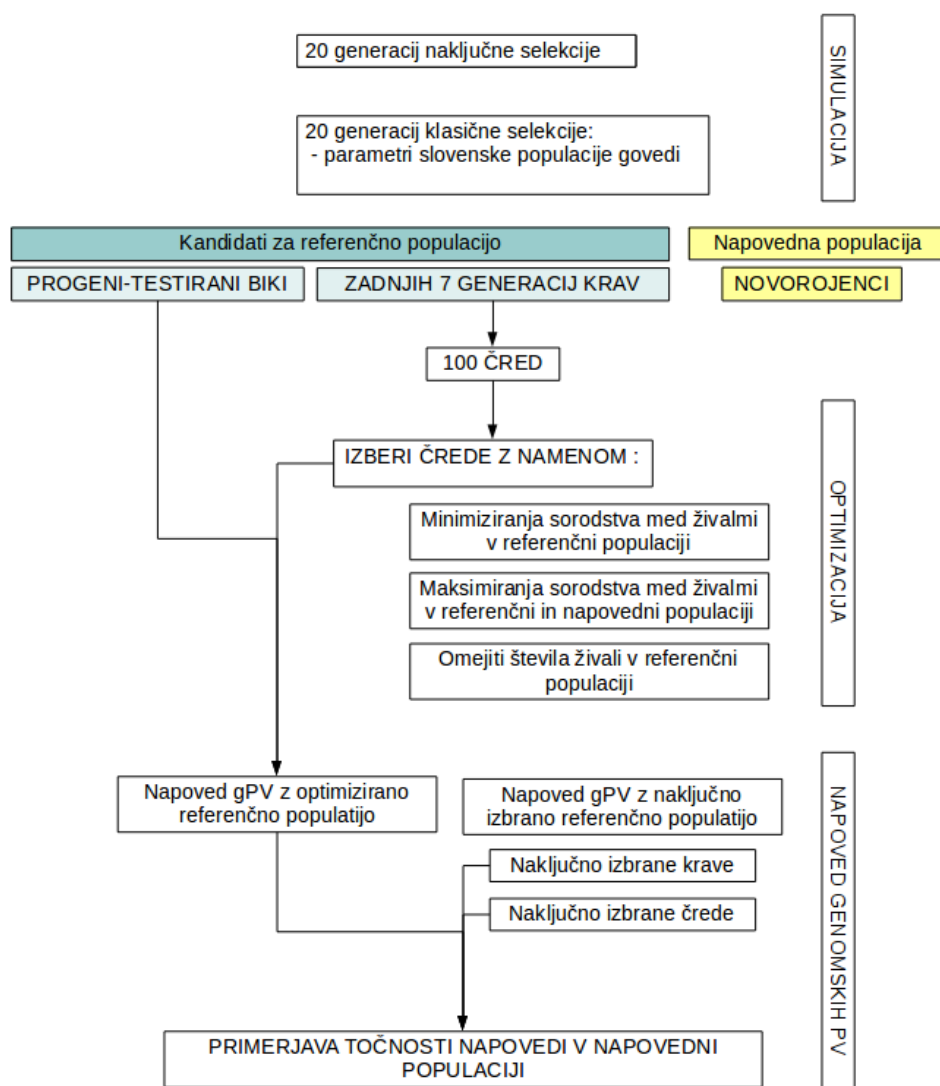
    sorodstvoTrening = mean(povprečno sorodstvo živali v referenčni
    populaciji)
    sorodstvoTreningTest = mean (povprečno sorodstvo med živalmi v
    referenčni in napovedni populaciji)
```

```

razlika = medA- znotrajA
fitnes = (reLu(razlika)) **2
penal = [-fitnes if (steviloZivali > 1.5*velikostReference or
steviloZivali < 0.85*velikostReference ) else 0]
return fitnes+penal

```

Za korak mutiranja smo izbrali IntegerBinary mutator iz paketa Pyevolve, ki naključno spreminja 0 in 1 elemente kromosoma oz. rešitve. Stopnjo mutacije (angl. *mutation rate*) smo nastavili na 0,01, stopnjo prekrivanja (angl. *cross-over rate*) pa na 0.001. Za izbiro staršev naslednje generacije smo izbrali »Tournament selection«, ki za starše izbere najboljše osebkke. Velikost populacije smo nastavili na 50 in število generacij na 900. Kot kriterij za konec optimizacije smo nastavili kriterij konvergence – kar pomeni, da je optimizacija končana, ko se določeno število iteracij rešitev ne izboljša oz. algoritem konvergira.



Slika 1: Shema dela.



Optimizacijo smo izvedli v desetih ponovitvah. V vsaki ponovitvi smo k optimizirani referenčni populaciji izbrali tudi dve naključno izbrani referenčni populaciji:

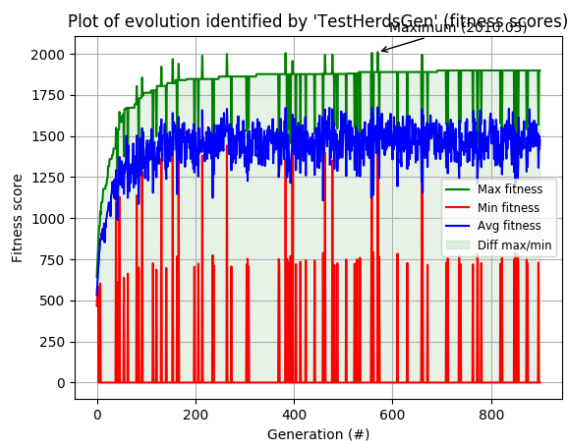
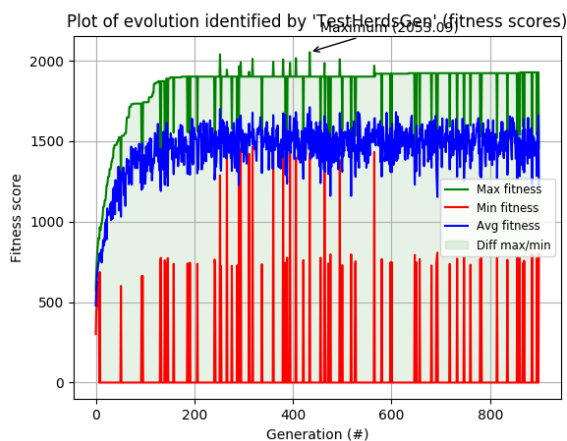
- a) naključna izbira krav iz nabora 26.131 krav, pri čemer je bilo število izbranih krav enako nastavljeni omejitvi za število živali v referenčni populaciji;
- b) naključna izbira čred krav, pri čemer je bilo število izbranih čred enako številu izbranih čred z optimizacije.

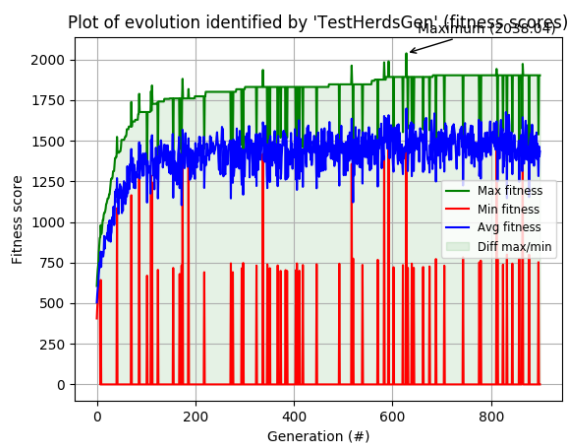
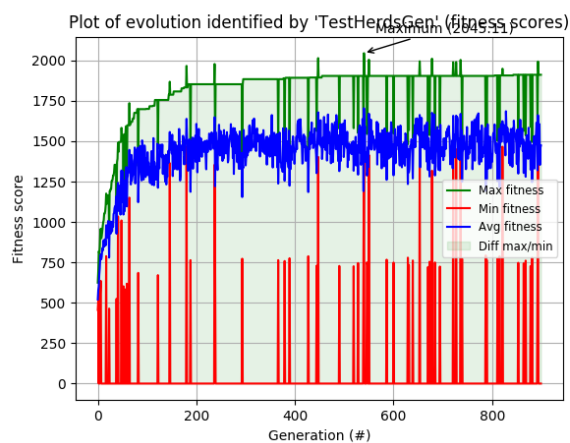
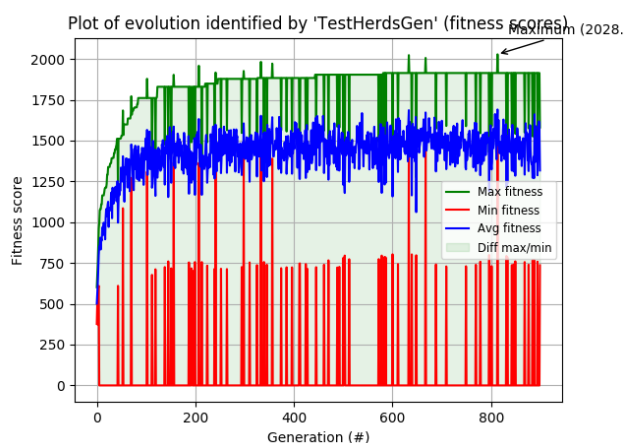
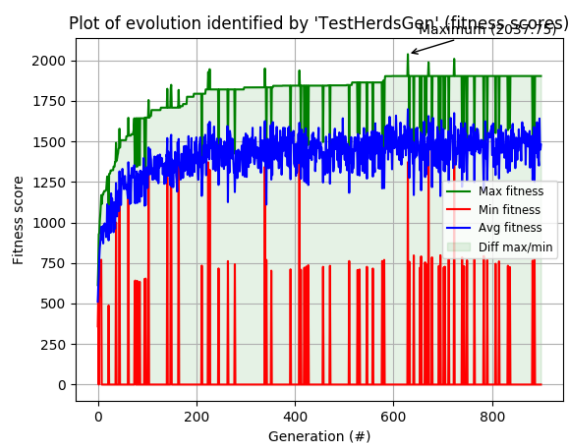
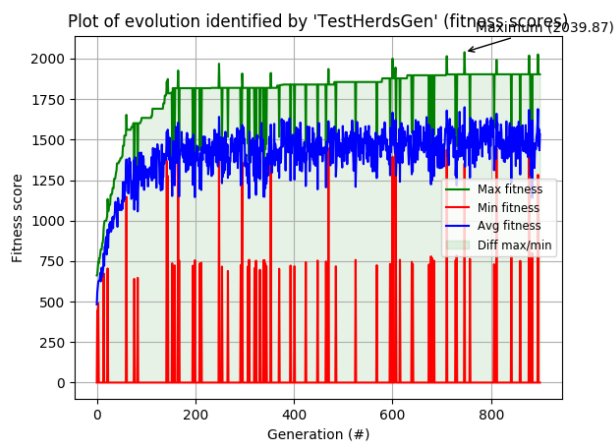
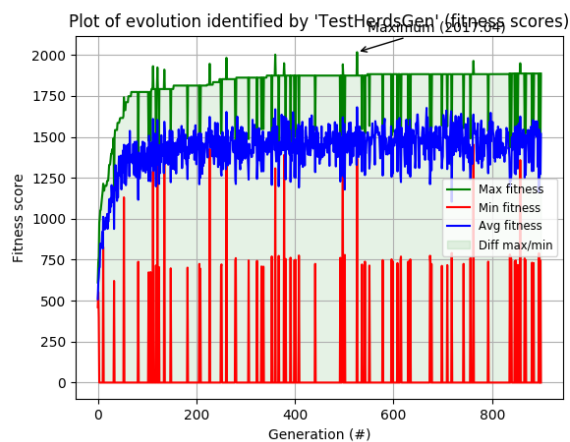
### 2.2.3 Napoved genetskih plemenskih vrednosti

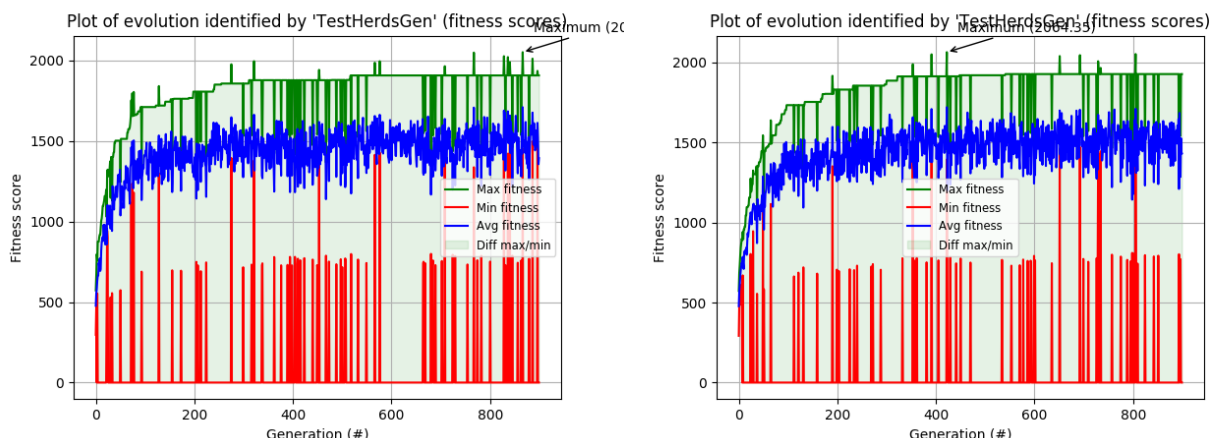
V vsaki od desetih ponovitev smo z vsako izmed optimiziranih in naključno izbranih referenčnih populacij napovedali gPV v napovedni populaciji (90 osebkov). Za napoved gPV smo uporabili program blupf90 (Misztal in sod., 2002). Točnost napovedi smo definirali kot korelacijo med pravimi genetskimi vrednostmi in napovedanimi gPV.

## 3 REZULTATI

Na sliki 2 je prikazano spreminjanje povprečnega fitnesa populacije skozi iteracije algoritma za vseh 10 ponovitev optimizacije. Vidimo, da v vseh ponovitvah povprečen fitnes populacije pri okoli 1500. V vseh optimizacijah pa vidimo tudi, da maksimalen fitnes populacije veliko bolj variira – pri čemer je bil maksimalen fitnes posameznega kromosoma dosežen v eni izmed vmesnih iteracij (slika 2).

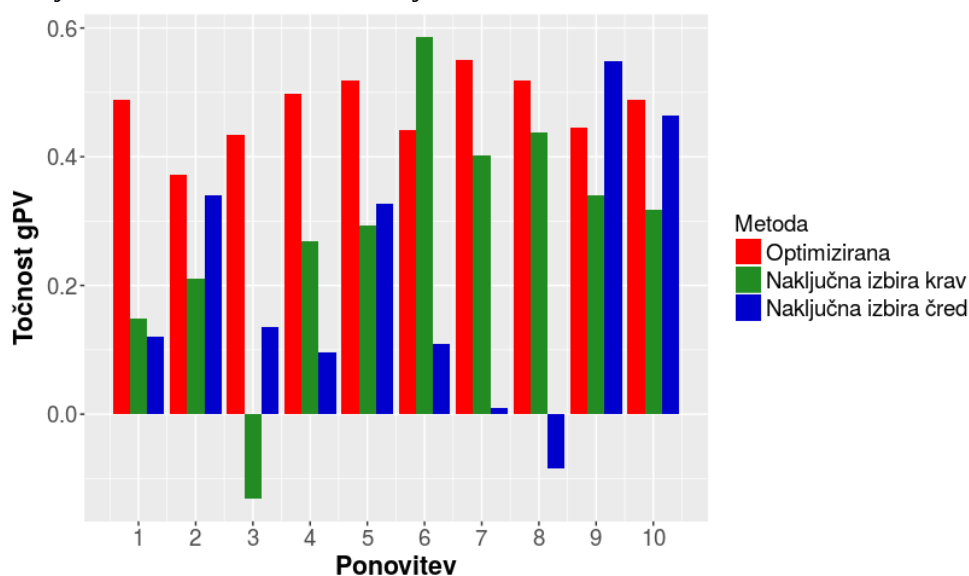






Slika 2: Spreminjanje povprečnega fitnesa populacije rešitev skozi iteracije genetskega algoritma v desetih ponovitvah.

Po optimizaciji smo napovedali gPV z optimizirano ali naključno izbrano referenčno populacijo. Točnosti gPV dosežene v vsaki izmed ponovitev, izračune kot korelacija med pravimi genetskimi vrednostmi in gPV, so prikazane na sliki 3. Vidimo, da smo v osmih od desetih ponovitev z optimizirano referenčno populacijo dosegli večjo točnost napovedanih gPV v napovedni oz. testni populaciji. Izjema sta ponovitvi 6 in 7, kjer z optimizirano referenco dosežemo manjše točnosti kot z referenco naključno izbranih krav oz. naključno izbranih čred.



Slika 3: Točnost napovedanih genomskih plemenskih vrednosti z optimizirano in naključno izbrano referenčno populacijo v vsaki izmed desetih ponovitev. Optimizirana: z genetskim algoritmom optimizirana izbira čred krav za referenčno populacijo; naključno izbrane krave: naključna odbira krav za referenčno populacijo; naključno izbrane črede: naključna odbira čred krav za referenčno populacijo.

V preglednici 1 pa so prikazane povprečne točnosti v desetih ponovitvah optimizacije. Vidimo, da v povprečju z optimizirano referenčno populacijo dosežemo 130 % večjo točnost napovedi gPV kot z

referenco naključno izbranih krav ter 65 % večjo točnosti kot z referenčno populacijo naključno izbranih čred. Pri tem je bila povprečno število izbranih čred v optimizaciji 36, povprečna velikost optimizirane referenčne populacije 4265, velikost reference z naključno izbranimi kravami je bila stalna, in sicer 5096, povprečna velikost referenčne populacije z naključno izbranimi čredami pa je bila 3853.

Preglednica 1: Povprečna točnost napovedi genomskih plemenskih vrednosti z optimizirano ali naključno izbrano referenčno populacijo.

| Metoda                | Povprečna točnost |
|-----------------------|-------------------|
| Optimizirana          | 0,475             |
| Naključna izbira krav | 0,287             |
| Naključna izbira čred | 0,207             |

Optimizirana: z genetskim algoritmom optimizirana izbira čred krav za referenčno populacijo; naključno izbrane krave: naključna odbira krav za referenčno populacijo; naključno izbrane črede: naključna odbira čred krav za referenčno populacijo.

## 4 ZAKLJUČKI

Majhne populacije se soočajo s problemom uvedbe genomske selekcije zaradi majhnega števila živali, ki ima za posledico točnosti gPV, ki so prenizke za uporabo gPV v selekciji. Prav zato je optimizacija sestave referenčne populacije še posebnega pomena prav v majhnih populacijah. Z optimizacijo lahko namreč dosežemo zadostne točnosti, ki nam omogočijo uporabo gPV za odbiro živali.

Že prehodne študije so razvile orodja za optimizacijo referenčne populacije, vendar pa kombinacija metod predstavljena v tem delu predstavlja nov pristop k problemu. V tej raziskavi smo za optimizacijo uporabili genetski algoritem, ki teži k minimiziranju sorodstva med živalmi v referenčni populaciji in maksimiranju sorodstva med živalmi v referenčni (trening) in napovedni (testni) populaciji. Pokazali smo, da lahko z optimizacijo referenčne populacije za do 130 % povečamo točnosti gPV v primerjavi z napovedjo z naključno izbrano referenčno populacijo.

Še vedno pa se povprečna točnost giblje okoli 48 %, kar je premajhna točnost za praktično uporabo v selekciji. Po drugi strani pa je bila omejitev 5000 živali za referenčno populacijo zelo stroga in tudi v praksi s takšno referenčno populacijo ne bi pričakovali dejanskega uspeha. Optimizacijo bi tako morali ponoviti z večjim številom za dovoljeno število živali.

V našem delu smo se ukvarjali le z napovedjo gPV v naslednji (eni) generaciji. Selekcija pa je kontinuiran proces, v katerem živali iz napovedne populacije po pridobitvi fenotipskih podatkov vstopijo v referenčno populacijo. Pri tem dosežemo ravno nasprotno učinek od zelenega, saj je bila referenčna populacija izbrana z namenom maksimiranja sorodstva s živalmi v napovedni populaciji

– ko pa se le-te pridružijo referenčni populaciji, pa želimo, da bi bilo to sorodstvo čim manjše. V prihodnjem delu bi lahko tako z razvitim orodjem poskušali nasloviti tudi ta problem.

## 5 VIRI IN LITERATURA

- Akdemir D., Sanchez J. I., Jannink J.-L. 2015. Optimization of genomic selection training populations with a genetic algorithm. *Genetics Selection Evolution*, 47, 1: 38.
- de Roos A.P.W. 2011. Genomic selection in dairy cattle, doktorsko delo. Wageningen University, Wageningen, Nizozemska: 184 str.
- Eiben A. in Schoenauer M. 2005. Evolutionary Computing. *Information Processing Letters*. 82, 1: 1-6.
- Falconer D.S., Mackay, T.F.C. 1996. *Introduction to Quantitative Genetics*. 4. izdaja. Harlow, UK, Longman: 464 str.
- Faux A.-M., Gorjanc G., Gaynor R.C., Battagin M., Edwards S.M., Wilson D.L., Hearne, S.J., Gonen S., Hickey J.M. 2016. AlphaSim: Software for Breeding Program Simulation. *Plant Genome*, 9, 3.
- Isidro J., Jannink J.L., Akdemir D., Poland J., Heslot N., Sorrells M. E. 2015. Theoretical and Applied Genetics, 128, 1: 145.
- Korošec P. 2004. Magistrska naloga: Metahevristično reševanje optimizacijskega problema s kolonijami mravelj. Ljubljana, Univerza v Ljubljani, Fakulteta za računalništvo in informatiko: 80 str.
- Lush J.L. 1945. *Animal Breeding Plans*. 3. izdaja, Ames, Iowa, ZDA, Iowa State University Press: 442 str.
- Maringer D. 2005. *Portfolio Management with Heuristic Optimization*. New York, ZDA, Springer-Verlag New York: 38 – 76.
- Meuwissen T.H.E., Hayes B.J., Goddard M.E. 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics*, 157, 4: 1819–1829.
- Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T., Lee D.H. 2002. BLUPF90 and re-lated programs (BGF90). V: *Proc. 7th World Congress on Genetics Applied to Livestock Production. WCGALP, Montpellier, France, 19-23 avg. 2002. Montpellier, WCGALP: 1-2.*
- Perone C. S. 2009. Pyevolve: a Python open-source framework for genetic algorithms. *ACM SIGEVOlution*, 4, 1: 12 – 2.0
- Pszczola M., Strabel T., Mulder H.A., Calus M.P.L. 2012. Reliability of direct genomic values for animals with different relationships within and to the reference population. *Journal of Dairy Science*, 95, 1: 389–400.
- Schaeffer L. R., 2006. Strategy for applying genome-wide selection in dairy cattle. *Journal of Animal Breeding and Genetics*, 123, 4: 218-223.

Obšteter, J. Uporaba genetskega algoritma za optimizacijo referenčne populacije v genovski selekciji. Ljubljana, Univ. v Ljubljani, Biotehniška fakulteta, 2018

---

Schöpke K., Swalve H.H. 2016. Review: Opportunities and challenges for small populations of dairy cattle in the era of genomics. *Animal*, 10, 6: 1050–1060.