

---

# METODE STATISTIČNIH PROSTORSKIH ANALIZ V GEOGRAFSKEM INFORMACIJSKEM SISTEMU

mag. Tomaž Podobnikar

ZRC SAZU – Inštitut za prostorske študije, Ljubljana

mag. Samo Drobne

FGG – Oddelek za geodezijo, Ljubljana

Prispelo za objavo: 1998-11-06

Pripravljeno za objavo: 1999-06-09

## Izvleček

Namen članka je predstaviti metode statističnih prostorskih analiz v geografskih informacijskih sistemih (GIS) oziroma opisati uporabo statističnih metod v kontekstu analitičnih sposobnosti tehnologije geografskih informacijskih sistemov. Najprej predstavimo vlogo statističnih prostorskih analiz v širšem področju prostorskih analiz v geografskem informacijskem sistemu. Statistične prostorske analize podrobneje obravnavamo po grafičnem (topološkem) pristopu.

**Gljučne besede:** geografski informacijski sistem, prostorske analize, statistične prostorske analize, statistika

## Abstract

This paper presents the methods of statistical spatial analysis in used geographic information systems (GIS) or to describe the use of statistical methods within the context of the analytical capacity of the GIS technology. The role of statistical spatial analyses within a wider field of spatial analyses in GIS is presented first. Statistical spatial analyses are discussed in detail using a graphical (topological) approach.

**Keywords:** GIS, spatial analysis, statistical spatial analysis, statistics

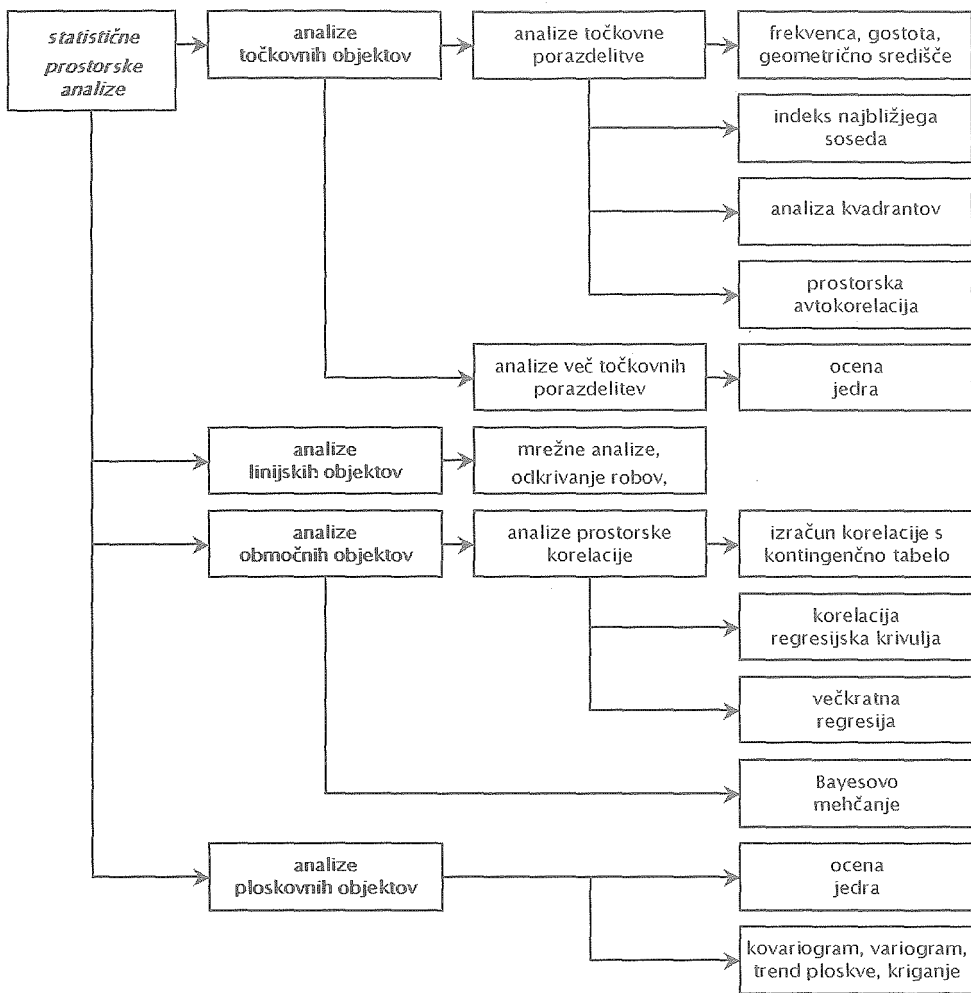
## 1 UVOD

V enem od prejšnjih Geodetskih vestnikov (Drobne et al., 1997) smo obravnavali splošen pregled prostorskih analiz v geografskih informacijskih sistemih. V tem prispevku podrobneje obravnavamo tisti del prostorskih analiz, ki temelji na statističnih metodah.<sup>1</sup> Ravno sposobnost analize podatkov loči prvo generacijo geografskih informacijskih sistemov od druge (Langran, 1989). Bailey je opredelil prostorske analize v geografskem informacijskem sistemu (1994) kot metode, s

katerimi analiziramo prostorske podatke in ustvarjamo nove informacije. Berry (1995), Manfred et al. (1996) ter Unwin (1997) pa so k tej definiciji dodali še splošno možnost upravljanja s prostorskimi podatki z namenom pridobivanja novih informacij.

Do danes sta se uveljavila predvsem dva pristopa izvajanja statističnih prostorskih analiz v geografskem informacijskem sistemu (Unwin, 1997). Po prvem pristopu izvajamo analize v posebnih statističnih paketih, v katere podatke uvozimo iz geografskega informacijskega sistema. Primera programskih paketov, ki podpirata takšen pristop, sta S+SpatialStats in SpaceStat<sup>2</sup>. Drugi pristop izkorišča nekatere že vgrajene (predvsem enostavnejše) statistične funkcije v orodjih geografskih informacijskih sistemov, kot so na primer: izračun števila opazovanj, vsote, maksimuma, minimuma, aritmetične in geometrične sredine, modusa, mediane, frekvenčne porazdelitve, standardnega odklona in variance. Vse več pa proizvajalci vgrajujejo v orodja geografskih informacijskih sistemov tudi že funkcije za izračun regresijskega modela, modela multiple regresije, trenda in avtokorelacije. Orodja, ki že vsebujejo nekatere statistične metode za analiziranje prostorskih podatkov, so na primer: Idrisi, Arc/Info, MapInfo in TNT<sup>3</sup>. Statistične prostorske analize opredelimo kot podskupino prostorskih analiz. Definiranje statističnih prostorskih analiz je zelo nevhvalno, saj mnoge metode prostorskih analiz, kot so npr. metode testiranja kakovosti podatkov, uporabljajo statistične metode, velikokrat obravnavamo kot posebno skupino metod. Metode statističnih prostorskih analiz bi lahko enostavno opredelili kot uporabo statističnih funkcij na prostorskih podatkih. V našem primeru nas zanimajo statistične analize, ki jih izvajamo na podatkih iz geografskih informacijskih sistemov, ali še bolje, vključimo med orodja za analizo podatkov v geografskih informacijskih sistemih.

Statistične analize prostorskih podatkov obravnavamo po grafičnem pristopu, torej glede na grafične (topološke) objektne tipe (Kvamme et al., 1997): točkovni objektni tip (0D), linijski objektni tip (1D), območni (arealni) objektni tip (2D), ploskovni objektni tip (3D). Slika 1 prikazuje nekaj skupin statističnih prostorskih analiz, opredeljenih po grafičnem pristopu. Večino jih obravnavamo v nadaljevanju. V članku pa je namenoma izpuščena velika skupina posebnih statističnih metod, ki jih uporabljamo pri analizi podatkov daljinsko zaznanih podob. Uporabljamo jih predvsem za razvrščanje podatkov v gruče, iz enega ali več podatkovnih slojev. Izpustili smo tudi obravnavo statističnih analiz, ki jih uporabljamo pri testiranju kakovosti prostorskih podatkov v geografskih informacijskih sistemih.



Slika 1: Grafični pristop delitve statističnih prostorskih analiz v geografskem informacijskem sistemu

## 2 STATISTIČNE ANALIZE TOČKOVNIH OBJEKTOV

Skupino statističnih prostorskih analiz točkovnih objektov lahko delimo na analize točkovne porazdelitve in analize več točkovnih porazdelitev (glej tudi sliko 1). V prvo skupino spadajo izračun indeksa najbližjega soseda, analiza kvadrantov, izračun prostorske avtokorelacije, v drugo skupino pa ocena jedra. V nadaljevanju obravnavamo le bolj enostavne analize, medtem ko si lahko bralec poišče opis metod zapletenejših statističnih analiz prostorskih podatkov v strokovni literaturi, npr. (Cressie, 1993; Bailey, Gatrell, 1995). Analize točkovnih vzorcev pogosto zahtevajo vzorčenje, še posebno, če je zbirka podatkov velika. V geografskem informacijskem sistemu poznamo dva načina vzorčenja, in sicer prostorsko in neprostorsko vzorčenje.

Pri prostorskem vzorčenju vzorčimo delovno (geografsko) območje in dobimo naključni dvodimenzionalni vzorec, pri neprostorskem pa vzorčimo podatke ne glede na položaj v prostoru.

Točke so ničrazsežni objekti, zato lahko njihovo porazdelitev merimo le kot število dogodkov v danem vzorcu s pripadajočimi položaji v prostoru. Pri tem predpostavljamo enakovrednost točkovnih objektov, kar poenostavi izvajanje analiz. Kljub temu pa lahko vzorcem točk v bolj zapletenih primerih pripišemo dodatne tematske podatke (vrsto dreves, naravo kriminala itd.). Pri analizah vzorcev točkovnih objektov obravnavamo tako vzorce točk v prostoru na splošno kot posebne primere obravnavanja (točk) rastrskih celic. Mere opisne statistike točkovnih objektov so lahko naslednje: frekvenca, gostota, geometrično središče, prostorska razpršenost (dispersija) in prostorska porazdelitev. Razen prostorske porazdelitve so metode izračuna prostorskih lastnosti skupine točk, omejene na splošno znane metode opisne statistike.

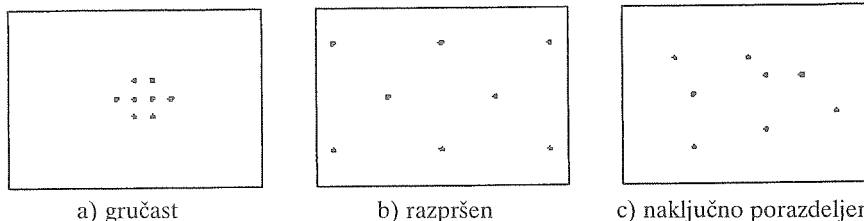
## 2.1 Frekvenca, gostota, geometrično središče in razpršenost točk

Frekvenca ali pogostost točkovne porazdelitve pomeni število točk na obravnavanem območju. Dva točkovna vzorca z isto frekvenco, ki ju obravnavamo na različno velikih območjih, lahko primerjamo z mero gostote vzorca. Ta predstavlja odnos med frekvenco točk in površino obravnavanega območja. Z geometričnim središčem in razpršenostjo (dispersijo) vzorca opišemo zgostitev točk na določenem delu obravnavanega območja. Geometrično središče porazdelitve izračunamo z aritmetično sredino koordinat  $x$  in  $y$ , prostorsko razpršenost pa lahko izrazimo s standardnim odklonom od aritmetične sredine.

## 2.2 Prostorska porazdelitev točk

Pogosto obravnavamo tri osnovne tipe točkovnih vzorcev (Chou, 1997; Slika 2), in sicer:

- gruče – točkovni objekti so zgoščeni na enem ali več manjših območjih,
- razpršen vzorec – pravilna porazdelitev in relativno velika razdalja med točkami,
- naključno porazdeljen vzorec – niti gručast niti razpršen vzorec.



Slika 2: Tipične skupine porazdelitve prostorskih točkovnih vzorcev

Prostorska porazdelitev točk je lahko še veliko bolj zapletena od omenjenih. Nekaj možnih analiz prostorske porazdelitve točk obravnavamo v nadaljevanju.

### 2.2.1 Indeks najbližjega soseda

Z indeksom najbližjega soseda (angl. nearest neighbor index – NNI) merimo stopnjo razpršenosti točk glede na minimalno razdaljo med obravnavanimi točkami. Indeks definiramo z upoštevanjem dejstva, da je povprečna razdalja med točkami v gručastem vzorcu krajša kot pri razpršenem (Chou, 1997). Algoritem izračuna indeksa najbližjega soseda najprej poišče vsaki točki najbližjega soseda ter izračuna razdaljo  $d_i$  med njima. Nato sledi izračun indeksa  $A_d$ , ki predstavlja povprečno razdaljo med najbližjimi sosedi točkovnega vzorca na obravnavanem območju. Parameter  $n$  v enačbi (1) predstavlja število točk, ki tvori vzorec na obravnavanem območju.

$$A_d = \frac{\sum_{i=1}^n d_i}{n} \quad (1)$$

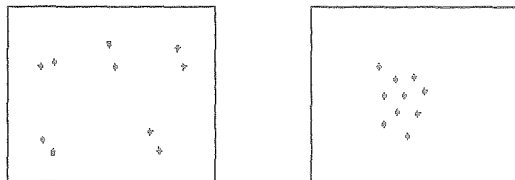
Manjši ko je indeks  $A_d$ , manjša je povprečna razdalja med najbližjimi točkami. Ob predpostavki, da je porazdelitev točk naključna, izračunamo pričakovano vrednost povprečne najbližje razdalje med točkovnimi objekti ( $E_d$ ):

$$E_d = \frac{1}{2} \sqrt{\frac{A}{n}}, \quad (2)$$

kjer  $A$  pomeni površino obravnavanega območja. Pričakovano polovično razdaljo dobimo iz razmerja  $A/n$ , ki ga določa vsaka točka na obravnavanem območju kot vrednost povprečne površine. Kvadratni koren te količine pretvori merjeno površino v medsebojno razdaljo med paroma sosednjih točk. Koefficient  $E_d$  je za kakršnokoli porazdelitev točk na obravnavanem območju vedno enak. Končno opredelimo indeks najbližjega soseda (NNI) kot kvocient  $A_d$  in  $E_d$ :

$$NNI = \frac{A_d}{E_d} \quad (0 \leq NNI \leq 2,1491). \quad (3)$$

V primeru, da vse točke sovpadajo, dobimo  $A_d = 0$  in  $NNI = 0$ . Manjša vrednost indeksa indeks najbližjega soseda (NNI) opisuje gručaste, večja pa razpršene vzorce. Pri vrednosti indeksa 1 obravnavamo porazdelitev vzorca kot naključno. Slaba stran indeksa najbližjega soseda (NNI) je, da ni občutljiv na zapletene vzorce. Primer takšnega vzorca je več ločenih gruč točk, ki so porazdeljene enako goste le v eni gruči, pri čemer indeks najbližjega soseda ne pokaže razlike (Slika 3).



Slika 3: Vzorca točk sta izrazito različna, vendar je njun indeks najbližjega soseda enak

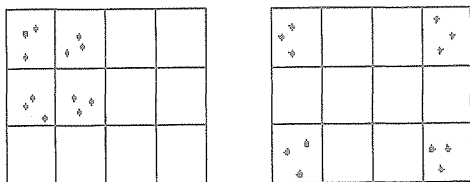
Indeks najbližjega soseda torej ne upošteva celotne prostorske porazdelitve točk. Za ugotavljanje vzorca lahko razdalje do najbližjega soseda izrazimo tudi z drugimi metodami, na primer z empirično kumulativno verjetnostno porazdelitveno funkcijo ali pa s funkcijo  $K$  (Bailey, Gatrell, 1995).

## 2.2.2 Analiza kvadrantov

Metode analize kvadrantov slonijo na seštevanju (frekvenca) oziroma porazdelitvi vzorca točk obravnavanega območja na podobno območja z enakimi površinami ali kvadranti (Bailey, Gatrell, 1995). Kvadrante razvrščamo glede na frekvenco. Za veljavnost analize mora vsak razred (kvadrant) vsebovati vsaj pet točkovnih objektov. Poznamo nekaj pristopov k takim analizam, ki se ločijo po tvorbi kvadrantov različnih velikosti ali oblik (največkrat je štirikotnikov) in po načinu postavitve kvadrantov na obravnavano območje za oblikovanje mreže celic (naključnega ali pravilnega rastra). Po preštevanju opazovanj po kvadrantih dobimo za vsak razred frekvenco  $f_i$ , kjer je  $i$  indeks posameznega razreda. Verjetnost, da je opazovanje v posameznem razredu, podamo z diskretno Poissonovo porazdelitveno funkcijo<sup>4</sup>:

$$p_x = \frac{v^x e^{-v}}{x!} \quad (v > 0), \quad (4)$$

kjer je  $x$  frekvenca v kvadrantu, pričakovana frekvenca v kvadrantu ( $v = n \cdot p$ ,  $n$  je število vseh opazovanj ( $n \gg 0$ ),  $p$  je verjetnost, da se opazovanje nahaja na obravnavanem območju) in  $e$  osnova naravnega logaritma. Analiza s kvadranti temelji na t. i. ničelni domnevi, ki pravi, da so točkovni objekti v Poissonovem procesu porazdeljeni naključno. Testiranje te domneve lahko izvedemo s testom  $\chi^2$ . Tudi analize kvadrantov imajo nekaj pomanjkljivosti. Metode slonijo le na ugotavljanju frekvence opazovanj po kvadrantih, pri tem pa ne upoštevajo prostorske porazdelitve kvadrantov (Slika 4).



Slika 4: Vzorca točk sta izrazito različna, vendar se po metodi kvadrantov ne razlikujeta

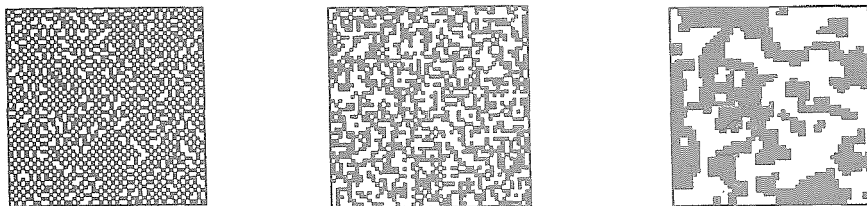
## 2.2.3 Prostorska avtokorelacija

Prostorska avtokorelacija je mera stopnje vpliva porazdelitve podobnih objektov v okolici objekta. Pri tovrstnem izračunu korelacije gre za obravnavanje vrednosti iste spremenljivke na različnih lokacijah – od tod izraz avtokorelacija. Prostorsko avtokorelacijo statistično največkrat opredelimo z Moranovim koeficientom  $I$ , redkeje pa z Gearyjevim koeficientom  $c$  (Chou, 1997; Bailey, Gatrell, 1995). Moranov koeficient oziroma indeks  $I$  definiramo:

$$I = \frac{n \sum_i \sum_j w_{ij} (x_i - \bar{X})(x_j - \bar{X})}{\left( \sum_i \sum_j w_{ij} \right) \left( \sum_i (x_i - \bar{X})^2 \right)} \quad (-1 \leq I \leq 1). \quad (5)$$

Pri tem je  $n$  število prostorskih enot (točk).  $w_{ij}$  predstavlja prostorski odnos med enotama  $i$  in  $j$  (v našem primeru rastrsko celico), kjer je:  $w_{ij} = 1$ , če sta celici  $i$  in  $j$  sosednji, in  $w_{ij} = 0$ , če nista.  $x_i$  predstavlja vrednost posameznega prostorskega pojava

in  $\bar{X}$  aritmetično sredino vseh vrednosti. Pozitivna vrednost koeficienta I blizu 1 pomeni, da se vzorec točk zbira v gručo, medtem ko predstavlja negativna vrednost blizu  $-1$  razpršen vzorec. Če se koeficient I ne razlikuje bistveno od 0, avtokorelacija ni značilna. Takšen prostorski vzorec pojmujeemo kot naključni vzorec (Slika 5).



a) koef. avtokorelacije  $I = -0,7$    b) koef. avtokorelacije  $I = 0$    c) koef. avtokorelacije  $I = 0,7$

*Slika 5: Različne stopnje prostorske avtokorelacije na dvojiških, naključno porazdeljenih rastrskih celicah (50 x 50 celic)*

Koeficient avtokorelacije je primeren za opis in analizo digitalnega modela reliefa (DMR) ali drugih ploskev v prostoru. Metode izdelave naključne in do določene stopnje avtokorelirane ploskve napak in odklonov digitalnega modela reliefa lahko uporabljamo pri prikazu njihovega vpliva na vidnost z določenih točk zemeljskega površja, pri prikazu spremembe nedoločenosti poteka optimalne poti, ki jo izračunamo z digitalnim modelom reliefa, za ponazoritev vpliva napak digitalnega modela reliefa na analize razvodij, za izračun stroškovnih ploskev glede na napake digitalnega modela reliefa itd. Večina omenjenih metod sloni na simulacijskih metodah Monte Carlo (Podobnikar, 1998 a, b). Moranov avtokorelacijski koeficient I pokaže razliko med prostorskima vzorcema, ki ju z analizo kvadrantov nismo mogli ločiti. Kljub temu se lahko rezultati z avtokorelacijo analiziranih vzorcev točk prav tako pokažejo kot nenatančni, največkrat zaradi nepravilnega definiranja ločljivosti in porazdelitve kvadrantov pri rastrskih točkah. Prav zato je priporočljivo Moranov koeficient I pri vrednotenju prostorske porazdelitve točk primerjati z indeksom najbližjega soseda (NNI) ter z rezultati analize kvadrantov.

### 3 STATISTIČNE ANALIZE LINIJSKIH OBJEKTOV

Statistične analize linijskih objektov so v splošnem mnogo bolj zahtevne kot druge tu omenjene. Enostavnejši je izračun mer, kot je na primer frekvenca presekov linij. Večina analiz linijskih objektov spada v širše področje prostorskih analiz v geografskem informacijskem sistemu, kjer so tukaj obravnavane statistične prostorske analize le njihova podmnožica. Deloma lahko med statistične prostorske analize štejemo mrežne analize, ki sicer temeljijo na teoriji grafov.

### 4 STATISTIČNE ANALIZE OBMOČNIH OBJEKTOV

Statističnimi analizami območnih (arealnih) objektov analiziramo odnose med sosednjimi območji v podatkovnem sloju. Površine območij lahko obravnavamo kot proizvod celic, razporejenih v pravilno mrežo ali v območja nepravilnih oblik. V to skupino analiz štejemo predvsem razne analize prostorske korelacije, kontingence ter regresije kot tudi ocene jedra ter razne analize štetja in razmerij, kamor spada na primer Bayesovo mehčanje (Bailey, Gatrell, 1995).

#### 4.1 Analiza povezanosti (za nominalne vrednosti)

Osnovni cilj tovrstne analize je odkrivanje odnosov med različnimi tipi prostorskih objektov, ki so organizirani v različnih podatkovnih slojih<sup>5</sup>. V primeru, da zasedajo spremenljivke v podatkovnih slojih nominalne (kategorične) vrednosti, potem lahko povezanost med njimi preverimo s kontingenčno tabelo (tabela dvodimenzionalne frekvenčne porazdelitve) in testa  $\chi^2$ . Vzemimo primer, ko želimo na obravnavanem območju ugotoviti povezanost tipa vegetacije in vrste tal (obe spremenljivki zasedata kategorične oz. nominalne vrednosti). S prekrivanjem dveh podatkovnih slojev in ugotavljanjem prekrivanja posameznih kategorij na obeh slojih sestavimo kontingenčno tabelo (glej preglednico 1).

		TIP VEGETACIJE (X)	
		gozd	ni gozda
TIP TAL (Y)	pesek	$f_{11}$	$f_{12}$
	ni peska	$f_{21}$	$f_{22}$

Preglednica 1: Primer kontingenčne tabele povezanosti tipa vegetacije in tipa tal

Sestavimo kontingenčno tabelo teoretičnih frekvenc  $f'_{ij}$  ter jih s statistiko  $\chi^2$  primerjamo z dejanskimi:

$$\chi^2 = \sum_{i=1}^{n_X} \sum_{j=1}^{n_Y} \frac{(f_{ij} - f'_{ij})^2}{f'_{ij}}, \quad (6)$$

kjer sta  $n_X$  in  $n_Y$  števili ravni naključnih spremenljivk X (tip vegetacije) in Y (tip tal),  $f_{ij}$  dejanske (empirične) ter  $f'_{ij}$  teoretične frekvence. Teoretična frekvenca je verjetnost  $P(X = x_i \cap Y = y_j)$  pomnožena s številom enot v vzorcu ( $n$ ):

$$f' = n \cdot P(X = x_i \cap Y = y_j) = n \cdot P(X = x_i) \cdot P(Y = y_j). \quad (7)$$

S primerjanjem eksperimentalne in tabelirane statistike  $\chi^2$  ugotovimo stopnjo zaupanja, s katero lahko trdimo, da sta podatkovna sloja oziroma spremenljivki povezani. V primeru, ko sta spremenljivki značilno povezani, lahko en podatkovni sloj (v našem primeru podatkovni sloj vegetacije) zavržemo iz nadaljnega postopka prostorskega modeliranja (Chou, 1997; Press et al., 1995).

#### 4.2 Korelacijski koeficient in linearna regresija

V primeru, da spremenljivki v dveh podatkovnih slojih zasedata zvezne vrednosti, lahko korelacijsko povezanost med njima analiziramo s Pearsonovim korelacijskim koeficientom in/ali regresijsko analizo. Pri tem Pearsonov korelacijski koeficient pokaže le obseg povezanosti med podatkovnima slojema, regresijski model pa tudi odvisnost. V postopkih prostorskega modeliranja zato pogosteje ocenjujemo parametre regresijskih modelov. Ti parametri namreč omogočijo kalibracijo in aplikacijo samih prostorskih modelov v GIS-u. Korelacijski koeficient v vzorcu opazovanj med dvema spremenljivkama X in Y izračunamo:

$$r_{XY} = \frac{\sum_i (x_i - \bar{X})(y_i - \bar{Y})}{n s_X s_Y} \quad (-1 \leq r_{XY} \leq 1), \quad (8)$$



kjer sta  $\bar{X}$  in  $\bar{Y}$  aritmetični sredini opazovanj,  $s_X$  in  $s_Y$  oceni standardnih deviacij spremenljivk  $X$  in  $Y$  in  $n$  število geografskih objektov.

Vzemimo primer izračuna linearne odvisnosti med razdaljo zemljišča do najbližje avtobusne postaje in prodajno ceno zemljišč. Podatke o obratni vrednosti razdalj parcel do najbližjih avtobusnih postajališč ter o prodajni ceni parcel pridobimo iz ustreznih podatkovnih slojev. Korelacijski koeficient izračunamo po enačbi (8). Vrednost korelacijskega koeficienta blizu +1 oziroma -1 pomeni, da spremenljivki močno korelirata (pozitivno oziroma negativno), vrednost blizu 0 pa ni v odvisnosti med spremenljivkama. Koeficient korelacije je navadno le eden izmed parametrov, ki ga izračunamo v postopku ocenitve parametrov linearnega regresijskega modela. Enačba (9) ponazarja splošen linearni regresijski model:

$$y_i = a + bx_i + e_i. \quad (9)$$

V primeru modeliranja linearne funkcijske zveze med oddaljenostjo zemljišč od avtobusnih postajališč in prodajno vrednostjo zemljišča predstavlja  $y$  prodajno ceno zemljišča,  $x$  obratno vrednost razdalje parcele do najbližjega avtobusnega postajališča,  $a$  konstanto premice,  $b$  naklon premice (prvi regresijski koeficient) ter  $e$  slučajne vplive (naključno porazdeljeno napako).

### 4.3 Multipla regresija

Pri proučevanju pojavov iz stvarnega sveta vpliva na rezultativni znak cel splet različnih bistvenih dejavnikov. Proučevanje odvisnosti  $y$  od enega samega znaka  $x$  je v tem primeru slabo, saj proučujemo vpliv drugih dejavnikov na odvisno spremenljivko  $y$  skupaj s slučajnimi vplivi. Zato pri reševanju tovrstnih problemov iz stvarnega sveta navadno uporabljamo metode multiple regresijske analize. Primer multiple regresijskega modela je odvisnost močne erozije površja od spremenljivk okolja, ki niso podane kot nominalne vrednosti, kot so na primer naklon površja, gostota vegetacije (ne tip vegetacije, ki je navadno nominalna vrednost), padavine itd. Multiple regresijske analize se uporabljamo tudi za obrazložitev variacije prostorskih zveznih vrednostih ali za interpolacijo vrednosti med vzorci točk v prostoru (Bailey, Gatrell, 1995).

Multipli regresijski model lahko zapišemo kot<sup>6</sup>:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + e, \quad (10)$$

kjer so  $Y$  proučevana (odvisna) spremenljivka,  $X_i$   $i$ -ta pojasnjevalna (neodvisna) spremenljivka,  $\beta_0$  presečišče,  $\beta_1$  pričakovani parameter spremenljivke  $X_i$  ( $i = 1, 2, \dots, n$ ) in  $e$  slučajni vplivi. Enačbo (10) poenostavimo tako, da postavimo  $X_1 = 1$ , in dobimo:

$$Y = X\beta + E, \quad (11)$$

kjer so  $Y$  vektor razsežnosti  $m \times 1$ , ( $m$  je število opazovanj),  $X$  je matrika razsežnosti  $m \times n$ , ( $n$  je število neodvisnih spremenljivk skupaj s konstantno vrednostjo  $X_1$ ),  $\beta$  je vektor razsežnosti  $n \times 1$  in  $E$  je vektor razsežnosti  $m \times 1$ . Ocene parametrov izračunamo po metodi najmanjših kvadratov, po enačbi (12):

$$\hat{\beta} = (X^T X)^{-1} X^T Y. \quad (12)$$

Izračun multiple regresije je zelo zahteven, zato večina statističnih paketov (npr. SAS,

SPSS)<sup>7</sup> omogoča izračun ocene pričakovane vrednosti spremenljivke  $\hat{\beta}_i$  ( $i = 1, 2, \dots, n$ ). Ta podatek (rezultat) lahko uporabimo kot relativno utež pri porazdelitvi odvisne spremenljivke. Sledi ključna odločitev o vključitvi tistih parametrov v model, ki so značilni in s tem nujni. Navadno vključimo v prostorski model majhno število kritičnih spremenljivk, s katerimi dobimo zadovoljiv rezultat.

#### 4.4 Logistična regresija

Pri logistični regresiji gre za povezavo med metodami regresijske analize in dvojiško logiko. V številnih primerih lahko prostorske pojave opišemo le s kategoričnimi (nominalnimi) vrednostmi. Logistična regresijska analiza je uporabna v postopkih prostorskega modeliranja, če je porazdelitev odvisne spremenljivke merjena z nominalnimi vrednostmi, medtem ko so pojasnjevalne (neodvisne) spremenljivke izražene z zveznimi vrednostmi. V takšnih primerih metoda multiple regresijske analize odpove. Primer uporabe logistične regresijske analize v postopku prostorskega modeliranja v geografskem informacijskem sistemu je študij odvisnosti požarov v naravi od več parametrov. Požgana in nepožgana območja definiramo kot nominalne vrednosti z dvojiško logiko (1 ali 0), verjetne pojasnjevalne parametre za nastanek požara pa kot zvezne vrednosti temperature, padavin, bližine cest itd.

Po zasnovi izhajajo logistična regresija iz metod verjetnostnega računa. Verjetnost obstoja nekega geografskega pojava (npr. požara) na obravnavanem območju označimo s  $P_a$  ter verjetnost odsotnosti s  $P_b$ , kjer velja  $P_a + P_b = 1$ . Metoda logistične regresije je torej primerna za izdelavo verjetnostne matrike. Logistični regresijski model je opredeljen z naslednjimi enačbami (Chou, 1997):

$$P_a = \frac{e^{U_a}}{1 + e^{U_a}} \quad (13)$$

$$U_a = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + e, \quad (14)$$

kjer so  $U_a$  količina izražena z linearno kombinacijo pojasnjevalnih spremenljivk  $X_1, X_2, \dots, X_n$  (pogosto imenovana tudi funkcija koristi dogodka  $a$ ), bi je ocena spremenljivke  $X_i$  in  $e$  slučajni vplivi. Večja je vrednost  $U_a$ , večja je verjetnost dogodka  $a$ .

## 5 STATISTIČNE ANALIZE PLOSKEV

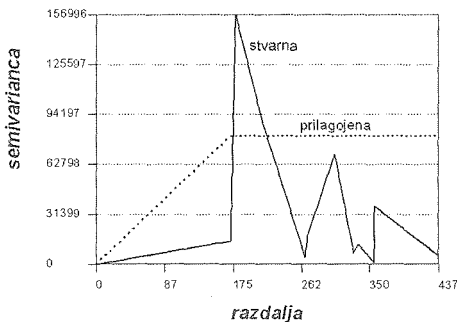
Pri statističnih analizah ploskev v prostoru gre večinoma za obravnavo točkovnih objektov (meritev), ki predstavljajo zvezne pojave v prostoru, s statističnimi metodami (Bailey, Gatrell, 1995). Medtem ko nas pri analizah vzorcev točk in območij v prostoru zanima lokacija opazovanj, nas pri analizah ploskev (zveznih vrednosti v prostoru) zanimajo predvsem porazdelitveni vzorci vrednosti atributa v prostoru. Poleg drugih metod uvrščamo med statistične analize ploskev še metode, ki temeljijo na teselaciji (delitvi prostora na izbrane like), kot so izdelava trikotniške nepravilne mreže (angl. triangulated irregular network – TIN) ali izračun Thiessenovih poligonov, metode ocene jedra ter razne metode modeliranja prostorskih zveznih vrednosti, kamor štejemo tudi izračun trenda površin ter kriging. V nadaljevanju obravnavamo le nekatere med njimi.

## 5.1 Ocena jedra

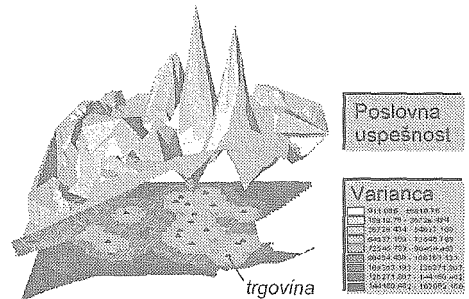
Z metodo ocene jedra pri prostorsko zveznih vrednostih iščemo srednjo vrednost  $Z_{\mu}(x_i, y_i)$  funkcijskih vrednosti ploskve, katerih vrednosti smo vzorčili za lokacije  $(x_i, y_i)$  v polmeru  $\tau$  (Bailey, Gattrel, 1995). Velikost polmera  $\tau$  ( $\tau > 0$ ) vpliva na stopnjo glajenja ploskve. Srednjo vrednost  $\mu(x_i, y_i)$  izrazimo kot kvocient skupnega števila atributov na enoto površine in števila opazovanj na enoto površine. Za razliko med oceno jedra pri točkovnih prostorskih vzorcih, kjer nas je zanimala intenziteta točkovnih vzorcev oziroma število opazovanj na enoto površine, gre pri oceni jedra pri prostorskih zveznih vrednostih za določitev srednje vrednosti  $\mu(x_i, y_i)$  tematskih podatkov, katerih vrednosti so bile vzorčene za lokacije  $(x_i, y_i)$ .

## 5.2 Kriging

Kriging (tudi kriganje) in spremljajoče vmesne rezultate – prostorske statistične Kmere razpršenosti pojava, kot so variogram, semivariogram in kovariogram ter odstopanja od variograma – uporabljamo pri modeliranju oziroma predikciji prostorskih zveznih vrednosti ali ploskev na osnovi danih točk, območij ali volumnov vzorcev. Za interpolacijo s krigingom je treba imeti veliko predhodnega znanja o prostorski statistiki. Kriging je optimalna metoda prostorske linearne predikcije, pri kateri ocenjujemo vrednosti z najboljšo linearno nepristransko oceno ali z utežnim linearnim premikanjem povprečja (Cressie, 1993). Na sliki 7, ki prikazuje po krigingu izračunan model poslovne uspešnosti trgovin z živili v starem mestnem jedru Kopra leta 1992, so nazorno poudarjene lokalne posebnosti; pod tridimenzionalnim modelom poslovne uspešnosti je dvodimenzionalna ploskev semivariograma, vključno z lokacijami trgovin.



a) semivariogram



b) lokacije trgovin, poslovna uspešnost in semivariogram

Slika 6: S krigingom izračunan model poslovne uspešnosti trgovin z živili na območju starega mestnega jedra Kopra leta 1992 (pogled z severa)<sup>8</sup>

## 6 ZAKLJUČEK

Med prostorskimi analizami v geografskem informacijskem sistemu zasledimo čedalje več statističnih analiz, prilagojenih zahtevam prostorskih podatkov in pričakovanim rezultatom. Tako lahko tudi v prihodnje pričakujemo vedno več

zapletenih statističnih funkcij, vključenih v standardna orodja geografskih informacijskih sistemov. Namen tega prispevka je bilo pregledno seznaniti bralce z metodami statističnih prostorskih analiz. Pri tem so bile predstavljene predvsem metode, ki jih lahko uporabljamo v orodjih standardnih geografskih informacijskih sistemov.

**Zahvala:** avtorja se zahvaljujeva recenzentoma, dr. Marku Krevsu in mag. Daliborju Radovanu, za ustvarjalne pripombe.

#### Literatura:

- Bailey, T. C., *A review of spatial statistical analysis in GIS*. Fotheringham, A.S., Rogerson, P.A. (eds.), *Spatial Analysis and GIS*. Taylor & Francis, London, 1994
- Bailey, T. C., Gatrell A. C., *Interactive Spatial Data Analysis*. Longman, London, 1995
- Berry J. K., *Spatial Reasoning for Effective GIS*. GIS World Books, Fort Collins, Colorado, 1995
- Chou, Y.-H., *Exploring Spatial Analysis in Geographic Information Systems*. OnWord Press, Santa Fe, 1997.
- Cressie, N. A. C., *Statistics for Spatial Data*. John Wiley & Sons, Inc, New York, 1993
- Drobne, S., Bogataj, M., *Lokacijski parametri v nalogah lastrinskega preoblikovanja trgovin na drobno*. IB revija, Ljubljana, 1995, XXIX (4-5), str. 46-57
- Drobne, S., *Katere formule so temelj prostorskim analizam v GIS-u? Geografski informacijski sistemi v Sloveniji 1997-98*. Zbornik referatov. Ljubljana, 1998
- Drobne, S., Podobnikar, T., Marini, S., *Prostorske analize v geografskih informacijskih sistemih*. Geodetski vestnik, Ljubljana, 1997, letnik 41, št. 4, str. 291-301
- Kvamme, K., Oštir-Sedej, K., Stančič, Z., Šumrada, R., *Geografski Informacijski Sistemi*. Znanstvenoraziskovalni center Slovenske akademije znanosti in umetnosti, Ljubljana, 1997
- Langran, G., *A Review of Temporal Database Research and Its Uses in GIS applications*. *International Journal of Geographical Information Systems*, 1989, št. 3, str. 215-232
- Manfred, M., Fischer, M., Scholten, H.J., Unwin, D., *Geographic Information Systems, Spatial Data Analysis and Spatial Modelling: An Introduction*. V M. Fischer, H.J. Scholten in D. Unwin, *Spatial Analytical Perspectives on GIS, GISDATA series: 4*. Taylor & Francis, London, 1996, str. 3-20
- Podobnikar T., *Monte Carlo simulacije napak digitalnega modela višin*. *Geografski informacijski sistemi v Sloveniji 1997-98*. Zbornik referatov, Ljubljana, 1998a
- Podobnikar, T. *Metode Monte Carlo simulacij v prostorskih analizah*. Magistrska naloga, Ljubljana, FG, Oddelek za geodezijo, 1998b
- Press, H. W., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P., *Numerical Recipes in C. The Art of Scientific Computing*, Cambridge University Press, druga izdaja, <http://cfatab.harvard.edu/nr/nronline.html>, (via internet), 1995
- Unwin, D., *GIS and Spatial Statistical Analysis*. V M.Craglia in H.Couclelis, *Geographic Information Research – Bridging the Atlantic*. Taylor & Francis, London, 1997, str. 399-411

#### Opombe

- 1 Glej tudi članek o pregledu nekaterih temeljnih formul v GIS-u (Drobne, 1998).
- 2 S+SpatialStats je zaščitena blagovna znamka MathSoft, Inc., SpaceStat je programska oprema avtorja Luc Anselina, razvita na National Center for Geographic Information and Analysis, University of California, Santa Barbara.
- 3 Idrisi je zaščitena blagovna znamka Clark Labs, Clark University. Arc/Info je zaščitena blagovna znamka ESRI Inc. MapInfo je zaščitena blagovna znamka MapInfo Corporation, TNT je zaščitena blagovna znamka MicroImages, Inc.
- 4 Poissonovo porazdelitev pogosto uporabljamo kot približek binomske porazdelitve diskretnih vrednosti. Ponazarja odnos med predpostavljeno verjetnostjo in številom resničnih dogodkov.

- 5 Razumevanje odnosov med različnimi prostorskimi objekti je pomembno izhodišče pri prostorskem modeliranju.
- 6 Spremenljivke in konstante, ki predstavljajo podatkovne sloje, pišemo z velikimi črkami (raster je matrika vrednosti).
- 7 SAS je zaščitena blagovna znamka SAS Institute, Inc., SPSS je zaščitena blagovna znamka SPSS, Inc.
- 8 Poslovna uspešnost je definirana v (Bogataj, Drobne, 1995).

*Recenzija: dr. Marko Krevs  
mag. Dalibor Radovan*