

**METODE****ANALIZA POSELITVE GLEDE NA GEOMORFOLOŠKE ZNAČILNOSTI Z METODO MONTE CARLO**

AVTOR

**Tomaz Podobnikar**

Naziv: mag., univerzitetni diplomirani inženir geodezije, asistent z magisterijem

Naslov: Prostorskoinformacijska enota ZRC SAZU, Gosposka ulica 13, SI – 1000 Ljubljana, Slovenija

E-pošta: tomaz@zrc-sazu.si

Telefon: 01 470 64 93

Faks: 01 425 77 95

UDK: 528:681.3:551.4

COBISS: 1.02

## IZVLEČEK

**Analiza poselitve glede na geomorfološke značilnosti z metodo Monte Carlo**

Za prostorske analize v geografskih informacijskih sistemih se vedno bolj uporabljajo prilagojene neprostorske statistične metode in tehnike. Ena od njih je tudi metoda Monte Carlo, ki je v članku predstavljena na primeru testiranja statistične značilnosti vzorcev točk v prostoru glede na geomorfološke lastnosti površja. Na testnem območju Polhograjskega hribovja je testirana prostorska porazdelitev hiš in prebivalstva glede na oddaljenost od vodnih izvirov in vodotokov ter glede na nadmorsko višino, naklon, ekspozicijo, osončenost in zakraselost površja. Rezultati empiričnih testiranj z metodo Monte Carlo so primerjani s tradicionalnimi statističnimi testi.

## KLJUČNE BESEDE

metoda Monte Carlo, simulacija, analiza točkovnih vzorcev v prostoru, statistična značilnost, geografski informacijski sistem, Polhograjsko hribovje, Slovenija

## ABSTRACT

**Settlement analysis regarding to geomorphological significances with Monte Carlo method**

New approaches include more and more non-spatial statistical techniques adapted for solving the spatial analyses tasks in geographic information systems. An example of such techniques is the Monte Carlo method. In the article it is represented with the case of testing statistical significance of the spatial point patterns regarding to geomorphological characteristic. In the case study area of Polhograjsko hribovje (Polhov Gradec hills) spatial distribution of houses and population with regard to the distance from water springs and streams, height above sea level, relief slope, relief incline, relief aspect, relief insolation and karst is tested. Results of the empirical Monte Carlo tests are compared with traditional statistical tests.

## KEY WORDS

Monte Carlo method, simulation, spatial point pattern analysis, statistical significance, geographic information system, Polhograjsko hribovje, Slovenia

Uredništvo je prispevek prejelo 10. februarja 2000.

## 1. Uvod

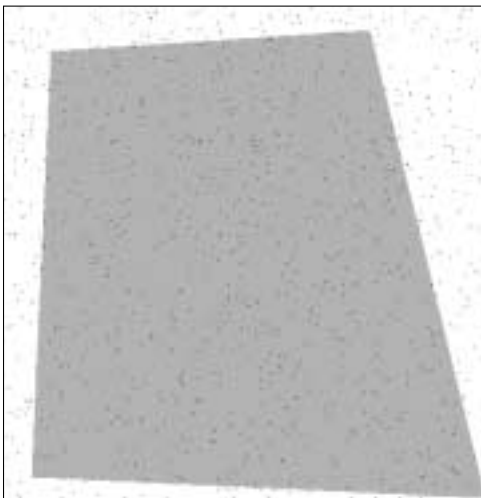
Pri regionalnih študijah pogosto vrednotimo različne prostorske porazdelitve z ugotavljanjem različnih statističnih povezav in raziskav vzorcev. Kadar se ne moremo zanašati na tradicionalne statistične teste, nam lahko pomagajo metode, ki temeljijo na naključnih številih in so znane pod imenom metode Monte Carlo. Uporabljamo jih predvsem takrat, ko ugotavljamo, ali obstaja določena prostorska porazdelitev točkovnih objektov tudi na neznačilnih vzorcih, pri katerih na prvi pogled ni opaziti korelacije. Metode Monte Carlo so še posebej primerne za testiranje statistične značilnosti prostorskih vzorcev, ki niso porazdeljeni normalno ali pa so majhni in neznačilni (Kvamme 1997).

V članku so metode Monte Carlo predstavljene na primeru regionalne prostorske analize sorazmerno majhnega in homogenega območja Polhograjskega hribovja. Preučili smo razmerja med naravo in družbo, ki jih znanost še razmeroma slabo pozna. Pri tem gre za integriranje podatkov fizične, družbene in predvsem regionalne geografije. Ugotavljali smo, ali je razmestitev hiš in prebivalstva povezana z geomorfološkimi značilnosti območja.

## 2. Metode Monte Carlo

Matematične, natančnejše statistične numerične metode Monte Carlo so se na najrazličnejših področjih znanosti močno razmahnile šele v zadnjem desetletju. Slučajnega pojava pri teh metodah ne opisujemo z analitičnimi zvezami, ampak s simulacijami (Vukadinović in Popović 1989). Naštejmo nekaj tipičnih področij uporabe omenjenih metod, ki se razvijajo na raznih področjih aplikativne znanosti: metode v statistični fiziki, naključna hoja, konstruiranje jedrskih reaktorjev, kvantna barvna dinamika, obsesvanje pri terapiji rakastih obolenj, pretok prometa, nastanek in razvoj zvezd, ekonometrija, napovedovanje borznega indeksa, raziskovanje naftnih vrelcev, ugotavljanje natančnosti geodetskih mrež, simulacija napak prostorskih podatkov (Podobnikar 1999) in podobno.

Pri tem uporabljamo najrazličnejše tehnike, na primer: integriranje Monte Carlo, reševanje sistemov linearnih enačb, ocenjevanje napak, tehnike redukcije variance, paraleliziranje in vektorizacija. Utežne (tehtane) metode Monte Carlo uporabljamo za zmanjšanje napak in za izboljšanje neodvisnosti funkcij (Mikhailov 1992), in sicer pri: optimizaciji postopkov ustvarjanja naključnih števil za oceno verjetnostnih karakteristik z naključnimi parametri, računalniških modelih naključnih polj in neme-



*Slika 1: Monte Carlo simulacija z metodo »zade- ni ali zgreši« za izračun površine nepravilnega štirikotnika (potemnjeno območje).*

ričnih simulacij, vektorskih postopkih za reševanje integralskih enačb, posebnih pristopih uporabe teorije perturbacij in podobno.

Enostaven primer tehnike integriranja z metodami Monte Carlo prikazuje izračun integrala s končnim številom dimenzij na primeru enostavne simulacije z metodo »zadeni ali zgreši«. Računali smo dvodimenzionalno površino nepravilnega štirikotnika. Z ustvarjanjem 1000 naključnih števil na območju s površino 1,000 enote smo dobili kot rezultat površino 0,708 enote, z analitično metodo pa površino 0,718 enote (slika 1).

## 2.1. Definiranje metod Monte Carlo

Metode Monte Carlo lahko zelo poenostavljeno predstavimo kot metode za računanje z naključnimi števili. Primernejšo definicijo metod Monte Carlo povzemamo po Kalosu in Whitlockovi (1986, 2): »... Metode Monte Carlo vsebujejo premišljeno uporabo naključnih števil pri izvedenju strukture stohastičnega procesa...«. S stohastičnim procesom mislimo na zaporedje položajev, katerih razvoj je določen z naključnimi dogodki. V računalniku jih ustvarjamo z naključnimi števili. Pri metodah Monte Carlo gre torej za reševanje problemov, ki niso povezani z verjetnostjo (na primer izračun vrednosti  $\pi$ ), z verjetnostnimi metodami. Glede na definicijo metod velja poudariti, da je treba razlikovati med naključnimi vrednostmi in metodami Monte Carlo. Z naključnimi števili le ustvarjamo sintetične podatke za testiranje, medtem ko pričakujemo od metod Monte Carlo tudi numeričen rezultat.

V nasprotju z metodami čiste oziroma teoretične matematike spadajo metode Monte Carlo med metode uporabne oziroma eksperimentalne matematike, ki se ukvarja s poskusi z naključnimi števili (Hammersley in Handscomb 1965). Glede na to, ali se pristopi do problemov neposredno nanašajo na vedenje ali rezultate naključnih procesov, jih delimo na dva tipa:

- verjetnostni pristop in
- deterministični pristop.

Enostaven verjetnostni problem je način opazovanja naključnih števil, ki omogoča neposredno simulacijo fizikalnega naključnega procesa originalnega problema. Glede na obnašanje naključnih števil v procesu lahko pridemo do zelene rešitve. Po tem postopku lahko raziskujemo na primer naraščanje populacije insektov glede na določeno privzeto vitalno statistiko preživetja in razmnoževanja. Pri takih analizah lahko napravimo model zgodovine življenja in razmnoževanja posameznih insektov. Za vsak posamezen insekt lahko ustvarimo naključna števila za starost in rojstvo potomca ter njegov pogin. Nato z naključnim procesom sledimo njegovim potomcem. Naključna števila pri tovrstni statistiki uporabimo za primerjavo določene količine naključnega vzorca glede na resnične rezultate, ki jih dobi entomolog v laboratoriju oziroma na terenu.

Deterministični pristop izkorišča moč teoretične matematike, na primer abstrakcijo in splošnost. Pri tem se izogiba šibkosti teoretične matematike, ki odpoveduje pri zamenjavi teorije s poskusi. Primer determinističnega pristopa je uporaba Laplaceove enačbe pri reševanju nekaterih problemov v elektromagnetni teoriji, kjer v določenih primerih odpove standardna analitična rešitev. Determinističnega pristopa se držimo tudi pri obravnavi našega primera uporabe metod v prostorskih analizah.

## 2.2. Zgodovinsko ozadje metod Monte Carlo

Ime za matematične metode Monte Carlo je nastalo okoli leta 1944, ko so jih začeli sistematično razvijati znanstveniki, ki so razvijali jedrsko orožje v projektu Manhattan v Los Alamosu (Kalos in Whitlock 1986; Computational Science Education Project 1995). Najprej se je metoda imenovala Monaco, in sicer po ruleti kot enostavnem generatorju naključij. Povod za izum metod je bila uporaba iger na srečo, ob katerih so začeli znanstveniki (in zasvojenici z igrami) študirati zanimive pojave in izide naključij. Večji razmah so metode Monte Carlo doživele po letu 1970 z razmahom digitalnih računalnikov (Pllana 1997), ki so se izkazali kot idealno sredstvo za njihovo izvajanje.

Poglejmo nekaj znanih del, ki so pripeljale do metod Monte Carlo. Prvo dokumentirano naključno vzorčenje najdemo v zvezi z rešitvijo integrala za izračun vrednosti  $\pi$  z naključnim spuščanjem žeblička na pravilno mrežo, ki jo je opisal francoski matematik Comte de Buffon leta 1768 v knjigi *Essai d'arithmétique morale* (Kalos in Whitlock 1986, 4; Pllana 1997). Leta 1886 je Laplace predlagal uporabo te ideje za izračun  $\pi$  in pri tem poudaril, da je konvergenca k rezultatu zelo počasna. Kasneje so se z naključnimi vzorčenji ukvarjali še Kelvin, Gosset, znan pod psevdonimom Student, in drugi. V tridesetih letih tega stoletja je fizik Fermi naredil nekaj numeričnih poizkusov z nevtronsko difuzijo, za katere bi že lahko rekli, da so prave simulacije Monte Carlo.

### 3. Teoretične osnove testov značilnosti pri metodah Monte Carlo

Pri testih značilnosti nas zanima odvisnost porazdelitve določene populacije v prostoru od značilnosti družbenega ali naravnega okolja. Populacija je lahko definirana z ekonomskimi ali obrednimi središči. Značilnosti družbenega okolja, kot so na primer vidnost, kognitivna pokrajina ploskev evklidske oddaljenosti, stroškovna ploskev, lahko modeliramo z orodji geografskega informacijskega sistema. Na podoben način lahko analiziramo prvine naravnega okolja, kot so na primer topografski podatki, značilnosti prsti, lastnosti kamnin in osončenost (Kvamme 1997).

Poznamo več tradicionalnih statističnih pristopov za testiranje značilnosti določene populacije, na primer testa  $\chi^2$  in  $t$ . V praksi testiramo statistično značilnost s primerjavo indeksa določenega prostorskega vzorca (na primer indeksa aritmetične sredine, indeksa najbližjega soseda ali koeficienta avtokorelacije) in skupine indeksov  $n$ -krat simuliranih naključnih porazdelitev pri uporabi populacije iste velikosti. V našem primeru smo se za obravnavanje statistične značilnosti omejili predvsem na indeks aritmetične sredine.

Privzemimo, da leži obravnavani prostorski vzorec populacije velikosti  $k$  na podobmočju  $S'$  (Kvamme 1997). Z ustvarjanjem  $k$  naključnih števil, kar je velikost populacije, ustvarimo  $n$  novih naključnih populacij:  $S_1, S_2, S_3, \dots, S_n$ . Nato za vsak vzorec, skupaj z originalnim, izračunamo določen indeks, ki ga imenujemo  $u$ , na primer srednjo vrednost. Tako statistiko torej izračunamo za  $n + 1$  vrednosti:  $u', u_1, u_2, u_3, \dots, u_n$ . Če predpostavimo, da je glede na obravnavano populacijo vsak vzorec enako verjeten, potem je statistična značilnost glede na originalni vzorec ter niz (rang) simuliranih vrednosti  $u$  naslednja:  $p = R(u') : (n + 1)$ . Če je na primer število naključnih populacij  $n = 999$  ( $n + 1 = 1000$  obravnavanih vrednosti, kjer 1 predstavlja vrednost za resnične podatke), je rang  $u'$ -ja (pri upoštevanju stopnje značilnosti testa  $\alpha$ ) pri obojestranski stopnji tveganja za obravnavani vzorec naslednji:  $R(u') \leq 50$  ( $\alpha \leq 5\%$ ) in  $R(u') \geq 950$  ( $\alpha \geq 95\%$ ). V tem primeru smo torej privzeli kot mero za značilnost obojestransko tveganje  $\leq 0,05$  ali  $\geq 0,95$ .

Centralni limitni izrek (poenostavljeno) pravi, da se porazdelitev pri dovolj velikem vzorcu približuje normalni porazdelitvi (Jamnik 1979, 509). Opazovana slučajna količina je namreč pogosto vsota neodvisnih in slučajnih prispevkov in s tem približno normalno porazdeljena. Razpršenost aritmetičnih sredin vzorcev, za katero mera je standardni odklon, se z večanjem vzorca manjša. Pri testiranju z metodami Monte Carlo, in tudi pri tradicionalnih testih značilnosti, izkoriščamo prednosti šibkega zakona velikih števil, ki pravi, da gre z naraščajočo velikostjo vzorca razlika odstopanja vzorčnega povprečja od povprečja populacije proti nič (Jamnik 1979, 502).

Na splošno dosežemo z metodami Monte Carlo precej boljše rezultate kot s tradicionalnimi statističnimi testi predvsem pri majhnih in neznačilnih vzorcih.

#### 3.1. Način vzorčenja za obravnavani primer

Rastrski pristop omogoča predstavitev populacije rastrskih celic obravnavanega območja kot celotno populacijo ( $K$ ). Hiše ( $k$ ), ki pripadajo le omejenemu številu rastrskih celic, obravnavamo kot vzorec

populacije  $K$ . Pri testiranju statistične značilnosti torej privzamemo celotno obravnavno populacijo hiš ( $k$ ) na danem območju za vzorec. Nato napravimo veliko število ( $n$ ) naključnih vzorcev enakomerno porazdeljene diskretne slučajne spremenljivke enake velikosti kot je število hiš ( $k$ ) na območju celotne populacije ( $K$ ). Za vsakega izmed  $n$  vzorcev izračunamo aritmetično sredino. Ker gre za veliko število vzorcev, se rezultati posameznih aritmetičnih sredin po centralnem limitnem izreku zbirajo okoli prave aritmetične sredine po normalni porazdelitvi (Petz 1985, 107), primerni za testiranje značilnosti. Normalnost porazdelitve aritmetičnih sredin, izračunanih iz  $n$  simulacij, lahko preverimo s testom Kolmogorov-Smirnov (Ambrožič in Leskošek 1998) ali kakim drugim testom.

#### 4. Metodologija testiranja značilnosti z metodo Monte Carlo

Za testiranje statistične značilnosti točkovnih vzorcev smo uporabili lasten AML-jev program, ki teče pod orodjem Arc/Info z uporabo modulov Arc in Grid. Program poganja podprogram, napisan v jeziku C, ki omogoča predvsem učinkovito generiranje naključnih števil. Uporabili smo generator psevdonaključnih števil, ki deluje na osnovi algoritma Park-Miller z Bays-Durhamovim mešanjem. Omenjeni algoritem prestane vse zahtevane statistične teste, razen testa ponovitve (periode), ki je sicer večja od  $10^8$  števil (Press, Teukolsky, Vetterling, Flannery 1995), vendar ni neskončna.

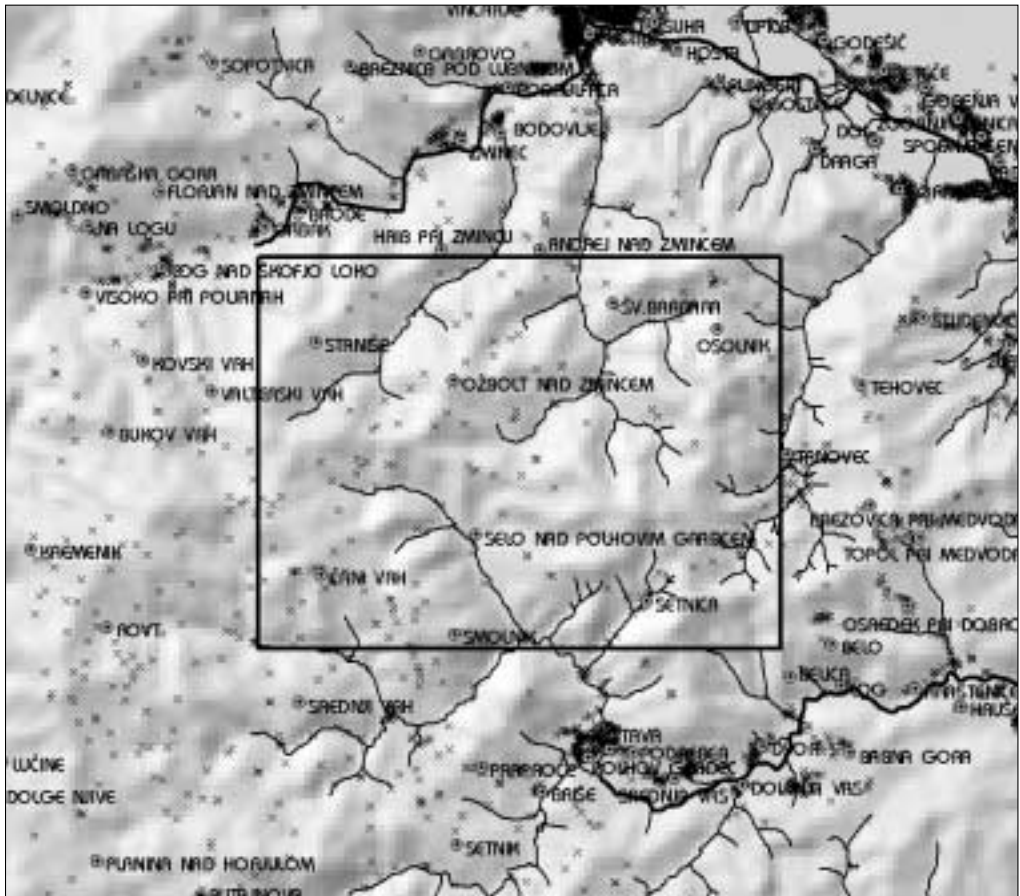
Testiranje značilnosti z metodo Monte Carlo smo izvajali za vektorsko podan točkovni sloj centroidov hišnih števil (RPE) na rastrsko podanih slojih nadmorskih višin, naklonov in osončenosti površja. Algoritem za testiranje je v osnovnih korakih naslednji:

- definiranje vhodnih parametrov: točkovne datoteke, ki predstavlja obravnavan vzorec, rastrske datoteke, na kateri preučujemo vzorec, in števila simulacij;
- ponavljanje spodaj navedenih korakov  $n$ -krat ( $n = 999$ ):
  - ustvarjanje naključnih števil,
  - linearno transformiranje ( $k =$  frekvenca obravnavanega vzorca) naključnih števil v dvorazsežni prostor obravnavanega območja,
  - prekrivanje  $k$  točkovnih vzorcev z atributi pripadajočih rastrskih celic obravnavanih slojev,
  - računanje aritmetične sredine  $m$  atributov rastrskih celic ob pogoju, da se prekrivajo s pripadajočimi točkami;
- izvajanje tretjega in četrtega koraka tudi za vhodne podatke točkovne datoteke glede na rastrske podatke;
- izdelava histograma vseh aritmetičnih sredin ( $n + 1$ );
- določitev meje ocene značilnosti z obojestranskim testiranjem ( $\alpha \leq 5\%$  in  $\alpha \geq 95\%$ ), ki se pri skupnem številu populacij  $n + 1 = 1000$  nanaša na manj kot 50 populacij z leve ali pa desne strani dobljene porazdelitve; v našem primeru je vrednost aritmetične sredine izvornih podatkov padla v območje značilnosti glede na simulirane podatke.

#### 5. Praktični primer testiranja značilnosti za območje Polhograjskega hribovja

Cilj statističnega testiranja z metodo Monte Carlo je ugotavljanje statistične povezave med razporeditvijo hiš in prebivalstva z značilnostmi pokrajine. Simulacijo izvajamo z naključnim vzorcem hiš v prostoru. Hiše so podane kot točke, sloji, ki predstavljajo značilnosti pokrajine, pa rastrsko. Rezultati testiranja so vizualno (opazovanje vzorcev hiš in histogramov) in statistično ovrednoteni.

Testiranje značilnosti prostorskih vzorcev z metodo Monte Carlo smo preizkusili na majhnem prostem pravokotnem območju Polhograjskega hribovja zahodno od Ljubljane in južno od Škofje Loke, podanim z naslednjimi Gauss-Krügerjevimi koordinatami v metrih: jugozahodno oglišče 442.100 in 104.000, severovzhodno oglišče 450.100 in 110.000), kar znaša 8 krat 6 km. Izbrano testno območje



Slika 2: Širše območje Polhograjskega hribovja s testnim območjem (8 krat 6 km). S križci so označene lege hiš, s krogi pa naselja.

obsega osrednji, zahodni in severni del Polhograjskega hribovja (Gabrovec 1989). Za območje so značilni zaobljeni vrhovi, široka slemena in ozke doline. Območje je zgrajeno predvsem iz dolomita in drugih karbonatnih kamnin, v katerih večinoma niso razviti površinski kraški pojavi, manj pa je nekarbonatnih kamnin.

Vsi sloji geografskega informacijskega sistema, na katerih ugotavljamo prostorsko razporeditev hiš, so podani z 1 ha velikimi rastrskimi celicami. Skupno število celic  $K$  je 4800 celic (48 km<sup>2</sup>), kolikor je teoretično možnih leg hiš. Iz RPE-ja Geodetske uprave Republike Slovenije iz leta 1993 ugotovimo, da je na omenjenem območju registriranih 171 hiš, kar pomeni 3,56 hiše na km<sup>2</sup>. Celotno populacijo hiš štejemo pri testiranju kot predstavnike vseh možnih leg. V hišah je po popisu iz leta 1991 živelo 542 prebivalcev (od 0 do 14 registriranih prebivalcev na posamezno hišo), kar pomeni 3,17 prebivalca na hišo in 12,9 prebivalca na km<sup>2</sup>, kar je več kot sedemkrat manj od slovenskega povprečja. Območje je brez večjih naselij, vizualno pa tudi ocenimo, da je območje zelo enakomerno poseljeno (slika 2). Testna porazdelitev obravnavanega vzorca hiš in geomorfoloških značilnosti je torej na pogled izrazito neznačilna. Pričakovati je torej, da v večini primerov ne moremo dobiti dobrih rezultatov statističnega testiranja, torej značilnega odstopanja od naključnih vzorcev.

### 5.1. Opis testiranih podatkov

Kot osnovo smo za testiranje statističnih značilnosti z metodo Monte Carlo uporabili vektorsko podan točkovni sloj centroidov hišnih števil. Testirali smo ga glede na rastrsko podane sloje z ločljivostjo 100 m. Uporabili smo pet podatkovnih slojev. To so:

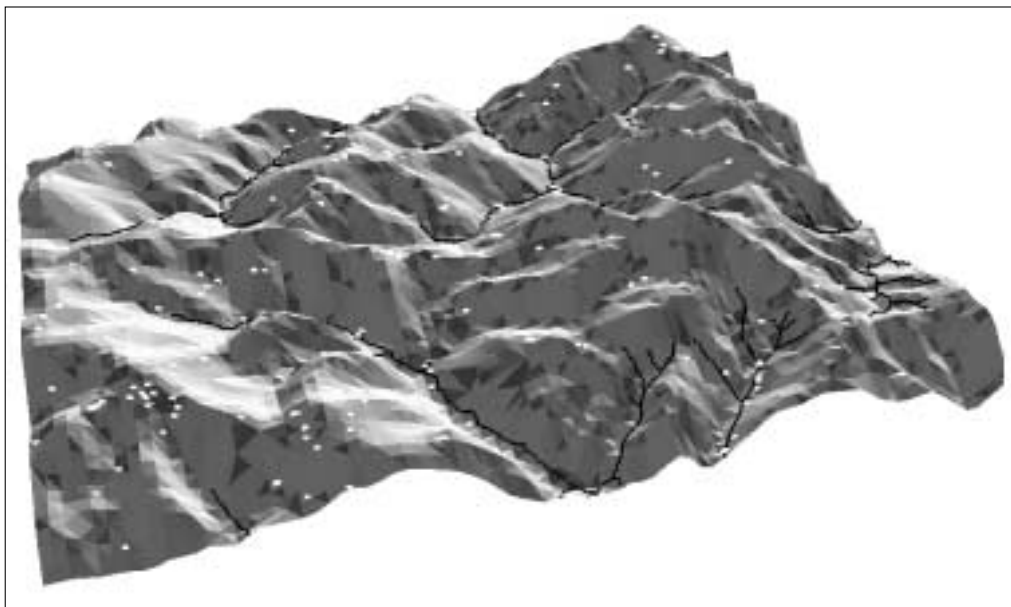
- centriodi hišnih števil (RPE), ki so bili zajeti s TTN-5, TTN-10, ortofoto načrta ali odmere objekta, objekti pa so opredeljeni kot točke (Geodetska uprava Republike Slovenije),
- digitalni model višin (DMR 100), ki je bil izdelan iz temeljnih topografskih načrtov v merilu 1 : 5000 in 1 : 10.000 (Geodetska uprava Republike Slovenije), objekti pa so opredeljeni kot rastrske celice (slika 3),
- osončenost Slovenije (Gabrovec 1996), pri kateri so objekti prav tako podani kot rastrske celice (Geografski inštitut Antona Melika ZRC SAZU),
- generalizirana kartografska baza za vodotoke in izvire v merilu 1 : 25.000 (GKB 25-H), pri kateri so izviri opredeljeni kot točke in vodotoki kot linije (Geodetska uprava Republike Slovenije, in
- metalogenetska karta SRS v merilu 1 : 220.000, ki je bila digitalizirana in rastrirana z ločljivostjo 100 krat 100 m (Geološki zavod Ljubljana in Raziskovalna skupnost Slovenije 1980).

Za testiranje smo iz digitalnega modela višin izdelali še tri rastrske sloje z ločljivostjo 100 m:

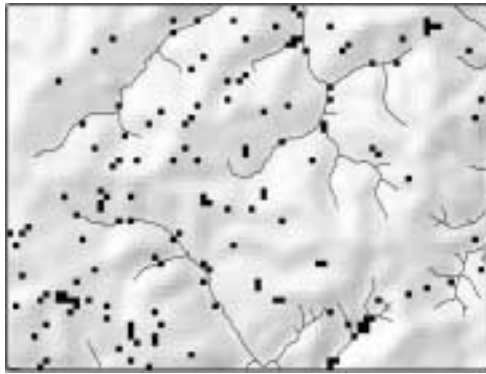
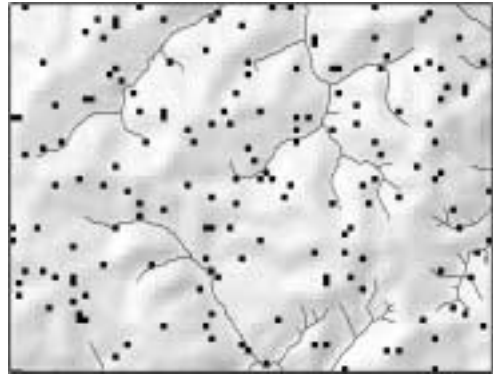
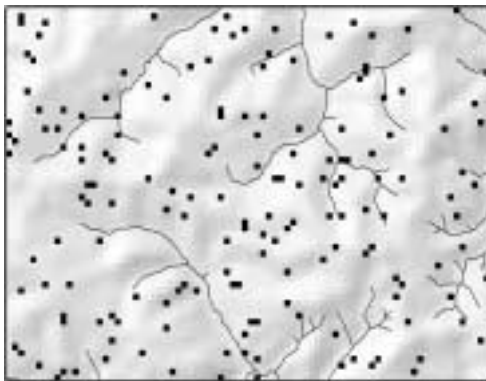
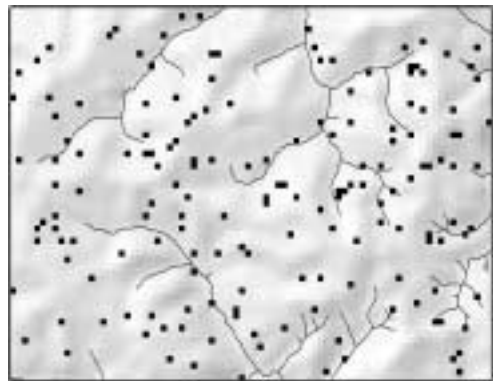
- sloj naklonov površja v kotnih stopinjah od 0° do 90°,
- sloj ekspozicij površja v kotnih stopinjah med 0° in 360°,
- sloj evklidske oddaljenosti od vodnih izvirov in vodotokov v metrih.

### 5.2. Izračun osnovnih statističnih parametrov razporeditve hiš

Odločimo se za enostavno testiranje dejanske porazdelitve hiš ( $k = 171$ ), (slika 4a) glede na dvo-raszežnostno enakomerno porazdelitev slučajne spremenljivke (slika 4b, 4c in 4č). S simulacijo



Slika 3: Perspektivni prikaz testnega območja Polhograjskega hribovja z označenimi legami hiš in rečno mrežo. Faktor povečave višin glede na položajne mere je 2.

a) lege hiš ( $k = 171$ )b) simulacija 1 ( $k = 171$ )c) simulacija 2 ( $k = 171$ )č) simulacija 3 ( $k = 171$ )

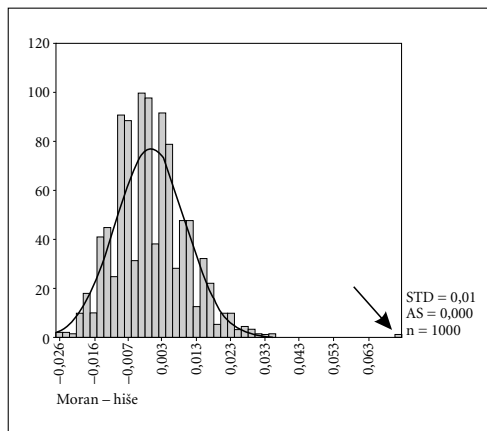
Slika 4: Primerjava prostorske razporeditve 171 hiš (a) glede na tri simulirane primere z enakim številom hiš na testnem območju.

ustvarjamo naključne vzorce enakomerno po vsem obravnavanem območju, neodvisno od geografskega položaja. Omenjeno (poenostavljeno) metodo testiranja seveda ni smiselno prenesti na tista območja Slovenije, kjer lahko v večini primerov že na prvi pogled opazimo, da porazdelitve hiš nikakor niso naključne. Na sliki vidimo, da na prvi pogled težko opazimo razliko med dejansko porazdelitvijo in simuliranimi porazdelitvami.

Rezultati testiranja so naslednji: geometrično središče leg hiš je glede na geometrično središče območja premaknjeno za 490 m proti zahodu in 285 m proti jugu, geometrično središče porazdelitve prebivalstva pa za 816 m proti zahodu in 358 m proti jugu. To lahko pomeni, da je več stalnih prebivalcev dlje od Ljubljane in Škofje Loke, lastniki počitniških hiš pa so bližje obeh mest.

Zanimivejše je testiranje statistične značilnosti pri ničelni domnevi, da so hiše porazdeljene naključno po obravnavanem območju. Za testiranje lahko primerjamo avtokorelacijski Moranov koeficient  $I$  med izvornim in simuliranim slojem leg hiš.  $ZI = 0,0705$  smo dobili majhno stopnjo pozitivne avtokorelacije med legami. To pomeni, da imajo hiše določeno tendenco zbiranja v skupine (zaselke in vasi). Histogram kaže, da se porazdelitev hiš bistveno razlikuje od naključne, saj jo najdemo kot najmanjši podatek (aritmetičnih sredin) z desne. S tem lahko z manj kot 0,1 % tveganjem sprejmemo alternativno domnevo, da se hiše zbirajo v gručeh (slika 5).





Slika 5: Histogram porazdelitve Moranovega koeficienta za naključno porazdeljene točke na testnem območju glede na vrednost tega koeficienta za izvirne podatke hiš (potemljen stolpec na skrajni desni).

### 5.3. Testiranje razporeditve hiš glede na geomorfološke značilnosti

Za testiranje smo uporabili:

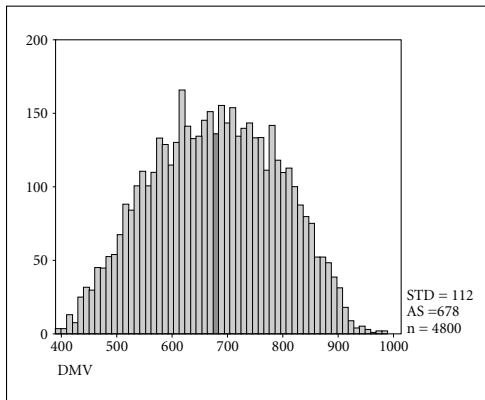
- nadmorsko višino površja,
- oddaljenost od vodnih izvirov in potokov,
- naklon površja,
- ekspozicijo površja,
- osončenost površja in
- zakraselost površja.

Zanimajo nas rezultati statističnih testov, ki odgovorijo na vprašanje, ali obstaja povezava med razporeditvijo hiš in prebivalstva ter geomorfološki značilnosti površja. Vsaki hiši, ki pripada posamezni rastrski celici, na katerih izvajamo simulacije, pripišemo odgovarjajočo vrednost celice. Nato izračunamo srednjo vrednost takega sloja tako za šest izvornih in vse simulirane sloje. Pripomniti velja, da kljub temu da smo pred tem ugotovili značilno zbiranje dejanskega vzorca hiš v gruče, pri testiranju značilnosti v članku ne simuliramo do dejanske stopnje avtokoreliranih vzorcev, ampak ostajamo pri naključnih vzorcih s stopnjo avtokorelacijskega koeficienta  $I = 0$ , kar poenostavi izvedbo naloge.

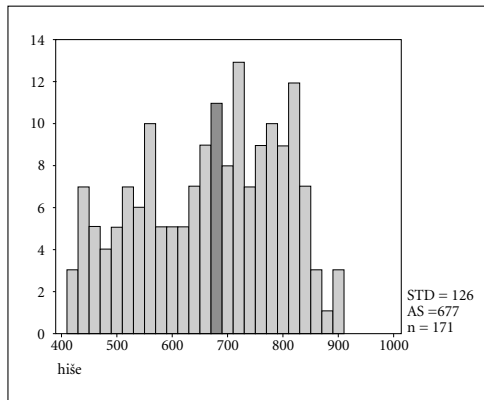
Končno ugotavljamo, ali se aritmetična sredina podatkov za izvorni sloj statistično značilno razlikuje od simuliranih. S tem potrdimo ali pa zavrnemo ničelno domnevo, da je se aritmetična sredina testirane izvirne porazdelitve, v našem primeru populacije 171 hiš, značilno ne razlikuje od aritmetične sredine celotnega območja 48.000 rastrskih celic oziroma aritmetične sredine, ki jo dobimo iz  $n = 999$  vzorcev simulacije, pri obojestranski stopnji tveganja  $\alpha \geq 5\%$  oziroma  $\alpha \leq 95\%$ . Predvidevamo torej, da so simulirani podatki porazdeljeni normalno in naključno. Pri testiranju se moramo zavedati, da je varianca povprečne lege hiš z večanjem vzorca  $k$  vedno manjša.

#### 5.3.1. Testiranje razporeditve hiš in prebivalstva glede na nadmorsko višino

Podrobneje obravnavamo prvega izmed šestih primerov testiranja porazdelitve hiš glede na geomorfološke značilnosti obravnavanega območja. Nadmorska višina obravnavanega območja se giblje med 403 in 984 m, s srednjo vrednostjo 678 m, kar je 25 m pod vrednostjo aritmetične sredine obeh skrajnih vrednosti intervala. Sklepamo, da nekoliko prevladajo nižji predeli (slika 6a). Histogram nadmorskih višin območja kaže na manjše odstopanje od normalne porazdelitve. Hiše so postavljene na nadmorskih višinah med 419 in 904 m (slika 6b), z aritmetično sredino nadmorske višine 677 m, medtem ko je povprečna vrednost nadmorske višine bivanja prebivalcev 689 m. Opazimo pa, da ima



a) DMV: aritmetična sredina je 678 m

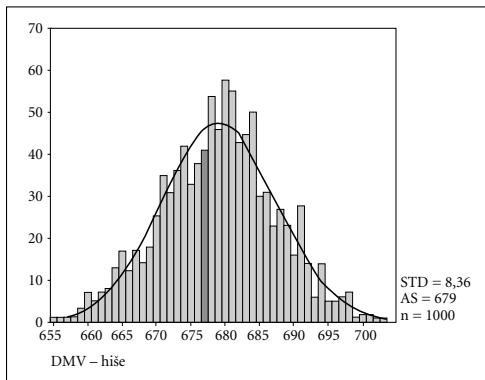


b) hiše: aritmetična sredina je 677 m

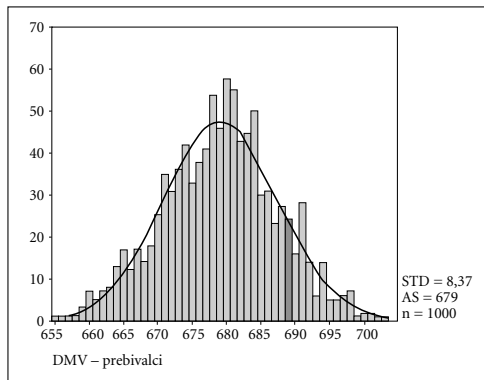
Slika 6: Histogram porazdelitve vseh nadmorskih višin z označeno povprečno nadmorsko višino (a) ter histogram porazdelitve nadmorskih višin hiš z označeno povprečno nadmorsko višino (b).

razporeditev hiš dva neizrazita maksimuma, in sicer na nadmorskih višinah med 500 in 550 m, kjer prevladujejo doline, ter med 650 in 850 m, kjer so zgornja pobočja in slemena hribov.

Z metodo Monte Carlo testiramo značilnosti z ničelno domnevo, da je povprečna hiša postavljena na značilno na povprečni nadmorski višini testnega območja, in prav tako, da povprečni prebivalec živi na povprečni nadmorski višini. Za hiše se izkaže, da postavljeno domnevo lahko potrdimo. Referenčna srednja vrednost je šele na 439. mestu z leve (slika 7a). Nekoliko drugačne rezultate dobimo, če analiziramo prebivalstvo glede na nadmorsko višino. Izkaže pa se, da nekoliko večji del prebivalstva (a ne statistično značilen) živi v hribih, saj se normirana referenčna vrednost pojavi na 118. mestu z desne (slika 7b). Kljub neznačilnim odstopanjem rezultatov od ničelne domneve pa lahko sklepamo, da je obstaja tendenca, da velike družine živijo v hribih, v dolinah pa najdemo opuščene hiše ali pa počitniške hiše.



a) simulacija: hiše je 677 m



b) simulacija: prebivalci je 689 m

Slika 7: Na histogramih povprečnih vrednosti nadmorskih višin na simuliranih vzorcih je označena povprečna vrednost nadmorskih višin izvornih podatkov hiš (a) in povprečna nadmorska višina bivanja prebivalcev (b).

### 5.3.2. Geografska interpretacija rezultatov testiranja

Glede na izvedene simulacije in druge analize, ki smo jih izvedli za hiše in prebivalstvo na vseh šestih podatkovnih slojev, lahko sklepamo, da so od obravnavanih naslednje spremenljivke za poselitev najpomembnejše: naklon površja, osončenost površja, ekspozicija površja, manj pa zakraselost in oddaljenost od potokov in izvirov, medtem ko je povprečna nadmorska višina hiš nepomembna. Značilna oziroma tipična hiša stoji torej na rahlo nagnjenem svetu, obrnjenem proti jugu, predvsem na nekarbonatni kamnini ter prej na hribu kot v dolini. Mimogrede, analize pokažejo tudi to, da na hribih ni potokov in je malo izvirov.

### 5.4. Primerjava metode Monte Carlo s tradicionalnimi metodami

Rezultate empiričnega testiranja značilnosti z metodo Monte Carlo smo ocenili prek primerjave s tradicionalnimi testi, ki smo jih izvedli za neparametrične (kategorične) vrednosti, kot je na primer kamninska zgradba površja, in za parametrične (zvezne) vrednosti, kot je na primer osončenost. Pri tem smo želeli potrditi domnevo, da so empirični testi, izvedeni s simulacijami metode Monte Carlo, boljši od testov pri tradicionalnih metodah. Med tradicionalnimi testi smo izbirali med testom  $\chi^2$ , ki je primeren za neparametrične statistične teste domnev, ter test  $t$ , ki je primeren za parametrične teste razlik med dvema populacijama. Glede na relativno velik vzorec populacije hiš ( $k = 171$ ), ki ga lahko prištevamo že kar med velike vzorce, smo uporabili tudi test  $z$ , ki ga namesto testa  $t$  uporabljamo za večje vzorce pri predpostavki, da se populacija porazdeljuje normalno, kar pa za obravnavane porazdelitve večinoma ne drži.

Pokazalo se je da pri konkretnem primeru uporabe testov značilnosti  $\chi^2$  in  $z$  za neparametrične vrednosti, kažeta na precej manjšo možnost, da postavljeno hipotezo zavrne kot pa z metodo Monte Carlo. Z drugimi besedami to pomeni, da se v tem primeru testiranje z metodo Monte Carlo precej boljše odziva na dano porazdelitev vzorca. Nekoliko obrnjen rezultat smo dobili pri primerjavi testov značilnosti  $\chi^2$ ,  $z$  in  $t$  za parametrične vrednosti. V tem primeru smo z s tradicionalnimi testi nekoliko prej zavrgli postavljeno hipotezo, vendar ne bistveno. Še najprej smo jo zavrgli s testom  $t$ . Največja napaka se pojavi na mejah intervala zaupanja (v naših primerih  $\alpha = 5\%$ ), ki so najzanimivejše za testiranje.

Metodo Monte Carlo, ki smo jo uporabili v prakazanih primerih, je pri vsakdanjem delu namesto tradicionalnih testov smiselno uporabiti šele takrat, ko ti kažejo na dvoumno rešitev problema. V obravnavanem testnem primeru se je na večini morfoloških značilnosti površja izkazala kot zelo koristna.

## 6. Sklep

Slaba stran tradicionalnih statističnih testov je, da ne glede na resnično porazdelitev najpogosteje privzemajo, da je porazdelitev normalna. Vzorcna porazdelitev aritmetične sredine  $m$ , ki smo jo uporabljali pri testiranju z metodo Monte Carlo, je vedno približno normalno porazdeljena, če je le velikost populacije  $n$  dovolj velika, na primer nad 30 ali 50 (Petz 1985, 110). Prednost rastrskih geografskih informacijskih sistemov pri naših testiranjih je, da vsebuje obravnavano območje končno število populacije ( $K$  rastrskih celic), kar omogoča lažji izračun populacijskih parametrov celotnega območja.

Poglavitne značilnosti in prednosti testov značilnosti z metodo Monte Carlo navajamo glede na oceno rezultatov našega testiranja, deloma pa jih povzemamo po Kvammeju (1997):

- V osnovi lahko testiranje z ustvarjanjem naključnih števil uporabljamo tudi z naključnim vzorcem populacije iz obravnavanega območja. Pri tem za testiranje ni treba privzeti normalne porazdelitve vhodnih podatkov kot pri tradicionalnih testih značilnosti.

- V testiranje lahko vključimo tudi zapleteno spremenljivko, kot je na primer vidnost. Pri tem lahko populacija vsebuje vse možne vidnosti (vrstica krat stolpec), ki povzročijo določitev populacijskih parametrov računsko težke tudi za srednje velika območja. Pristop naključnega spreminjanja  $n + 1$  vidnosti predstavlja bolj obvladljiv problem. Na podoben način bi lahko simulirali tudi naključni digitalni model višin in potem ocenjevali določene spremenljivke na njem.
- Metodo Monte Carlo za testiranje značilnosti, ki je omejena le na simulacijo naključnih točk pri poznavanju problema, bi lahko razširili tudi na simulacijo linij in območij ter celo ploskev.
- Bistvena potencialna prednost simulacijskih postopkov Monte Carlo pri testiranju značilnosti je, da lahko ob dobrem poznavanju geografskih značilnosti obravnavanega območja oziroma vzorca uporabimo druge modele vzorčenja. Kot primer navajamo, da včasih ne moremo privzeti, da je obravnavana populacija v prostoru razporejena neodvisno. Mnogo večja možnost je, da obstaja med elementi populacije, največkrat točkami, neka soodvisnost. Z drugimi besedami; lega posamezne točke je delno odvisna od lege drugih točk. Če znamo podrobno opisati pravila odvisnosti, lahko dani vzorec primerjamo s simuliranimi vzorci, ki jih dobimo po primerljivih pravilih vzorčenja. Eden izmed možnih načinov vzorčenja je upoštevanje koeficienta avtokorelacije vzorca. Lahko pa iz postopka testa značilnosti vzorca enostavno izločimo tista območja, ki ne zadoščajo predpisanim pogojem. Primer je testiranje porazdelitve hiš glede na naklon površja, pri čemer izločimo naklone, večje od  $30^\circ$ . Nadalje lahko pri simulaciji upoštevamo tudi to, da je na različnih naklonih površja različna verjetnost pojavljanja hiš. Eno od smiselnih nadaljevanj opisanega primera bi bilo modeliranje, na primer območij, najprimernejših za poselitve. Metode ugotavljanja značilnosti vzorca se v postopku modeliranja nanašajo na opisni model. Pri tem bi morali ugotoviti še, če je posamezno spremenljivko smiselno uporabiti za model, na primer z izračunom večkratne regresije.

## 7. Viri in literatura

- Ambrožič, F., Leskošek, B. 1998: Uvod v SPSS (verzija 7.5 za Windows 95/NT). Ljubljana.
- Computational Science Education Project, 1995: Introduction to Monte Carlo Methods. Medmrežje: <http://csep1.phy.ornl.gov/CSEP/MC/MC.html> (8. 2. 2000).
- Gabrovec, M. 1989: Vpliv reliefa za geografsko podobo Polhograjskega hribovja. Magistrska naloga. Oddelek za geografijo Filozofske fakultete Univerze v Ljubljani. Ljubljana.
- Gabrovec, M. 1996: Solar Radiation and the Diverse Relief in Slovenia. Geografski zbornik 36. Ljubljana.
- Hammersley J. M., Handscomb, D. C. 1965: Monte Carlo methods. New York.
- Jamnik, R. 1979: Verjetnostni račun in statistika, Višja matematika II. Ljubljana.
- Kalos, M. H., Whitlock, P. A. 1986: Monte Carlo methods, Volume I: Basics. New York.
- Kvamme, K. L. 1997: GIS and Statistical Inference in Arizona: Monte Carlo Significance Tests. Archaeological Applications of GIS, Proceedings of Colloquium II, UISPP 13<sup>th</sup> Congress, Forli, Italy, September 1996. Sydney. (zgoščenka)
- Mikhailov, G. A. 1992: Optimization of Weighted Monte Carlo Methods. Berlin.
- Petz, B. 1985: Osnovne statističke metode za nematematičare. Zagreb.
- Pllana S. 1997: History of Monte Carlo method. Medmrežje: <http://stud2.tuwien.ac.at/~e9527412> (8. 2. 2000).
- Podobnikar, T. 1999: Monte Carlo Simulations in Slovenia. Modelling and Visualisation of Spatial Data Error. GIM International, 13–7. Lemmer.
- Press, H. W., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P. 1995: Numerical Recipes in C, The Art of Scientific Computing, Cambridge University Press. Medmrežje: [http://www.ulib.org/web-Root/Books/Numerical\\_Recipes/](http://www.ulib.org/web-Root/Books/Numerical_Recipes/) (9. 2. 2000).
- Vukadinović, S., Popović, J. 1989: Metoda Monte-Karlo. Beograd.

## 8. Summary: Settlement analysis regarding to geomorphological significances with Monte Carlo method

(translated by Tomaž Podobnikar)

Monte Carlo methods were rapidly developing in the last two decades. Mostly they are implying in cases when analytical methods are failed. Monte Carlo Methods comprise branch of experimental mathematics which is concerned with experiments on random numbers. Similar situations are occurring when performing spatial analyses in geographical information system, but we mostly ignore or simplify such problems. For example Monte Carlo methods are effective tools for error estimation and visualisation of geographical information system data for vector data, as well as for grid data.

Problems handled by Monte Carlo methods are of two types called probabilistic or deterministic according to whether or not they are directly concerned with behaviour and outcome of random processes. In case of a probabilistic problem the simplest Monte Carlo approach is to observe random numbers, chosen in such a way that they directly simulate the physical random processes of the original problem, and to infer the desired solution from the behaviour of these random numbers. The idea behind the Monte Carlo approach to deterministic problems is to exploit the strength of theoretical mathematics, its concern with abstraction and generality, while avoiding its associated weakness by replacing theory by experiment whenever the former falters. Being deterministic, this problem has no direct association with random process, but we can solve the deterministic problem numerically by a Monte Carlo simulation.

In this paper we describe deterministic approach of a Monte Carlo simulation for evaluation of statistical significance of population spatial distribution with regard to social or natural environment. Social environment may be reflected as viewsheds, cognitive landscapes, cost surfaces etc., while natural (physical) environment may be reflected as topographical data, soil or geological characteristics, sun exposition etc.

Practically statistical significance is tested with comparison index of appointed spatial sample (for example index of mean value, index of the nearest neighbour, coefficient of autocorrelation etc.) with  $n$  additional randomly simulated indexes using the same statistical population. In our case the index of the mean value is mostly employed.

Let spatial sample with  $k$  cases from population lie on subset  $S'$ . Additional  $n$  samples of size  $k$  may be selected at random from the same population ( $S_1, S_2, S_3, \dots, S_n$ ). For each sample, including the sample of interest particularly statistic index  $u$  (in our case mean value), is computed yielding  $n + 1$  values ( $u', u_1, u_2, u_3, \dots, u_n$ ). Assuming that each sample is realizable and equally probable outcome from population, the statistical significance of differences in the sample of interest relative to population may be estimated by ranking the values of  $u$  and computing  $p = R(u') : (n + 1)$ . If  $n = 999$  ( $n + 1 = 1000$  samples and a sample of real distribution) and rank of  $u$  for the sample of interest is  $R(u') \leq 50$  ( $\alpha \leq 5\%$ ) and  $R(u') \geq 50$  ( $\alpha \geq 95\%$ ), the significance of this outcome is either less than or equal to 0.5 or greater or equal to 0.95.

Practical case study of Monte Carlo simulation has been implied to the very regularly hilly region of »Polhograjsko hribovje«, about 25 km west of the capital of Slovenia, Ljubljana, measuring 8 by 6 km, encoded within raster composed of 100 by 100 m grid cells, for a finite population of  $K = 4800$  locations (48 km<sup>2</sup>). Raster data layers have been chose as following: height above sea level, distance from water springs and streams, relief slope and aspect, sun exposition, and karst or non-karst ground. Those geomorphological characteristics are visually very inexpressive. Within this region are registered  $k = 171$  houses. There are not perceived any larger settlements and population distribution seems to be insignificant.

First of all, we applied simple Monte Carlo simulation to test the distribution of the houses only. Geometrical centre of distribution has been computed. It is significantly shifted 490 m to the west and 285 m to the south with regard to geometrical centre of the area of interest. With Monte Carlo meth-

ods we evaluated also autocorrelation coefficient  $I$  (Moran's) of the distribution of houses. It has been significantly different than randomly set, so it is significantly that houses unite to clusters (villages).

After initial stages, we applied Monte Carlo simulation for evaluation of statistical significance of settlement analysis with regard to geomorphological characteristics. The main question was: Does relation between distribution of houses and geomorphological characteristics exist? We implied previously described method of Monte Carlo significance test with index of mean value for all six available raster data. In the article, we are describing test of houses distribution with regard to height above sea level only.

Mean value of height of the tested area is 678 m. Histogram of the heights is almost normal distributed. Mean value of heights of the houses is 677 m. There are noticed two inexpressive maximums, between 500 and 550 m (referring to the valleys), between and 650 and 850 m (referring to upper slopes and ridges of the hills). With Monte Carlo simulation ( $n = 999$ ) it has been evaluated that the references mean value is on the position 439 from the left (insignificant deviation from the mean height level of surface) what confirms assumption that average house is placed on the average height above sea level.

The results of all six data sets testing shows that the most important variables (data layers) which influence to the settlement distribution are: relief slope, aspect and sun exposition, partially also karst or non-karst ground and distance from water springs and streams, while height above sea level is less important. A recapitulation of the analyses could be a geographical interpretation: average (typical) house lies on a little slope relief, faces to the south, mainly on the non-karst ground and mostly on the hills. Another ascertainment (which not include settlements) is that on the hills there are not water streams and not many springs.

There are many traditional statistical approaches for significance tests, as well known  $\chi^2$  and  $t$ . What we gained with Monte Carlo methods are much better results compared to traditional, especially in cases when samples are small and insignificant or when complex population distributions are simulated. We applied comparative testing for described case study also with traditional approaches. Data layers were divided to parametrical (e. g. continuous sun exposition) and non-parametrical (e. g. categorically classified relief aspect). Results of the tested cases show that Monte Carlo method is much more reliable for such tests, mainly because it is not so sensitive to the size and the distribution of the sample. In such cases, we can much easier accept uncertain value at the boundary of decision as significant with traditional tests than with Monte Carlo Method, what could not lead to good solution of the problems.

In general, the Monte Carlo method is mostly recommended when we find out that results of previously applied traditional tests are ambiguous. In our case, the method proved oneself as to be very useful in most cases. It could be even more useful when some improvements are applied in simulation, especially if there are a priori known geographical characteristics.