

Civilizacije

Anže Marinko, Klara Golob, Ema Jemec in mentor dr. Matjaž Gams
 Inštitut Jožef Stefan (E9 - oddelek za inteligentne sisteme), Ljubljana
 Samo za interno uporabo
 maj 2020

CONTENTS

I	Uvod	1
II	Osnovni modeli in njihova posplošitev	1
II-A	Model 1	1
II-B	Model 2	2
II-C	Model 3	2
II-D	Model 4	2
II-E	Implementacija	2
III	Gručenje	2
III-A	PCA - Principal component analysis . .	3
III-B	Analiza gru	3
IV	Rezultati	3
V	Razprava	3

I. UVOD

Oceniti želimo porazdelitev časa, ki ga inteligentna civilizacija preživi v naši galaksiji (in posplošeno v celotnem vesolju). Civilizacija se šteje za inteligentno, če je sposobna radio-komunikacije z drugimi planeti. Porazdelitev ocenjujemo z uporabo Drakeove enačbe. Želimo pa zajeti čim večji prostor relevantnih modelov. Relevantni modeli so tiste porazdelitve, ki ne dajejo prevelike verjetnosti temu, da bi inteligentna civilizacija obstajala več milijonov let in ki imajo s časom padajočo gostoto, torej da je bolj verjetno, da bi obstajali (bili sposobni radio-komunikacije) le tisoč let kakor pa med tisoč in dvatisoč let.

Pripravili smo štiri bolj splošne modele, katerim smo spremenjali parametre, da bi dobili čim več različnih porazdelitev, ki ustrezajo Drakeovi enačbi. Vse te porazdelitve smo potem gručili po podobnosti in dobili nekaj značilnih porazdelitev, ki jih lahko posamično obravnavamo in izključujemo z razvojem znanosti in vedenja o vesolju. To nam pomaga oceniti, koliko let obstoja še preostaja naši civilizaciji na tem planetu. Če je verjetnost, da obstajamo še dolgo, majhna, se moramo toliko bolj potruditi, da ne propademo že prej.

Vsa koda je pripravljena v programskem jeziku Python 3, kakih 470 vrstic urejene kode s komentarji.

II. OSNOVNI MODELI IN NJIHOVA POSPLOŠITEV

V nadaljevanju vam bomo predstavili razširjen Sandbergov model in dva druga modela dobljena s spremembo prvega ter četrti model po teoriji redke zemlje. Vsak model pa je

odvisen od nekaj parametrov, za katere znamo ceniti spodnjo in zgornjo mejo, ne vemo pa njihove dejanske vrednosti ali celo porazdelitve na tem intervalu. Za razliko od Sandberga bomo namesto fiksnih vrednosti uporabili različne porazdelitve na izbranem intervalu. Vsak parameter porazdelimo z eno izmed petih izbranih porazdelitev ter tako dobimo možnih izbranih porazdelitev $5^{\#parametrov}$ za vsako zgornjo mejo za N za vsak model.

Izbrane porazdelitve so:

- 1) loguniformna (uniformno porazdeljeno glede na x os v logaritski skali),
- 2) uniformna (enakomerno po celem intervalu),
- 3) pol Gaussova (polovica Gaussove porazdelitve z vrhom pri spodnji meji),
- 4) lognormal (normalna porazdelitev vzdolž x osi v logaritski skali s povprečjem na sredini intervala),
- 5) fiksna (fiksirana vrednost na zgornji meji).

Poleg tega ne poznamo zgornje meje za število inteligentnih civilizacij v nekem trenutku N . Edino, kar vemo, je, da obstaja vsaj ena civilizacija (to smo mi) in domnevamo, da inteligentnih civilizacij v naši galaksiji ni več kot 10 000, sicer bi kakšne druge civilizacije že videli ali vsaj zaznali njihove radijske valove, kar bi šteli za inteligentno civilizacijo.

A. Model 1

Imamo 6 parametrov:

- 1) R_* stopnja nastanka nove zvezde, kar je porazdeljeno nekje med 1 in 100 na leto
- 2) f_p verjetnost, da ima zvezda vsaj en planet, je porazdeljena med 10 in 100 odstotki
- 3) f_e verjetnost, da je ta planet podoben zemlji, je nekje med 10 in 100 odstotki
- 4) f_i verjetnost, da se na zemlji podobnem planetu razvijejo inteligentna bitja, je porazdeljena med 0,1 in 100 odstotki
- 5) f_c verjetnost, da ta bitja naredijo civilizacijo sposobno radio-komunikacije z drugimi planeti, je med enim in 100 odstotki
- 6) N - število takih civilizacij v naši galaksiji porazdeljena med 1 in izbrano zgornjo mejo za število civilizacij

Število let obstoja civilizacije L je po Drakeovi enačbi enako $L = N/(R_*f_p f_e f_i f_c)$. Ker so vsi parametri slučajne spremenljivke je tudi L slučajna spremenljivka, katere porazdelitev empirično dobimo z veliko generiranimi primeri.

Tako dobimo za izbrane porazdelitve in zgornjo mejo za N možno porazdelitev, torej 5^6 različnih porazdelitev za vsako

izbrano zgornjo mejo za N . Enako velja tudi za preostale modele.

B. Model 2

V drugem modelu zberemo nekatere parametre iz prvega modela skupaj tako, da imamo le tri parametre. Za razliko od prvega modela, ki po Drakeovi enačbi računa glede na število inteligentnih civilizacij v tem trenutku, drugi model uporablja namesto stopnje nastanka nove zvezde kar število vseh zvezd. Parametre pa združi zgolj v tri:

- 1) astro-fizična verjetnost f_a porazdeljena lognormalno na sredini intervala od ob robovih pa z izbrano porazdelitvijo
- 2) bio-tehnična verjetnost f_b porazdeljena med 10^{-11} in 0.1 odstotka,
- 3) N - število takih civilizacij v naši galaksiji porazdeljena med 1 in izbrano zgornjo mejo za število civilizacij.

Tu je L ponovno izračunan kot $N/(f_a + f_b)$.

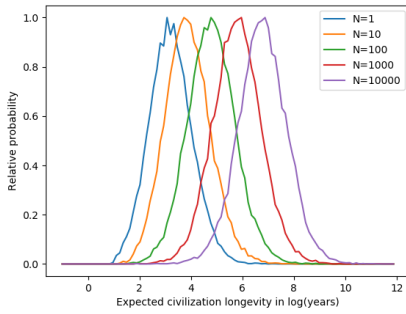


Fig. 1. Histogrami v logaritmiki skali za prvi model, kjer so vsi parametri porazdeljeni loguniformno.

C. Model 3

Modelu 1 dodamo možnost širjenja na druge planete. Izračunamo parametre za model 1: N in $f := R_* f_p f_e f_i f_c$, kjer je f stopnja nastanka nove inteligentne civilizacije.

$E = 5.1334 * 10^{10+f_p+f_e} S$ je ocenjeno število zemlji podobnih planetov, kjer je $S = 4.7233 * 10^{-42}$ gostota zvezd glede na Wikipedijo.

Iščemo ničle polinoma $p(L) = EfL^4 + fL - N$ oz. rešitev enačbe $f * (L + EL^4) = N$, kjer je L pozitivna realna ničla polinoma oz. pozitivna realna rešitev enačbe.

D. Model 4

Poleg tega je že v pripravi nov model (tega še ni v naslednjem poglavju), ki veliko bolj pesimistično ocenjuje vrednosti parametrov in pravi, da je razvoj tovrstne civilizacije praktično nemogoč, torej da smo skoraj gotovo edini v vesolju. Ker pa ne vemo, kaj bi bilo realistično, ne vemo ali to tudi zares je preveč pesimistično. Ta teorija se imenuje Rare Earth theory oz. Rare Earth hypothesis. Pravi, da je Zemlja zelo edinstven in redek planet zaradi vseh svojih lastnosti, ter je zaradi tega malo njej podobnih planetov. Število let L računamo s pomočjo Rare

earth enačbe: $N = N_* \times n_g \times f_p \times f_{pm} \times f_i \times f_c \times f_l \times f_m \times f_j \times f_{me}$ in Drakeove enačbe. Kjer so naslednji še neznani parametri:

- 1) N_* je število zvezd v naši galaksiji, med 250 in 150 milijard.
- 2) n_g število planetov v galaktični naseljitveni coni, približno 40 bilijonov.
- 3) f_{pm} delež planetov, ki so bogati s kovinam, med 1 in 10 odstotki.
- 4) f_m delež naseljivih planetov, z veliko luno, med 0.3 in 3 odstotki.
- 5) f_j delež planetarnih sistemov z velikim jupitrovskim planetom, med 5 in 10 odstotki.
- 6) f_{me} delež planetov z zelo niskom številom izumirajočih dogodkov, med 1 in 10 odstotki.

Število let obstoja civilizacije smo z izpeljavo izračunali po naslednji formuli:

$$L = \frac{N_* \times f_{pm} \times n_g \times f_m \times f_j \times f_{me}}{R_*}$$

E. Implementacija

Za pripravo vsakega histograma za vsako kombinacijo porazdelitev in zgornje meje za N smo generirali 200 000 točk in jih združili v histograme z intervalom razrezanim na 100 delčkov. Izračunali smo histogram, kjer smo imeli linearno in logaritmico skalo na osi L .

Izračunali histograme za vse možne kombinacije porazdelitev pri zgornjih mejah za število inteligentnih civilizacij N nastavljenih na 1, 3, 10, 30, 100, 300, 1000, 3000 in 10 000.

Model 1 s porazdelitvijo (0, 0, 0, 0, 0, 0) v logaritmiki skali

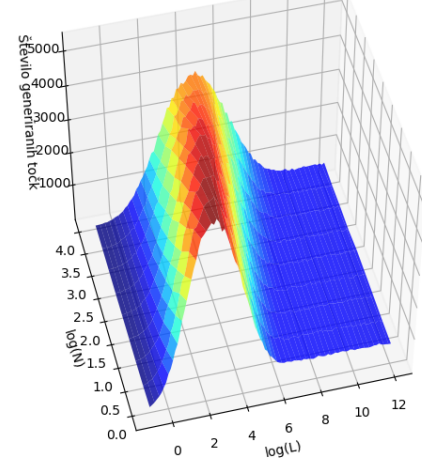


Fig. 2. Histogrami v logaritmiki skali za prvi model, kjer so vsi parametri porazdeljeni loguniformno.

Primer dobljenih porazdelitev lahko vidimo na sliki 1.

Histograme smo naredili na intervalu $[0, 10^10]$ pri linearni skali in $[-1, 12]$ pri logaritmiki skali, ker večje vrednosti niso realne in so le teoretične dobljene zaradi posplošitev v modelu.

Pri generiranju podatkov smo izkoristili lastnost, da model 1 izračunamo tekom generiranja modela 3. Problem dopušča

popolno paralelizacijo računanja vsakega histograma na svojem procesorju. V našem primeru smo uporabili 10 procesorskih enot. Histograme smo si tekom izračunavanja zapisovali v datoteke in jih na koncu združili v 4 datoteke. Posebej linearne in logaritsmske histograme in za vsake histograme še sezname parametrov v drugi datoteki.

III. GRUČENJE

Vse dobljene histograme želimo posplošiti na le nekaj značilnih histogramov, ki nam povedo nekaj možnih porazdelitev vrednih razprave. Če želimo tudi vizualno prikazati množico histogramov, kjer bodo gruče kar se da opazne, moramo izrisati le tri dimenzije izmed vseh 100 (dolžina histograma). V nadaljevanju vam bomo predstavili transformacijo prostora tako, da bo slika kar se da povedna. Torej iščemo novo bazo za prostor histogramov.

A. PCA - Principal component analysis

Izračunamo korelacijsko matriko za vse histograme (dobimo 100×100 matriko) in izračunamo lasne vrednosti ter lastne vektorje korelacijske matrike. Vse histograme zamaknemo za povprečen histogram in jih preslikamo z matriko iz lastnih vektorjev.

Novo dobljene točke so v lastnem prostoru in če so lastne vrednosti urejene padajoče, so prve dimenzije najbolj informativne, ker nosijo največ variance. Velikost lastne vrednosti v primerjavi z vsoto lastnih vrednosti nam pove, koliko variance v podatkih nam ta dimenzija prinese. Torej vzamemo le največje tri lastne vrednosti za izris. Lastni vektorji pa so baza prostora. Inverzna transformacija je enostavna.

PCA ne vpliva na podobnost med histogrami (razdalje se pri rotaciji in translaciji ohranjajo), ampak se le znebi linearne odvisnosti med komponentami.

B. Analiza gruč

Z uporabo algoritma k-means, ki točke iterativno gruči na podlagi k povprečij gruč, histograme razvrstimo v izbrano število gruč (smiselno je do 10). Vsako gručo analiziramo v smislu porazdelitve osnovnih modelov in izbranih zgornjih mej za N . Porazdelitve izpišemo na sledeč način:

Na primer za neko gručo:

Model		1	2	3	
		523	425	3	
$\log(N_{max})$	0	0.5	1	1.5	2
	25	45	60	81	110
$\log(N_{max})$	2.5	3	3.5	4	
	148	177	165	140	

Poleg tega za vsako gručo izračunamo povprečno porazdelitev f_i (in uporabimo inverzno transformacijo za izris), njeno preživetveno funkcijo ($G(t) = 1 - \int_0^t f_i(s)ds$) in konveksno ovojnico gruče, da lahko izberemo še nekaj robnih primerov iz gruče.

Ob izrisu prvih treh dimenzij transformiranega prostora histogramov s PCA točke obarvamo glede na osnovni model in velikost pike določimo kot $\log(N_{max})$. Na sliko ne izrišemo

vseh točk, da se lepše vidi in ker ima model 2 občutno manj primerov jih pa vse uporabimo pri računanju PCA, gruči in aproksimacije točk v 3-dimenzionalnem prostoru s ploskvijo. Aproksimiramo z linearno regresijo na podlagi prvih dveh dimenzij na tretjo dimenzijo. Implementirano imamo do ploskev stopnje 3, torej za izbrano gručo izračunamo koeficiente b za ploskev:

$$x_3 = b_0 + b_1x_1 + b_2x_2 + b_3x_1x_2 + b_4x_1^2 + b_5x_2^2 + b_6x_1^2x_2 + b_7x_1x_2^2 + b_8x_1^3 + b_9x_2^3.$$

Izračunano ploskev tudi narišemo na 3-dimenzionalno sliko, kjer so točke in ploskve obarvane po gručah.

IV. REZULTATI

Oglejmo si nekaj zanimivih rezultatov. Ker je poročilo predvsem namenjeno diskusiji, vsebuje tudi nekatere slike, ki jih ne bomo dovolj dobro predstavili in so namenjene predvsem bralcu za lastna opažanja. Najprej si oglejmo gručenje in analizo gruč histogramov v linearni skali. Slika 2 nam prikazuje, kako pomembna je kakšna dimenzija transformiranega prostora, kar vidimo iz velikosti lastne vrednosti. Opazimo, da je že druga sploh pa tretja lastna vrednost zelo majhna v primerjavi s prvo torej nam ta dimenzija prinese največ variance. Kumulativno nam prve tri dimenzije razložijo že skoraj vso varianco med histogrami. Graf desno spodaj nam prikazuje lastne vektorje za prve 4 lastne vrednosti, torej bazo transformiranega prostora.

Slike 3, 5, 8 in 10 nam prikazujejo oblak točk obarvan po modelih (levo) in po gručah (desno). Gruče so tudi aproksimirane s ploskvijo. Opazimo, da so gruče razdeljene predvsem vzdolž prve komponente pri linearnih histogramih, pri logaritsmskih pa dobimo zanimive luske.

Na slikah 4 in 9 vidimo povprečne histograme in tri histograme s konveksno ovojnico vsake od gruč. Hitro opazimo osnovne lastnosti gruč.

Na slikah 6 in 11 vidimo povprečja gruč in njihove preživetvene funkcije, ki nam za vsako gručo povedo, kakšno vverjetnost nam gruča pripisuje, da preživimo konkretno število let.

Tudi slika 7 nam prikazuje, kako pomembna je kakšna dimenzija transformiranega prostora, a tokrat logaritsmskih histogramov. Opazimo, da tudi v tem primeru lastne vrednosti hitro padajo proti nič, a nam prve tri dimenzije povejo manj o celotni varianci podatkov.

V. RAZPRAVA

Želimo zapolniti še luknje v tem oblaku točk, da bi dobili čim bolj celotno podobo možnih histogramov. V nadaljevanju bomo dodali še peti model, ki je povprečje modela 1 in 3 pri istih porazdelitvah in isti zgornji meji za N ter eno izmed izmed porazdelitev iz modela 2 upoštevajoč porazdelitev iz prvih dveh modelov.

Šesti model bi bil za vsako porazdelitev iz modela 2 povprečje z vsemi histogrami iz modela 1 in 3 s primerljivimi porazdelitvami. Če pa kakšen izmed novih dveh modelov ne prinaša dovolj nove informacije oz. le zgosti porazdelitev točk

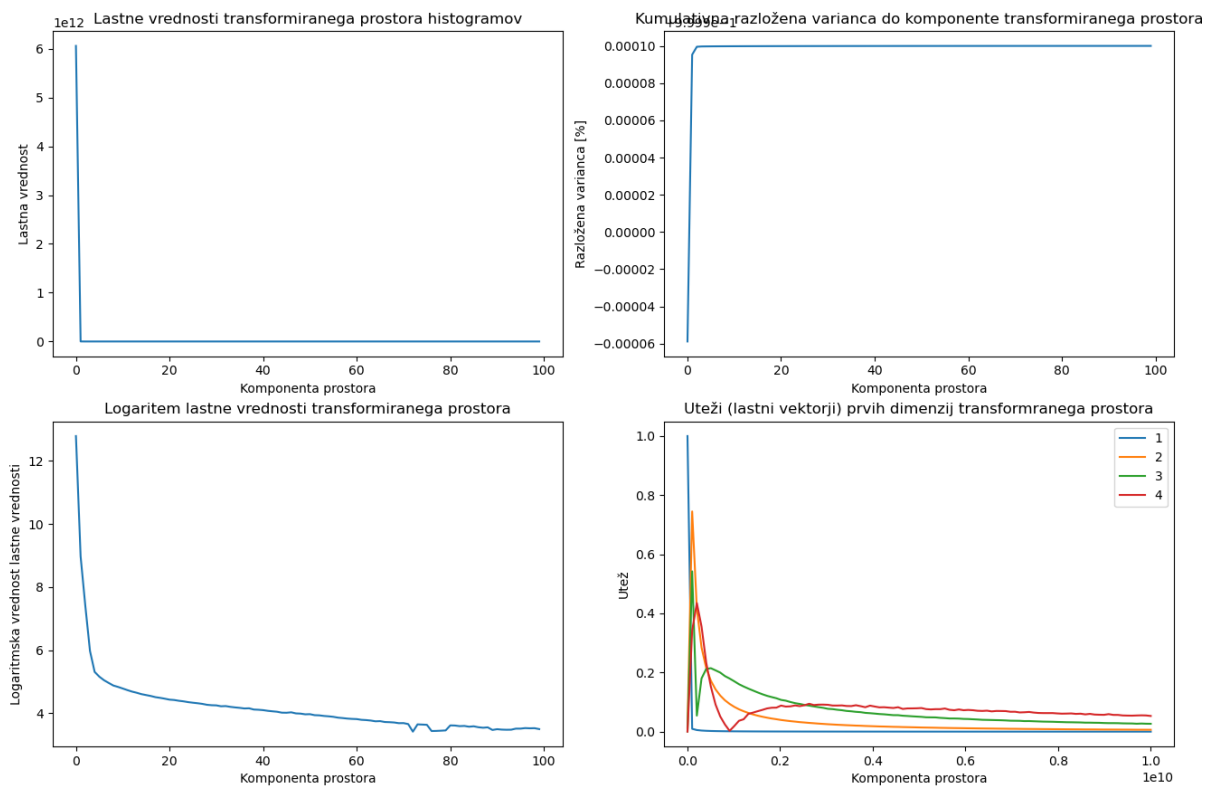


Fig. 3. Lastne vrednosti in lastni vektorji prostora linearnih histogramov.

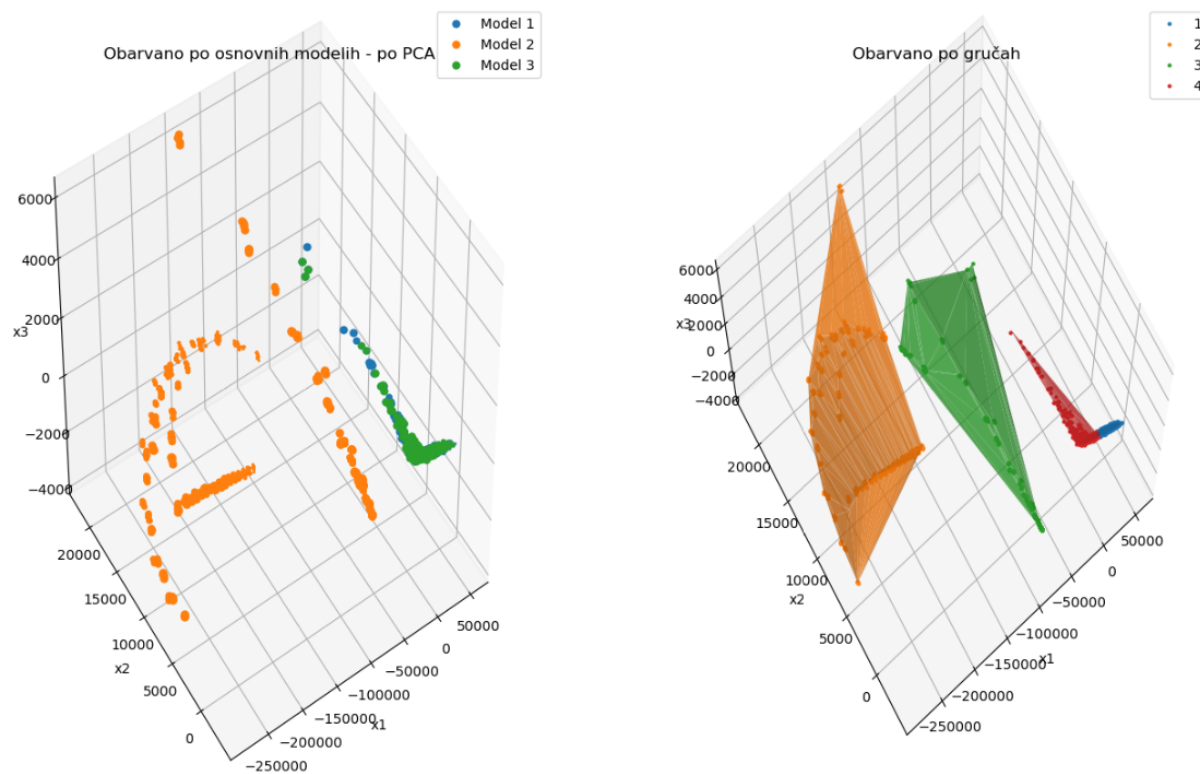


Fig. 4. Prve tri komponente transformiranega prostora.

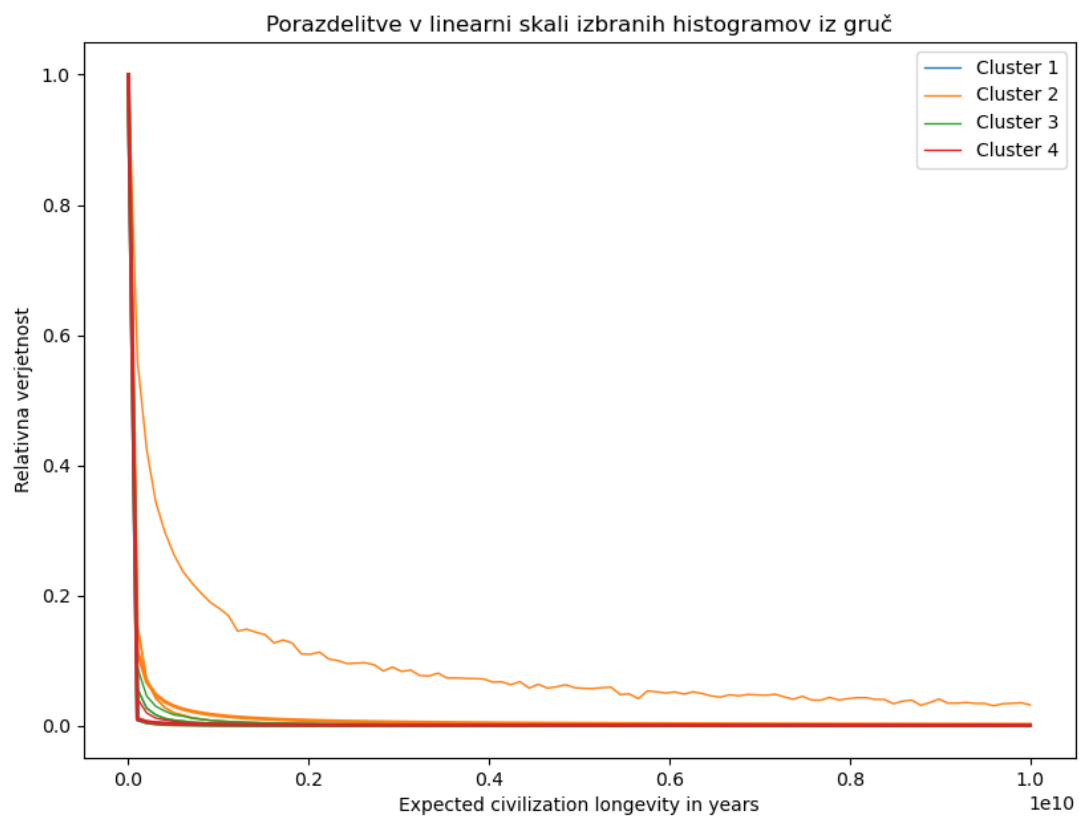


Fig. 5. Povpreĉja in robni primeri iz vsake gruĉe.

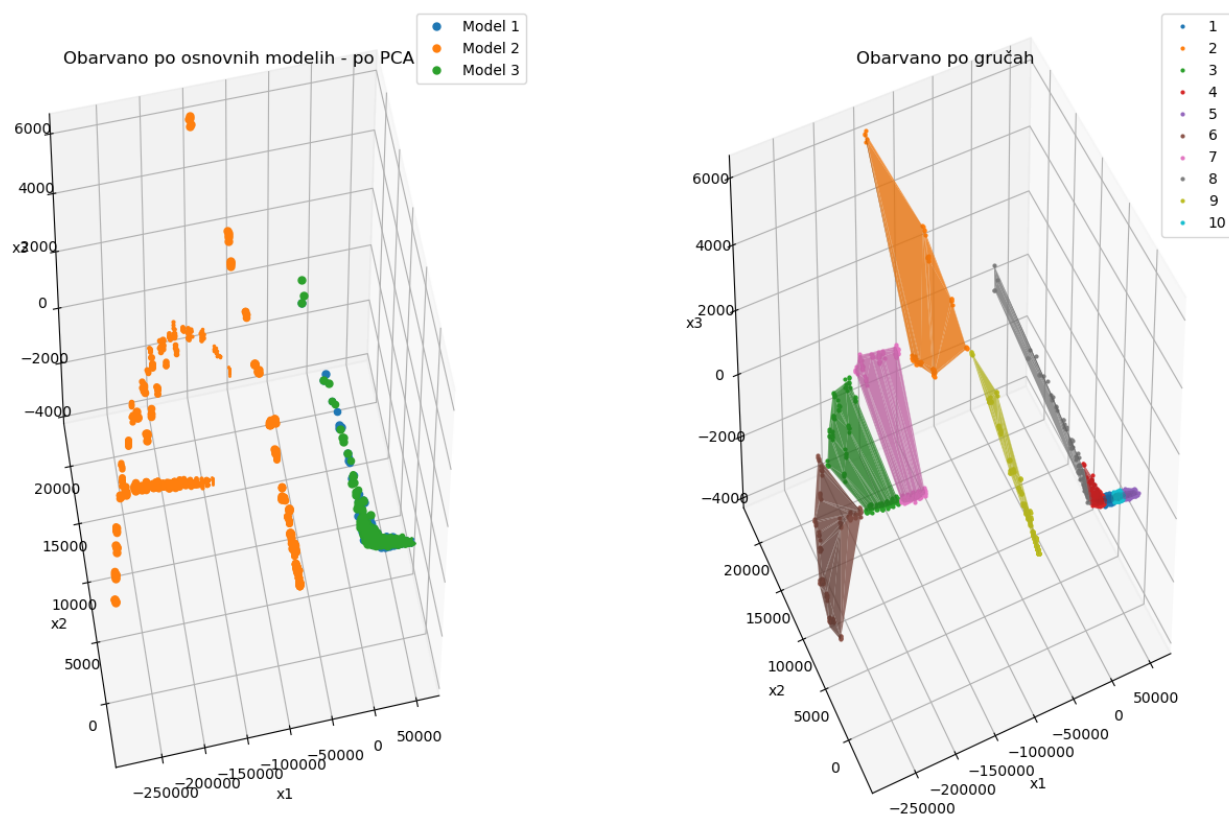


Fig. 6. Prve tri komponente transformiranega prostora.

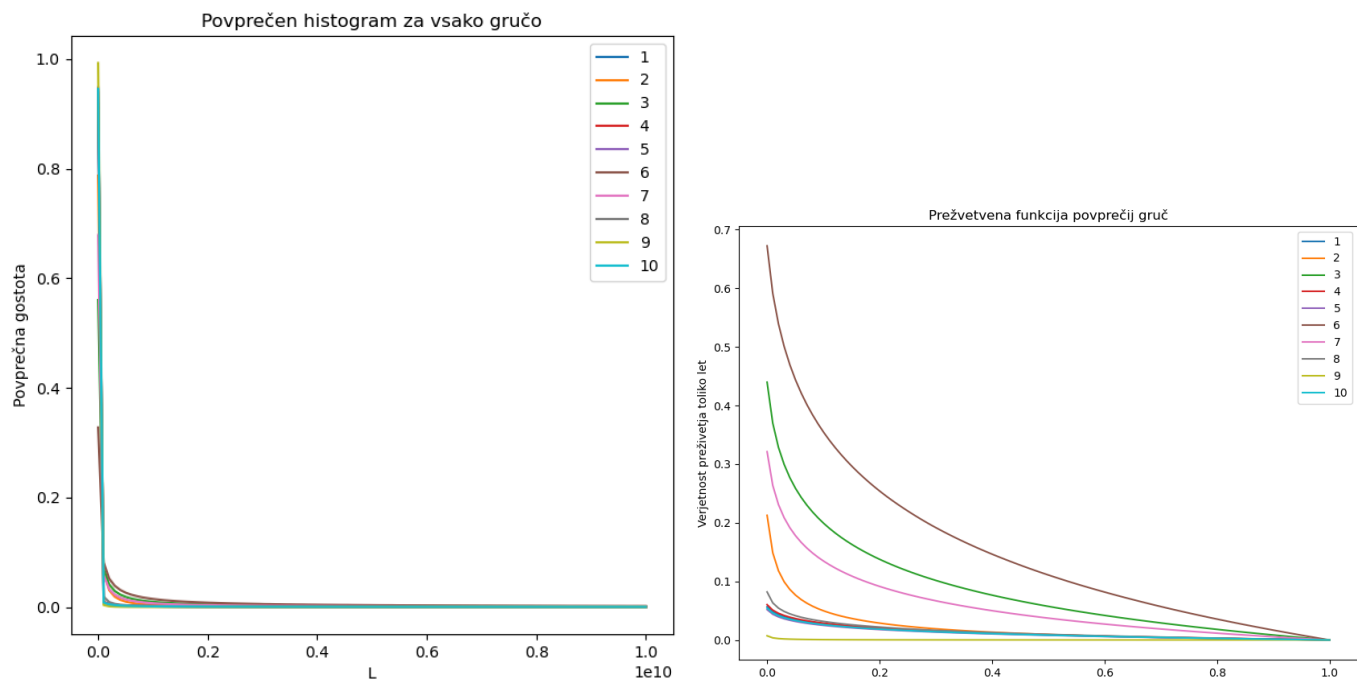


Fig. 7. Povprečni histogrami in njihove preživetvene funkcije.

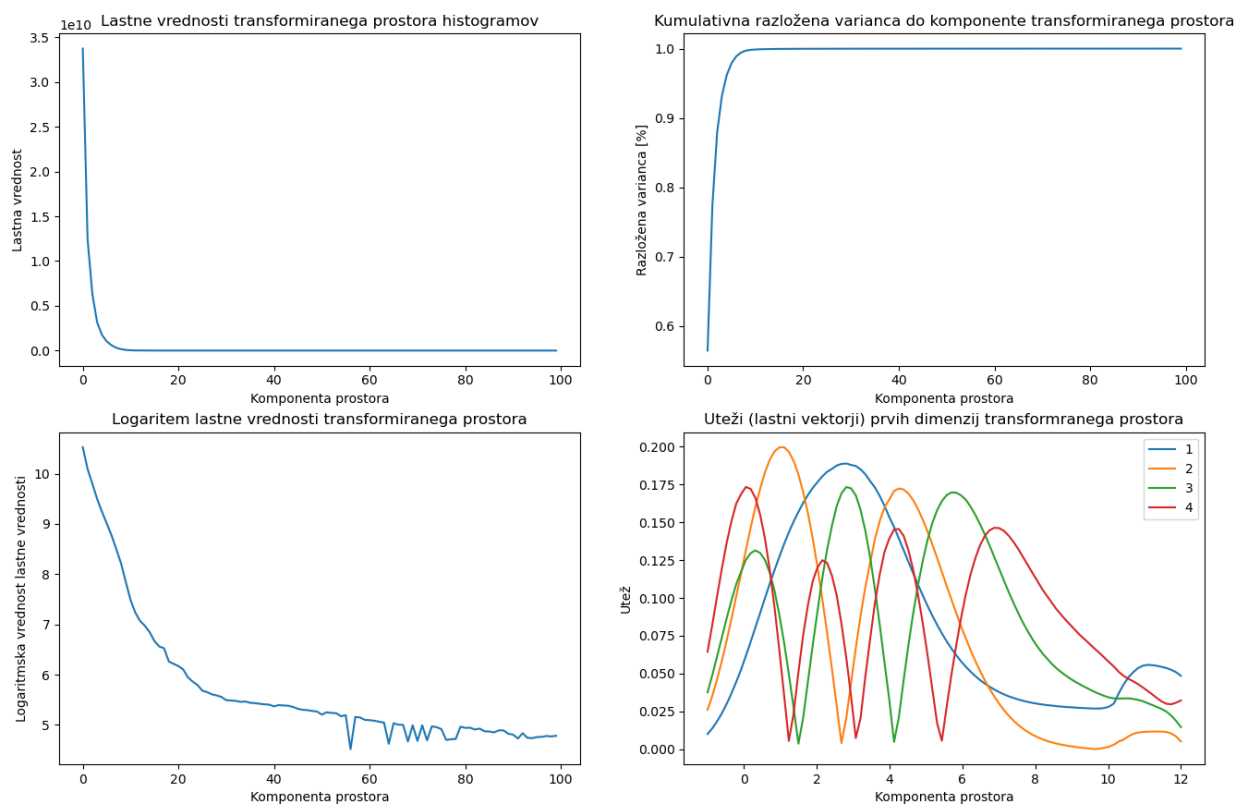


Fig. 8. Lastne vrednosti in lastni vektorji prostora logaritmskih histogramov.

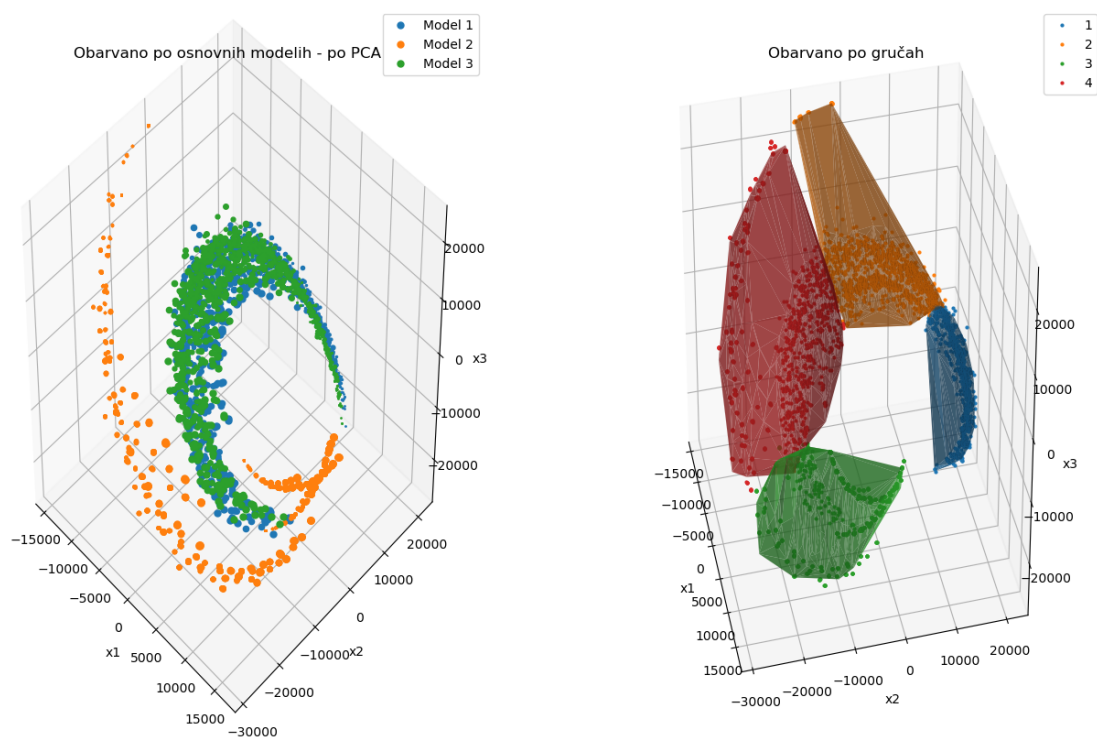


Fig. 9. Prve tri komponente transformiranega prostora.

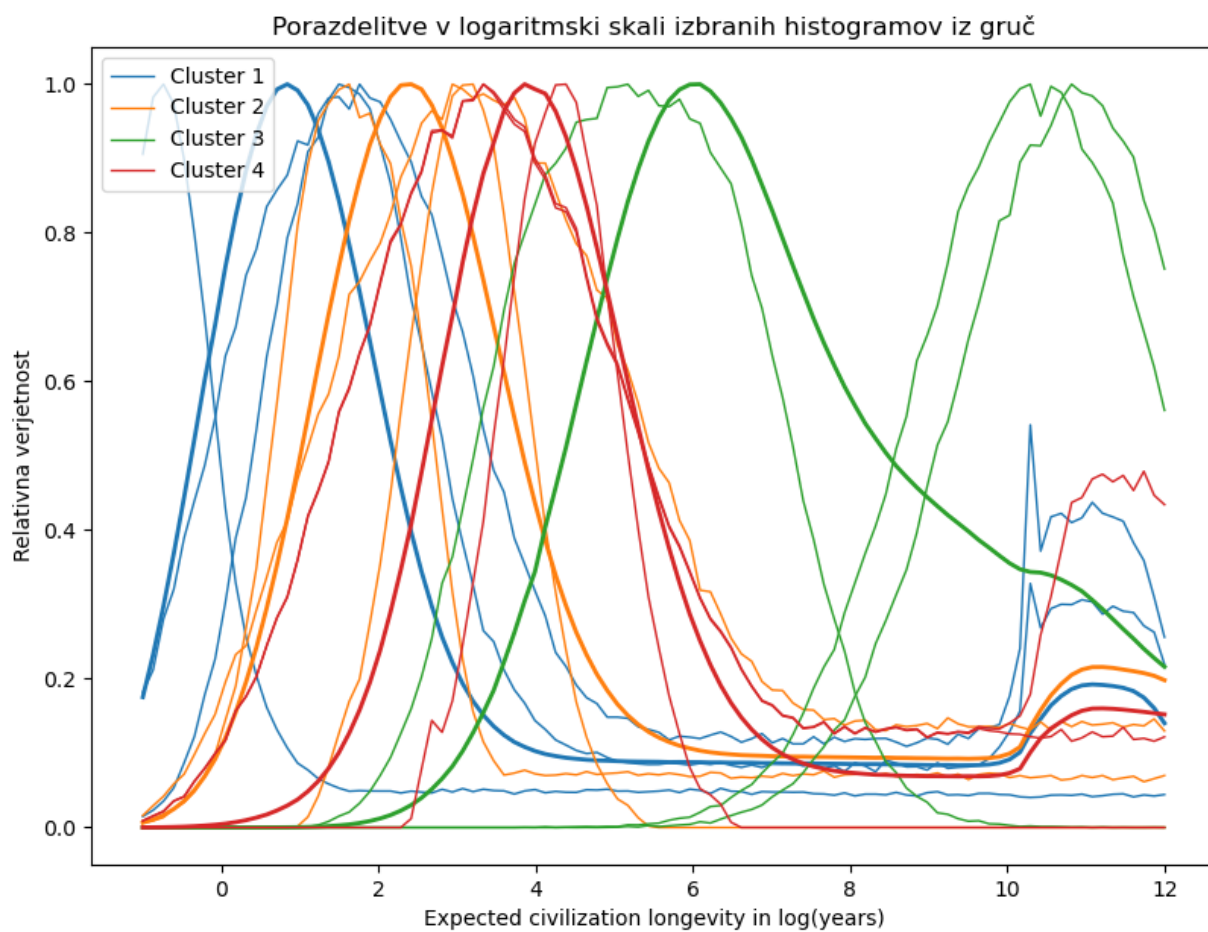


Fig. 10. Povpreĉja in robni primeri iz vsake gruĉe.

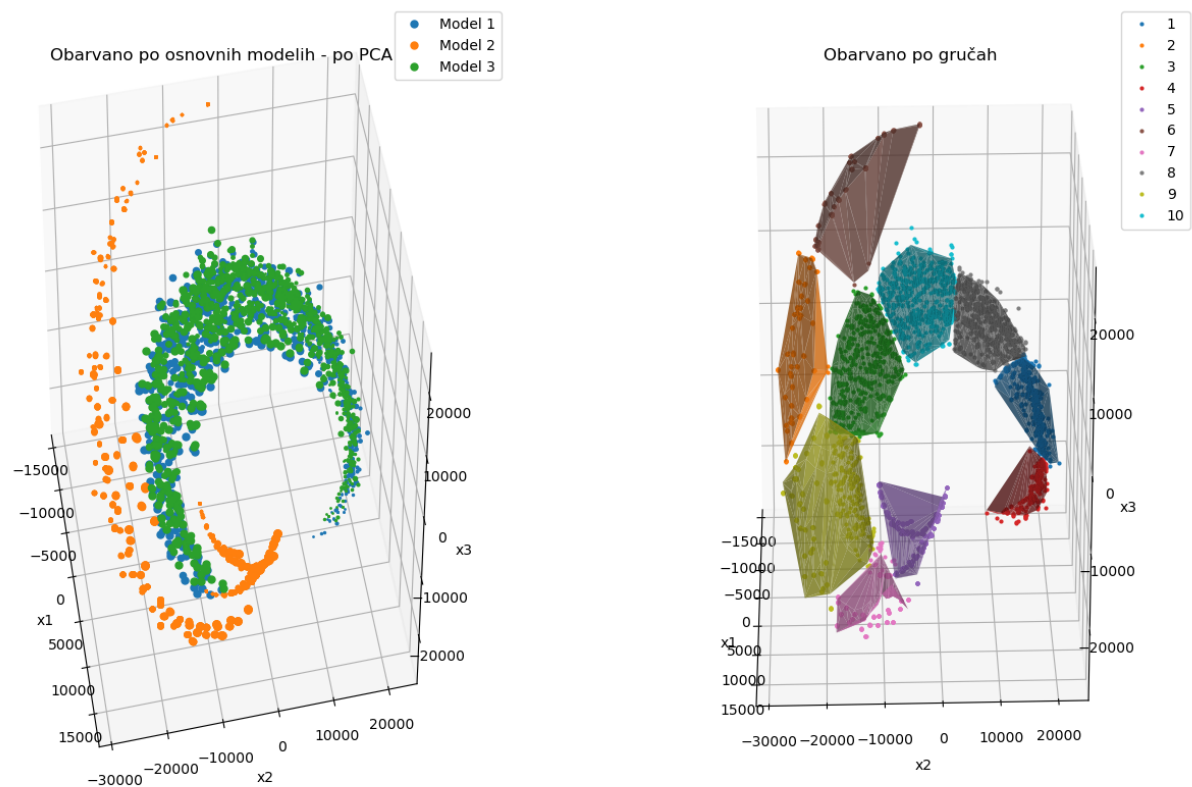


Fig. 11. Prve tri komponente transformiranega prostora.

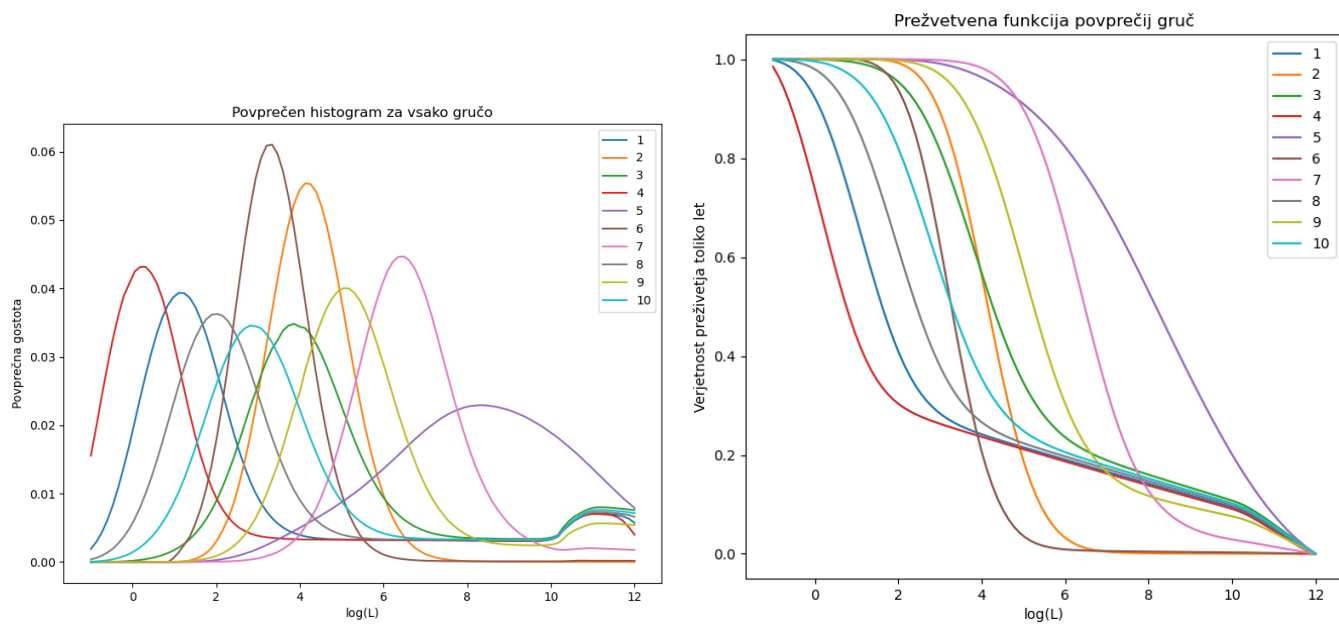


Fig. 12. Povpreĉni logaritmski histogrami in njihove preĉetvene funkcije.

v že gostih delih oblaka točk, se ga znebimo, ker želimo točke čim bolj enakomerno porazdeliti.

Ko bomo imeli vse te modele pripravljene, bomo izračunali histograme na podlagi vsaj 5 milijonov generiranih točk za bolj gladke histograme in na večji množici zgornjih mej za N .