

UNIVERZA V LJUBLJANI
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Klavdija Krivec

**Vizualizacija in analiza prostorskih podatkov
z orodjem ArcMap**

DIPLOMSKO DELO

NA VISOKOŠOLSLEM STROKOVNEM ŠTUDIJU

Ljubljana, 2010



Št. naloge: 00504/2010

Datum: 15.03.2010

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko izdaja naslednjo nalogo:

Kandidat: **KLAVDIJA KRIVEC**

Naslov: **VIZUALIZACIJA IN ANALIZA PROSTORSKIH PODATKOV Z ORODJEM
ARCMAP**

VISUALIZATION AND SPATIAL DATA ANALYSIS WITH ARCMAP

Vrsta naloge: Diplomsko delo visokošolskega strokovnega študija

Tematika naloge:

Orodje ArcMap je del programske opreme ArcGIS, ki je svetovno prevladujoč nabor orodij, namenjenih delu s prostorskimi podatki. Kandidatka naj predstavi možnosti vizualizacije prostorskih podatkov, izvajanja geostatističnih analiz in vizualizacije analitičnih rezultatov v tem orodju. Opiše naj možnosti razširitev nabora analitičnih metod z lastnimi metodami. Za ilustracijo opisanih možnosti naj kandidatka pridobi primerne podatke s področja Slovenije in širše, ter na primerih prikaže vizualizacijske in analitične možnosti orodja ArcMap.

Mentor:


doc. dr. Matjaž Kukar



Dekan:


prof. dr. Franc Solina

UNIVERZA V LJUBLJANI
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Klavdija Krivec

**Vizualizacija in analiza prostorskih podatkov
z orodjem ArcMap**

DIPLOMSKO DELO

NA VISOKOŠOLSKEM STROKOVNEM ŠTUDIJU

Mentor: doc. dr. Matjaž Kukar

Ljubljana, 2010

IZJAVA O AVTORSTVU

diplomskega dela

Spodaj podpisani/-a **Klavdija Krivec**,

z vpisno številko **63050448**,

sem avtor/-ica diplomskega dela z naslovom:

Vizualizacija in analiza prostorskih podatkov z orodjem ArcMap

S svojim podpisom zagotavljam, da:

- sem diplomsko delo izdelal/-a samostojno pod mentorstvom (naziv, ime in priimek)

doc.dr. Matjaža Kukarja

- so elektronska oblika diplomskega dela, naslov (slov., angl.), povzetek (slov., angl.) ter ključne besede (slov., angl.) identični s tiskano obliko diplomskega dela
- soglašam z javno objavo elektronske oblike diplomskega dela v zbirki »Dela FRI«.

V Ljubljani, dne _____ Podpis avtorja/-ice: _____

ZAHVALA

Za izkazano strokovno podporo, nasvete in pomoč pri izdelavi naloge se zahvaljujem mentorju, doc. dr. Matjažu Kukar.

Svoji družini ter še posebej svoji sestri, Mojci Krivec, se zahvaljujem za potrpežljivost ter vsestransko podporo pri študiju.

Nenazadnje bi se rada zahvalila tudi vsem prijateljem, ki so me pri pisanju podpirali z vzpodbudnimi besedami. Brez njih te naloge sploh ne bi bilo, brez drugih bi bila drugačna. Hvala vsem.

KAZALO

POVZETEK	1
ABSTRACT	2
1 UVOD	3
1.1 Geografski informacijski sistem.....	4
1.1.1 Opredelitev pojma geografski informacijski sistem.....	4
1.1.2 Prostorski podatki	5
1.1.3 Podatkovna baza.....	7
1.2 Programska oprema ArcGIS za delo s prostorskimi podatki	8
1.2.1 ArcGIS produkti.....	8
1.3 Statistični pristopi v analizi prostorskih podatkov.....	11
1.3.1 Razširitev Geostatistical Analyst.....	12
1.4 Samostojen razvoj razširitev	13
1.5 Podatkovno rudarjenje na prostorskih podatkih	14
1.5.1 Druge razširitve v ArcMap-u.....	15
1.6 Izmenjava podatkov iz ArcMap-a z zunanji orodji	18
2 TESTNI PODATKI	19
3 VIZUALIZACIJA PODATKOV	20
3.1 Priprava podatkov in združevanje prostorskih podatkov s podatki brez prostorskih atributov	21
3.2 Možnosti vizualizacije	25
4 GEOSTATISTIKA IN RAZŠIRITVENI MODUL GEOSTATISTICAL ANALYST	31
4.1 Predstavitev podatkov	31
4.2 Kreiranje površine povprečne koncentracije ozona	32
4.3 Porazdelitev podatkov	37
4.4 Identifikacija trendov.....	40
4.5 Karta verjetnosti, kjer vrednost atributa presega kritično mejo.....	42
4.6 Končna karta s površino povprečne koncentracije ozona in območji preseganja kritične meje ozona.....	44
5 ZAKLJUČEK	49
6 KAZALO SLIK.....	50
7 LITERATURA	52

SEZNAM KRATIC IN SIMBOLOV

GIS	<i>Geographic(al) information system</i>
NCGIA	<i>National Center for Geographic Information and Analysis</i>
ArcSDE	<i>Spatial Database Engine</i>
ArcIMS	<i>Internet Mapping Server</i>
GPS	<i>Global Positioning System</i>
ADF	<i>Application Development Framework</i>
API	<i>Application Programming Interface</i>
XML	<i>Language</i>
EDN	<i>ESRI Developer Network</i>
COM	<i>Component object model</i>
ETL	<i>Extract, transform, load</i>

POVZETEK

Vizualizacija in ustrezne analize prostorskih podatkov predstavljajo pomemben vir informacij, ki so v pomoč pri razumevanju in napovedovanju prostorskih pojavov ter vplivajo na sprejemanje prostorskih odločitev. Za učinkovito odkrivanje informacij, ki jih vsebujejo prostorski podatki, je danes na voljo mnogo programskih rešitev. V diplomski nalogi sem raziskala zmožnosti orodja ArcMap. Orodje je del programske opreme ArcGIS, namenjene delu s prostorskimi podatki. Cilj diplomske naloge je bil predstaviti možnosti vizualizacije ter statistične analize prostorskih podatkov v tem orodju.

Za vizualizacijo ter izvajanje analiz prostorskih podatkov v ArcMap-u morajo biti podatki podani v ustrezni obliki. Prostorskim podatkom je mogoče pridružiti podatke brez prostorskih atributov ter tako povečati vsebino informacije. V diplomski nalogi je predstavljena priprava prostorskih podatkov ter povezava z neprostorskimi podatki. Prikazani so različni načini vizualizacije na karti.

V diplomski nalogi sem se osredotočila na statistične pristope v ArcMap-u. Na praktičnem primeru sem predstavila zmožnosti ene izmed razširitev ArcMap-a, Geostatistical Analyst. Razširitev temelji na metodah za prostorsko interpolacijo ter omogoča raziskovanje statističnih lastnosti prostorskih podatkov, izbiranje modela za kreiranje površine ter ovrednotenje, kako dobro je model napovedal neznane vrednosti. Vsi omenjeni koraki so predstavljeni na konkretnem primeru.

ArcMap je vsestranska aplikacija z mnogo funkcionalnostmi. Na podlagi preizkušanja raznih načinov vizualizacije ter statistične analize prostorskih podatkov lahko povzamem, da je za učinkovito rabo tega orodja zaželen osnovni nivo znanja s tega področja oziroma pripravljenost na sprotno učenje. Posebej pomembna je predpriprava podatkov v ustrezno obliko za izvajanje analiz.

Ključne besede:

- ArcMap,
- ArcGIS,
- vizualizacija prostorskih podatkov,
- geostatistika,
- geografski informacijski sistemi.

ABSTRACT

Spatial data visualization and analyses are an important source of information that are helpful in understanding and prediction of spatial phenomena and have an impact on spatial decisions. Nowadays there are many software solutions available for an effective detection of information contained in spatial data. In my thesis I explore the capabilities of ArcMap tool, which is a part of the ArcGIS software designed to work with spatial data. The goal of this study was to present the capabilities of visualization and statistical analysis of spatial data in this tool.

To analyze and visualize spatial data in ArcMap, data must be given in the proper format. It is possible to join spatial data and data without spatial attributes in order to increase the content of information. In the thesis itself the preparation of spatial data and its connection with non-spatial data is presented. Various ways of visualizing on the map are shown.

In my thesis I have focused on statistical approaches in ArcMap. In a case study I presented the capabilities of one of the ArcMap extensions, called Geostatistical Analyst. The extension is based on methods for spatial interpolation and allows exploration of statistical properties of spatial data, choosing a model for creating surface and evaluating how well the model predicted the surface. All these steps are presented on a practical example.

ArcMap is a comprehensive application with a lot of functionalities. Based on testing various methods of visualization and statistical analysis of spatial data, I can summarize that for an effective use of this tool a basic level of knowledge in this area is welcome or at least preparation for ongoing learning. Moreover to perform analyses it is especially important to prepare the data beforehand.

Keywords:

- ArcMap,
- ArcGIS,
- spatial data visualization,
- geostatistics,
- geographic information systems.

1 UVOD

V današnjem času, ko imamo na voljo ogromno podatkov, na podlagi katerih prihajamo do informacij, vizualizacija podatkov postaja nujnost. Vizualni prikaz podatkov je neprimerno bolj učinkovit in primeren kot numerična predstavitev, saj okoli sedemdeset odstotkov informacij sprejmemo na ta način [19]. Ravno zato postaja poudarek na vizualni predstavitvi podatkov vedno večji. Podatki, predstavljeni na grafični način, omogočajo boljše povezovanje in preglednost ter kvalitetnejše razumevanje posredovane informacije. V diplomski nalogi bom obravnavala področje vizualizacije podatkov, vezanih na prostor.

Po nekaterih ocenah se preko osemdeset odstotkov vseh odločitev človeka oziroma institucij posredno ali neposredno nanaša na prostor. Z razvojem informacijske tehnologije in posledično geografskih informacijskih sistemov so postali prostorsko opredeljeni podatki izredno pomembna podpora pri sprejemanju številnih odločitev človeka oziroma družbe in spremljanju dejavnosti v prostoru, saj jih je mogoče analizirati in spremeniti v nove informacije. Metodološki pristopi k analizi podatkov ter sistematično spremljanje oziroma opisovanje prostorskih podatkov omogočajo kakovostno uporabo in povezovanje številnih prostorskih in neprostorskih podatkov.

Na voljo so razna programska orodja, s katerimi lahko prostorske podatke analiziramo, vizualiziramo, razkrivamo njihove medsebojne odnose in napovedujemo trende. Posebej zanimive in koristne so možnosti interpoliranja vrednosti za poljubne točke v prostoru, kjer ni meritev, pri čemer ima velik pomen statistična analiza prostorskih podatkov oziroma geostatistika, katere del bomo raziskali na konkretnem primeru v okviru diplomske naloge.

V diplomski nalogi si bomo ogledali področje vizualizacije prostorskih podatkov. V praktičnem delu diplome bomo obravnavali grafični način prikaza prostorskih podatkov v obliki kart in grafov. Posebej se bomo posvetili pomenu povezovanja prostorskih podatkov s podatki brez prostorskih atributov. Ogledali si bomo analizo konkretnega primera, kjer bodo predstavljeni geostatistični pristopi pri analizi podatkov ter vizualizacija rezultatov analize. Program, ki ga bomo pri tem uporabili in je predstavljen v diplomski nalogi, je aplikacija ArcMap iz programskega paketa ArcGis orodij, namenjenih delu s prostorskimi podatki.

1.1 Geografski informacijski sistem

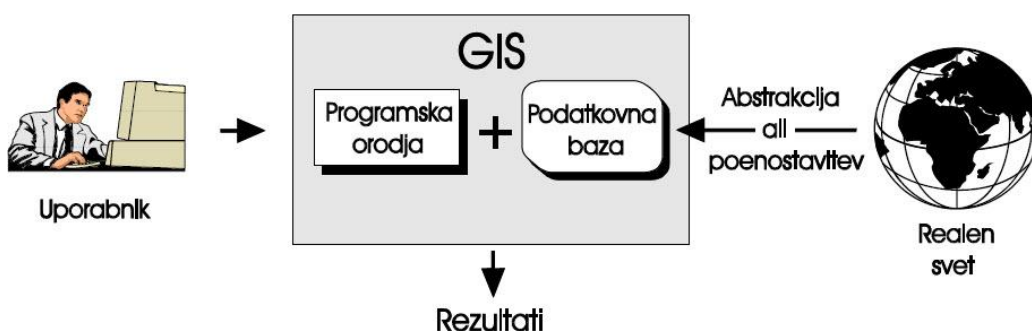
Zaradi preučevanja neposrednega človekovega okolja je geografija družbeno zelo pomembna znanost. Gospodarski razvoj sodobnega sveta prinaša korenite spremembe v geografsko okolje [21]. Geografski pristop preučuje povezanost in pomen dejavnikov, ki oblikujejo prostor, in je nepogrešljiv za študijo vpliva vedno obsežnejših človekovih posegov v naravno okolje. Posredno je izjemnega pomena pri različnih vrstah gospodarskega in prostorskega načrtovanja.

Geografski pristop k reševanju problemov zagotavlja boljše poznavanje sveta in razumevanje drugih kultur in družb ter posledično vodi k boljši komunikaciji ter potencialnemu sožitju in sodelovanju. Širok spekter predmetov preučevanja geografije omogoča poznavanje globalnega gospodarstva, političnih razmer in stanja tako naravnega kot družbenega okolja. Geografsko orientirane oziroma prostorske podatke shranjujemo, urejamo, analiziramo in prikazujemo s pomočjo geografskih informacijskih sistemov.

1.1.1 Opredelitev pojma geografski informacijski sistem

Geografski informacijski sistem, krajše GIS (*angl. geographic(al) information system*), je kompleksen pojem, za katerega se je v svetu vzporedno razvijalo več definicij.

Drobne navaja tri definicije tega pojma, tehnološko, procesno funkcijsko in vsebinsko definicijo [21]. Po *tehnološki definiciji* je geografski informacijski sistem opredeljen kot zbirka tehnoloških orodij za zbiranje, shranjevanje, iskanje, pretvorbe in prikazovanje prostorskih podatkov za določene namene. Geografski informacijski sistem je sinonim za ustrezno računalniško tehnologijo, metodologijo in orodja, ki so nastala in se razvila z integracijo tehnologije podatkovnih baz ter računalniško podprte kartografije.



Slika 1: Geografski informacijski sistem po tehnološki opredelitvi.

Procesno funkcijska definicija poudarja pomen podatkovnih analiz in opredeljuje GIS kot računalniško podprti prostorski informacijski sistem, ki nudi upravljavsko, organizacijsko in poslovno osnovo za zajemanje, shranjevanje, iskanje, obdelavo, analiziranje, prikazovanje ter distribucijo prostorskih podatkov.

Vsebinska definicija se osredotoča na sestavne komponente geografskega informacijskega sistema, ki so naslednje: splošna in posebna strojna oprema, sistemska in posebna programska oprema, sistem uporabniških aplikacij, integrirana podatkovna baza prostorskih podatkov ter vzdrževalci in uporabniki informacijskega sistema.

NCGIA (*angl. National Center for Geographic Information and Analysis*) opredeljuje geografski informacijski sistem kot skupek strojne opreme, programske opreme in postopkov, ki omogočajo urejanje, upravljanje, analiziranje, modeliranje, predstavitev in prikaz geografsko referenciranih podatkov z namenom reševanja kompleksnih problemov planiranja in upravljanja virov [17].

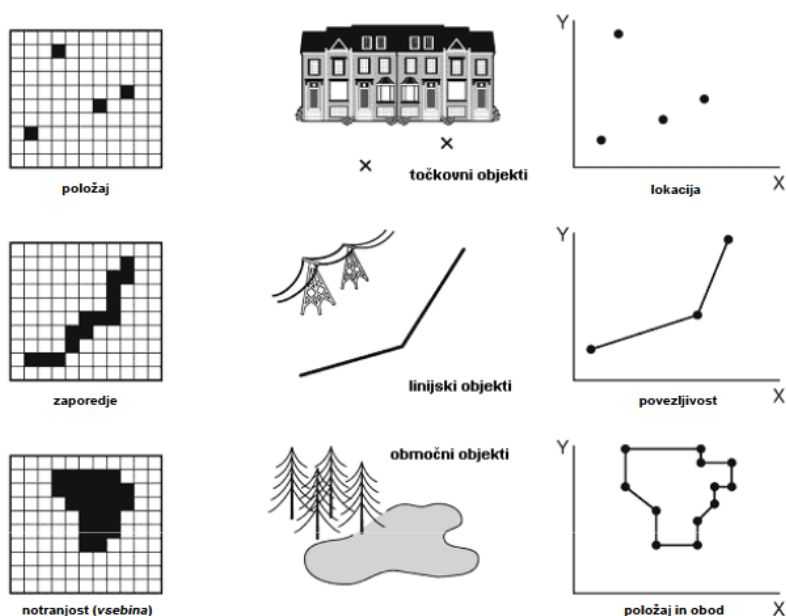
GIS je učinkovito orodje za upravljanje s prostorskimi podatki ter izvajanje prostorskih analiz. GIS lahko opredelimo kot pametno karto, ki nam omogoča pridobivanje odgovorov na najrazličnejša vprašanja, naprimer katera področja imajo primerno kvaliteto tal in so na prisojnih pobočjih, da bi jih lahko uporabili za vinograde; kako postaviti radijske oddajnike za optimalno pokritost prebivalstva in podobno. GIS ne odgovarja zgolj na enostavna vprašanja, ki se tičejo pozicije, pač pa kombinira najrazličnejše podatke - tako prostorske kot neprostorske [19]. Zato so prostorski podatki v današnjem času postali osnova za kvalitetno odločanje.

1.1.2 Prostorski podatki

Geografski informacijski sistem vsebuje prostorske podatke. Ti so opredeljeni z geometrični in alfanumeričnimi atributi. Med geometrične attribute sodijo oblika, lega in velikost, alfanumerični atributi pa so pripadajoče lastnosti [16].

Prostorski podatki so lahko grafično organizirani v vektorski ali rastrski obliki [16]. *Rastrske podatke* si najlažje predstavljamo kot funkcionalno povezano dvodimenzionalno pravokotno mrežo celic ali slikovnih elementov (pikslov), kjer je vsaki celici prirejena ena številka. Ta številka pomeni vrednost nekega parametra, položaj te številke v mreži pa geografski položaj njene vrednosti relativno glede na ostale vrednosti v mreži. Velikost celice določa pravokotno območje v naravi, za katerega velja vrednost celice. Celica v satelitski rastrski sliki ima lahko velikost deset metrov, kar pomeni, da predstavlja področje v naravni velikosti 10 x 10 metrov. Bistveno za rastrske podatke je, da jih lahko zelo učinkovito in nazorno prikažemo grafično.

Uporaba *vektorskih podatkov* prevladuje med GIS-i. Vektorska organizacija temelji na treh osnovnih grafičnih gradnikih: točkah, linijah in likih. Grafični gradniki so opredeljeni s točkami v koordinatnem sistemu in povezavami med točkami. Na prostorsko opredeljene grafične gradnike se navezujejo ostali opisni ali atributni podatki.



Slika 2: Primerjava rastrske in vektorske ponazoritve točkovnih, linijskih in območnih objektov.

Rastrske in vektorske podatkovne sloje pojmuje kot dve različni izvedbi kartografskega pristopa. Sodobni GIS sistemi lahko hkrati manipulirajo z rastrskimi in vektorskimi geografskimi podatki. Vektorski model se lažje prilagaja načelu tridimenzionalnega modeliranja prostora kot izrazito ravninski rastrski pristop. Dostopnost do rastrskih podatkov, njihova velika količina in potrebne predobdelave so glavne pomanjkljivosti rastrskega podatkovnega modela v primerjavi z vektorskim. Vektorski podatkovni model pa na drugi strani zahteva časovno zamudno in delovno zahtevno izvorno zajemanje [16].

Funkcija	Rastrski podatk. model	Vektorski podatkovni model
zajemanje podatkov	zelo hitro in enostavno	večinoma zelo zamudno
količina podatkov	zelo velika	majhna in zmerna
grafična obdelava	povprečna	odlična
podatkovna struktura	enostavna	zelo zapletena in zahtevna
geometrična natančnost	slaba (resolucija)	teoretično neomejena
mrežne analize	zelo slabe	zelo dobre
površinske analize	zelo dobre	povprečne, a precej težavne
generalizacija	enostavna	zapletena in zahtevna

Slika 3: Primerjava funkcij rastrskega in vektorskega podatkovnega modela.

1.1.3 Podatkovna baza

Prostorske podatke shranjujemo v prostorski podatkovni bazi. Ta vsebuje entitete, ki so vezane na geometrijo prostora preko atributov, ki predstavljajo geografske koordinate. Prednosti prostorske podatkovne baze v primerjavi s konvencionalnimi podatkovnimi bazami so naslednje [17]:

- grafično vnašanje podatkov (vektorska in rastrska grafična digitalizacija),
- asociacija entitet preko skupne lastnosti – geografske pozicije,
- geografske analize,
- grafični prikaz podatkov in njihovih medsebojnih odvisnosti.

Starejši ESRI produkti so uporabljali podatke v ESRI shapefile podatkovnem sloju. Gre za digitalni vektorski format za shranjevanje prostorske lokacije objektov ter drugih atributov. Danes ArcGIS programska oprema temelji na ArcSDE (*angl. Spatial Database Engine*) prostorski bazi podatkov, ki za shranjevanje prostorskih podatkov uporablja objektno relacijsko podatkovno bazo [8].

Prostorska baza podatkov vsebuje razrede objektov (*angl. feature classes*) s prostorskimi atributi. Vsebuje tudi topološke informacije. V shapefile podatkovnem sloju je možno shraniti samo en tip objektov, medtem ko je v prostorski bazi podatkov možno shraniti skupine objektov, ponazorjene s točkami, linijami ali poligoni.

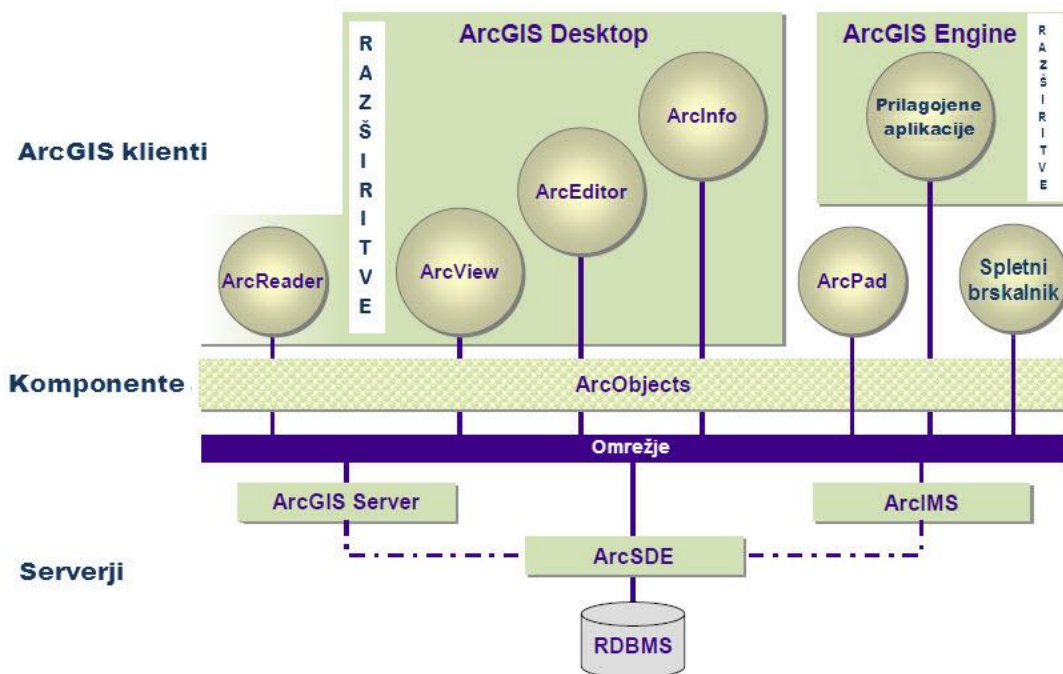
Prostorska baza podatkov v ArcGIS-u je lahko datotečna, osebna ali ArcSDE prostorska baza podatkov [8]. *Datotečna prostorska baza podatkov* shranjuje zbirko različnih prostorskih podatkov v mapi z .gdb končnico. Vsak dataset oziroma skupek razredov objektov je shranjen kot posamezna datoteka na disku, katere velikost je omejena na 1 TB. Datotečna prostorska baza podatkov je mapa, ki shranjuje poljubno število datotek z dataseti. *Osebna prostorska baza podatkov* shranjuje vse datasete znotraj Microsoft Access podatkovne datoteke (.mdb), ki je omejena z velikostjo 2 GB. *ArcSDE podatkovna baza* shranjuje zbirko različnih tipov datasetov kot tabele znotraj objektno relacijske podatkovne baze. Lahko je postavljena na Oraclu, Microsoft SQL Serverju, IBM DB2 ali IBM Informix.

Datotečna ter osebna prostorska podatkovna baza omogočata editiranje le enemu uporabniku naenkrat in ne podpirata verzioniranja. V datotečni podatkovni bazi je editiranje omogočeno večjemu številu uporabnikov, vendar le v primeru, ko uporabniki editirajo vsak svoj dataset.

V primeru, ko potrebujemo veliko prostorsko podatkovno bazo, ki bo hkrati na voljo večjemu številu uporabnikov, se odločimo za ArcSDE. ArcSDE zagotavlja okolje, v katerem je možno sočasno editiranje brez kreiranja množice kopij podatkov. To okolje editiranja se imenuje verzioniranje.

1.2 Programska oprema ArcGIS za delo s prostorskimi podatki

Za analizo in vizualizacijo prostorskih podatkov v diplomski nalogi sem uporabila produkt ArcMap, ki je del ArcGIS-a. ArcGIS je integrirana zbirka programske opreme za GIS tehnologijo. Zagotavlja standardo platformo za prostorske analize, upravljanje s podatki in kartiranje. Gre za produkt podjetja ESRI, ki je vodilni svetovni proizvajalec programske opreme na področju GIS tehnologije. ArcGIS je namenjen reševanju kompleksnih vprašanj na različnih nivojih [9].



Slika 4: ArcGIS programska oprema.

ArcGIS sam po sebi ni GIS aplikacija, pač pa je nabor programske opreme za izgradnjo ArcGIS sistemov. Temelji na skupni knjižnici GIS komponent programske opreme, ki se imenuje ArcObjects. ArcGIS sestavljajo odjemalske in strežniške aplikacije. Vsaka aplikacija lahko kreira, upravlja, analizira in nudi podatke, shranjene v enem ali več formatih.

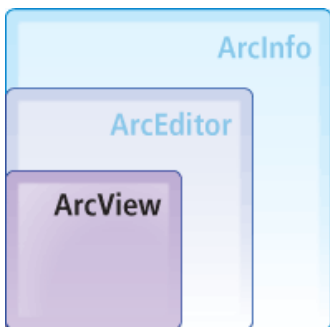
1.2.1 ArcGIS produkti

ArcGIS produkti se izvajajo v namiznem, strežniškem in mobilnem okolju. ArcGIS vsebuje tudi orodja za razvoj lastnih aplikacij.

1.2.1.1 ArcGIS Desktop

ArcGIS Desktop izdelki uporabnikom omogočajo urejanju, obdelavo, prikaz in objavo geografskih podatkov [9]. Trenutna ESRI-jeva namizna zbirka GIS izdelkov je različica 10, ki je bila objavljena junija 2010. Integrirana zbirka ArcGIS Desktop aplikacij je sestavljena iz treh programskih izdelkov, ki se razlikuje glede na licenčne nivoje:

- ArcView,
- ArcEditor in
- ArcInfo.



Slika 5: Družina produktov ArcGIS Desktop.

ArcView nudi robustno GIS funkcionalnost za večino namenov uporabe. ArcEditor ima razširjeno funkcionalnost za urejanje in obdelavo podatkov. ArcInfo predstavlja najvišji nivo licenciranja in ponuja poln nabor funkcionalnosti ESRI-jevega namiznega GIS-a, poudarek je predvsem na geostatistični obdelavi in podpori napredni topološki obdelavi podatkov.

Na vseh treh nivojih licenciranja so dosegljivi naslednji vmesniki: ArcMap, ArcCatalog in ArcToolbox. Dodatek ArcGIS 3D Analyst vsebuje še ArcScene in ArcGlobe vmesnika.

ArcGIS Explorer, ArcReader in ArcExplorer so osnovni in brezplačni izdelki za pregledovanje GIS podatkov. Dodatki kot so Spatial Analyst, 3D Analyst in veliko drugih, pa so namenjeni naprednejšim uporabnikom.

Namizni izdelek ArcGIS Desktop trenutno deluje samo v Windows okolju.

1.2.1.2 GIS spletni servisi

Vključitev GIS tehnologije na splet omogoča množični dostop do prostorskih podatkov brez dodatne programske opreme. Naglo širjenje prodora GIS tehnologije na splet dokazuje, da široka dostopnost, porazdeljenost podatkov in obdelav, enostaven prenos podatkov in standardne dostopne strategije pretehtajo poznane slabosti in pomanjkljivosti spleta.

Strežniški GIS izdelki dovoljujejo GIS funkcionalnost v spletnem okolju. Internet in spletne tehnologije omogočajo napredne prikaze prostorskih podatkov preko spletnih brskalnikov. Izmenjava podatkov med strežnikom in odjemalcem poteka preko http protokola. Prenos obdelanih podatkov na strežniku se izvede v rastrski, tabelarični ali XML obliki [10].

Odjemalčev brskalnik mora biti dopolnjen z dodatnimi možnostmi, kar mu omogoča sporazumevanje z GIS strežnikom ter nudi možnost manipulacije z geografskimi podatki in kartografskimi prikazi. Do GIS aplikacij na spletu je možno pristopati na sledeče načine [10]:

- z uporabo posebnih brskalnikov,
- preko nadgradnje obstoječih brskalnikov za kartografska poizvedovanja,
- z uporabo navideznega javanskega stroja (gre za javansko okolje na strežniku v povezavi z brskalnikom, ki omogoča izvajanje javanskega programa).

S pojavom spletnih GIS strežnikov je uporaba prostorskih informacij postala enostavnejša, javna, splošna in posledično tudi masovna. Prodor GIS tehnologije in uporaba kartografskih prikazov na spletu je v porastu in predstavlja novo področje uporabe.

ArcIMS (*angl. Internet Mapping Server*) ponuja pregledovanje podatkov preko spleta. Izdelek ArcSDE je uporabljen in nameščen nad relacijsko podatkovno bazo RDBMS, ki je optimiziran za delo z GIS podatki. Trenutno ArcSDE podpirajo Oracle, DB2, Informix, Microsoft SQL Server in PostgreSQL podatkovne baze. Izdelek ArcGIS Server je servisno orientirana spletna aplikacija, ki razširja funkcionalnosti namiznega ArcGIS Desktop v spletno okolje. Izdelek ArcGIS Server je dosegljiv na operacijskih sistemih Windows in Linux, kot tudi na Solaris [9].

1.2.1.2.1 Mobilni GIS

Mobilni GIS se nanaša na uporabo geografskih podatkov na mobilnih napravah. Združuje tri bistvene sestavine: GPS (*angl. Global Positioning System*), robustni ročni računalnik in GIS programsko opremo. Zagotavlja tehnološko podporo pri zajemanju prostorskih podatkov v realnem času neposredno na terenu. Uporabnik lahko pri tej tehnologiji vzame potrebne digitalne prostorske in opisne podatke s seboj na teren, jih tam uporablja in posodablja. Internet v povezavi z mobilno telefonijo omogoča brezžične omrežne povezave s strežnikom v pisarni. Takšen pristop pomeni pretvorbo tradicionalnih pasivnih meritev na terenu v aktivno obdelavo in vzdrževanje podatkov med samim zajemanjem [11].

ArcPad omogoča zajemanje, urejanje, analiziranje in predstavitev geografskih informacij na terenu. Ostali produkti so še Mobile ArcGIS Desktop Systems, ArcGIS Server (strežniško usmerjeni API) in ArcWeb Services (spletno usmerjeni API). ArcGIS mobile ima tudi svoj ADF (*angl. Application Development Framework*) za razvoj aplikacij na mobilnih Windows platformah za različne naprave (Pocket PC, smart phones) [11].

1.2.1.2.2 GIS za razvijalce

Ta orodja nudijo programski vmesnik oziroma API (*angl. Application Programming Interface*) in programsko opremo za razvoj inovativnih namiznih, strežniških, mobilnih in spletnih GIS rešitev [12].

Produkti so ESRI Developer Network ali EDN, ArcEngine (namizno usmerjeni API), ArcGIS Server (strežniško usmerjeni API in spletno usmerjeni ADF, ki je del ArcGIS Serverja) in ArcWeb Services (spletno orientiran API) [12].

1.3 Statistični pristopi v analizi prostorskih podatkov

Statistični modeli se uporabljajo za ocenjevanje, opisovanje in napovedovanje na podlagi verjetnostne teorije [4]. Osnovna statistična predpostavka je, da so podatki v vzorcih neodvisni. Vendar je prepostavka o neodvisnosti vzorcev v splošnem napačna, saj prostorski podatki težijo k visoki avtokorelaciji. Naprimer ljudje s podobnimi karakteristikami, poklicem ter poreklom živijo v isti soseski. Tudi spremembe naravnih nahajališč, temperatur in ekonomskih dejavnosti regij se spreminjajo postopno preko prostora. Lastnost grupiranja podobnosti v prostoru je tako temeljna, da je oblikovan prvi zakon geografije: vse je med sabo povezano, vendar so bližnje stvari povezane v večji meri kot oddaljene stvari [15]. Na področju prostorske statistike se ta lastnost imenuje prostorska avtokorelacija.

Za iskanje zakonitosti v prostorskih podatkih so v uporabi različne tehnike:

- *Klasifikacijske metode*
Podatki so preslikani kategorije, ki so vnaprej definirane.
- *Klasterska analiza*
Klasterska analiza omogoča združevanje podatkov v gruče oziroma skupine glede na podobnosti v določenih značilnostih. Rezultat analize so podatki, urejeni v razrede, kjer so razlike med objekti znotraj razredov manjše od razlik med objekti med razredi. Metoda omogoča oblikovanje klastrov na enostaven način, brez predhodnega znanja.
- *Iskanje asociativnih pravil*
Tehnike temeljijo na karakteristikah, ki povezujejo enega ali več objektov z drugimi objekti. Poznamo več vrst prostorskih predikatov, ki tvorijo asociativno pravilo za prostorske podatke kot naprimer podatek o razdalji do najbližjega sosedu, topološke relacije (sečišča, prekrivanja) ali prostorska orientacija (severno od, desno od).
- *Agregacijske in aproksimacijske metode*
Te metode uporabljajo topološke ali geometrične prostorske relacije za kreiranje novih kompleksnih objektov ali za iskanje vzorcev in struktur v prostorski bazi. Geometrija kompleksnega objekta je opisana s primerno aproksimacijo agregiranih geometrij. Primer ponazarja naslednji opis: *80% hiš v tej skupini ima zelo položno streho.*
- *Analiza odklonov*
Metoda išče odklone od pričakovanih vrednosti, izjeme.
- *Sumarizacija*

Tehnike sumarizacije povzamejo obstoječe podatke in kot rezultat nudijo njihov povzetek. Te metode vključujejo vizualizacijo podatkov, statistične funkcije in generalizirana pravila.

- *Metode napovedovanja*

Te metode se koristijo v linearnih in nelinearnih regresijskih modelih.

Mnogi pristopi v GIS in digitalni kartografiji koristijo principe združevanja, naprimer generalizacija v modelu in podatkovni bazi, ter prezentirajo prehod iz nivoja posameznega podatka na nivo skupin podatkov s pomočjo uporabe določenih pravil. Pri tem je ključnega pomena opredelitev pravil, ki klasificirajo podatke v skupine [15].

1.3.1 Razširitev Geostatistical Analyst

ArcGIS Geostatistical Analyst je razširitev aplikacije ArcMap, ki vključuje orodja za raziskovanje prostorskih podatkov in generiranje površin iz podatkov meritev, opravljenih v območjih, kjer zbiranje podatkov za posamezno lokacijo ni možno ali pa so stroški previsoki [15].

Omogoča statistično analizo geografskih podatkov, ugotavljanje nepravilnosti in napovedovanje dogodkov, ki so vezani na zbrane podatke. Z ekstenzijo lahko pregledujemo spremenljivost podatkovnega niza, ugotovimo robne pogoje in odnose med različnimi podatkovnimi nizi, izdelujemo napovedi, standardne napake, verjetnosti doseganja mejnega praga ter geostatistične karte.

Zagotavlja stroškovno učinkovito, logično rešitev za analizo različnih prostorskih podatkovnih baz, ki bi sicer bila časovno in stroškovno potratna. Razširitev polni vrzel med področjem geostatistike in področjem GIS analize. Omogoča:

- vizualizacijo, modeliranje in napovedovanje prostorskih odnosov,
- dinamično povezovanje podatkov, grafov in kart,
- izvajanje deterministične in geostatistične interpolacije,
- evaluacijo modelov in verjetnostno napovedovanje za oceno tveganja.

ArcGIS Geostatistical Analyst omogoča celostno raziskovanje vzorčnih podatkov, ocenjevanje negotovosti, ustvarjanje edinstvenega vpogleda in ustvarjanje prilagojenih interpoliranih površin ter s tem nudi podlago za bolj utemeljeno odločanje. Je v pomoč pri reševanju spornih vprašanj s področja podnebja, okolja, kmetijstva in živalstva [15].

Generiranje kontinuiranega površja za predstavitev določenega atributa je osnovna zmogljivost GIS aplikacij. Glavni izziv je ustvariti najbolj natančno možno površino iz obstoječih podatkov ter opredeliti napako in variabilnost napovedane površine. ArcGIS Geostatistical Analyst uporablja vzorec točk na različnih lokacijah v pokrajini in ustvari oziroma interpolira zvezno površino. Vrednosti na posameznih lokacijah služijo za napovedovanje vrednosti za vsako mesto v pokrajini. Vzorčenja so meritve nekaterih pojavov kot naprimer sevanje jedrske elektrarne, zvišanje vodne gladine, ipd.

ArcGIS Geostatistical Analyst nudi orodja za raziskovalno analizo prostorskih podatkov, ki so v pomoč pri predstavitvi in analizi podatkov s statističnimi metodami. Sestavlja ga vrsta orodij, ki omogočajo vpogled v podatke. Posamezen pogled je povezan z vsemi drugimi pogledi ter z ArcMap aplikacijo. Z vsakim pogledom je mogoče upravljati, ga raziskovati in omogočiti različne vpoglede v podatki. Preverimo lahko porazdelitev podatkov, globalne trende, prostorske avtokorelacije in kovariacije med več nizi.

Razširitev temelji na dveh vrstah interpolacijskih tehnik [15], determinističnih in geostatističnih. Obe skupini kot osnovo delovanja izkoriščata podobnost najbližjih točk. Deterministične tehnike za interpolacijo uporabljajo matematične formule, medtem ko geostatistične tehnike zraven matematičnih vključujejo še statistične pristope. Deterministične interpolacijske tehnike določajo površje iz izmerjenih točk na podlagi podobnosti (npr. inverzna tehtana razdalja) ali stopnje glajenja (npr. radialne osnovne funkcije). Geostatistične interpolacijske tehnike temeljijo na statističnih pristopih in se uporabljajo za bolj napredno napovedovanje modeliranja površja, ki vključuje tudi napake ali negotovosti napovedi.

Kartam z napovedmi negotovosti Geostatistical Analyst zagotavlja orodja za validacijo, ki omogočajo evaluacijo modela in napovedi. Orodja ovrednotijo natančnost modela, tako da lahko model in njegove parametre potrdimo ali pa spremenimo ter ustvarimo boljšo površino. Obenem pa lahko na podlagi primerjave modelov izberemo optimalni model ter površje.

1.4 Samostojen razvoj razširitev

Za razvoj inovativnih ArcGIS rešitev je na voljo razvojno okolje ArcObjects. Razvijalcem omogoča razviti razširitve osnovnih funkcionalnosti ArcGIS produktov. Razvojno okolje ArcObjects je delo razvojne ekipe ESRI [14]. ArcObjects knjižnice zagotavljajo bogat nabor funkcionalnosti, ki jih lahko razvijalci uporabijo za izgradnjo močne GIS aplikacije. Omogočajo razvoj razširjenih aplikacij in rešitev, ki presegajo ArcGIS funkcionalnosti. ArcObjects okolje se uporablja za gradnjo funkcionalnosti namiznih, strežniških, mobilnih in spletnih GIS rešitev.

Veliko uporabnikov, ki nimajo programerskih izkušenj, želi imeti orodja za reševanje težav, s katerimi se srečujejo v ArcGIS produktih. Ta orodja so enostavna za uporabo, imajo enostaven grafični vmesnik, skriptne objekte in programerska orodja, primerna za prilagajanje. V ta namen ESRI podpira različne skriptne jezike, ki uporabljajo ArcObjects. Uporabniki s programerskim znanjem pa lahko prilagodijo funkcionalnosti ArcGIS orodij s pomočjo različnih programskih jezikov: Visual Basic for Applications (VBA), C, C#, C++, Java...

Eden izmed skriptnih jezikov za enostavno prilagajanje je Python, ki je nameščen v ArcGIS 9 in naprej vključno z drugimi komponentami. Python odlikujeta prilagodljivost različnim platformam ter enostavna integracija z drugimi programskimi jeziki kot so C, C++, FORTRAN in Java. Vsebuje mnoge knjižnice in orodja, ki so že razviti za delo z GIS podatki. Python podpira manipulacijo s formatom ESRI shapefile, geografskimi mrežami ter slikami, kot tudi skriptni ArcSDE in interakcijo s spletnimi storitvami in podatkovnimi bazami.

ArcObject so gradniki aplikacije ArcMap ter drugih namiznih aplikacij kot so ArcCatalog, ArcScene, ArcGlobe. Ti osnovni gradniki so GIS objekti, do katerih je možno dostopati preko kode in jih implementirati v lastnih razredih. Uporabniki aplikacij in razvijalci uporabljajo enake objekte. Objektni razredi oziroma komponente ponazarjajo osnovne GIS objekte:

- Map
- Layer
- Point
- Line
- Polygon
- Table
- Row
- Field

Koraki pri programiranju enostavnega orodja v ArcMap-u (naprimer za klik na gumb, ki izvede neko akcijo na karti) so sledeči:

- Najdemo podobno ArcGIS orodje.
Za funkcionalnost, kjer kliknemo na gumb, bi to bil naprimer razred Button.
- Raziščemo, katere vmesnike razred implementira in jih implementiramo tudi v naši komponenti.
V našem primeru Button implementira ICommand, s čemer pravzaprav implementiramo lastnosti in metode ICommand vmesnika, ki jih sami sprogramiramo.

ArcObjects objekti so COM komponente (Component object model). Gre za objekte, ki so razviti v skladu s standardi, ki omogočajo prenosljivost med različnimi programskimi jeziki. COM razredi lahko preko vmesnikov komunicirajo z drugimi programskimi jeziki. ArcObjects razredi so napisani v C++. Vmesniki omogočajo njihovo rabo v C++, VB.NET, C#.NET, VBA, Python-u ter nekaterih drugih programskih jezikih.

Razvoj enostavne COM komponente za ArcMap aplikacijo poteka po korakih:

- kreiranje COM/.NET projekta,
- kreiranje COM razreda,
- referenciranje na ArcGIS knjižnico,
- implementacija ArcGIS vmesnikov,
- registracija v ArcGIS kategoriji komponent.

1.5 Podatkovno rudarjenje na prostorskih podatkih

Podatkovno rudarjenje oziroma odkrivanje zakonitosti [15] v podatkih lahko definiramo kot odkrivanje novega znanja v bazah z veliko podatki. Podatkovno rudarjenje v prostorskih bazah je prenos principov podatkovnega rudarjenja s področja relacijskih in transakcijskih podatkovnih baz na področje prostorskih podatkovnih baz. Danes imamo v različnih aplikacijah na voljo ogromno prostorskih podatkov. Možnost, da bi te prostorske podatke analizirali ročno, je neizvedljiva. Ravno zato je avtomatizacija odkrivanja znanja v podatkih izjemno pomembna.

Podatkovno rudarjenje v prostorskih bazah povezuje mnoga področja, od strojnega učenja, shranjevanja podatkov, vizualizacije podatkov, statistike, informacijske teorije, pa do računalniške geometrije. Temelje podatkovnega rudarjenja v prostorskih bazah predstavljata statistika in podatkovno rudarjenje [7]. Tehnike prostorskega podatkovnega rudarjenja se lahko uporabljajo za razumevanje prostorskih podatkov, z njimi odkrivamo povezave med prostorskimi in drugimi neprostorskimi podatki, ponazorimo lahko porazdelitev in vzorce prostorskih podatkov nekaterih pojavov ter napovemo trend oblikovanja teh vzorcev, optimiziramo lahko poizvedovanje na bazi, ipd.

Naloge odkrivanja zakonitosti v prostorskih podatkih lahko razdelimo na naloge opisovanja, raziskovanja in napovedovanja. Za razumevanje je potrebno prostorske podatke in pojave najprej opisati in analizirati ter raziskati vse skrite vzorce in povezave med prostorskimi in neprostorskimi podatki. Na osnovi vzorcev prostorskih porazdelitev in razumevanja prostorskih povezav lahko napovemo trende prostorskih vzorcev and porazdelitev.

Odkrivanje zakonitosti v prostorskih podatkih vključuje (vendar ni omejeno le na) vizualno interpretacijo in analizo, prostorsko poizvedovanje in selekcijo atributov, generalizacijo in klasifikacijo, detekcijo prostorskih in neprostorskih asociativnih pravil in klustersko analizo.

Podatki so kompleksnejši kot pri podatkovnem rudarjenju, saj vključujejo razširjene objekte: točke, linijah in poligone. Objekti imajo dva različna tipa atributov, neprostorske ter prostorske attribute. Neprostorski atributi opisujejo neprostorske značilnosti objekta kot so recimo ime, populacija in stopnja nezaposlenosti v izbranem mestu. Enaki so atributom, ki se uporabljajo v klasičnem podatkovnem rudarjenju. Prostorski atributi opredeljujejo prostorsko lokacijo ter obseg prostorskih. Prostorski atributi prostorskih objektov najpogosteje vsebujejo informacije, ki so vezane na lokacijo: zemljepisna dolžina in širina, nadmorska višina, ipd.

Razmerja med neprostorskimi podatki so eksplicitna (je član skupine, je podrazred, je nadrazred). Nasprotno pa so pri prostorskih objektih razmerja pogosto implicitna (prekrivanje, sečišče).

1.5.1 Druge razširitve v ArcMap-u

ArcMap vsebuje več razširitev, ki nudijo različne funkcionalnosti [8].

3D Analyst

Omogoča učinkovito predstavitev in analizo podatkov v 3D okolju. Razširitev omogoča kreiranje površine, kreiranje 3D objektov, konvertiranje 2D objektov v 3D objekte, 3D analize in geoprociranje [8].

Spatial Analyst

Razširitev ponuja več funkcionalnosti [8]:

- Pridobivanje novih informacij iz obstoječih podatkov.
Omogoča izračunavanje razdalj med objekti, razvrščanje obstoječih podatkov v razrede, kreiranje reliefa, ipd.
- Iskanje primernih lokacij za določene namene.

Naprimero analiziranje lokacij za gradnjo novih stavb z upoštevanjem nevarnosti poplav in plazov.

- Identifikacija optimalnih povezav med določenimi lokacijami.
Identificirati je mogoče optimalne poti za gradnjo cest in vodnega omrežja, migracijo živali. Najkrajše poti niso vedno najcenejše, razširitev upošteva ekonomske, ekološke in druge dejavnike.
- Analiza razdalj in stroškov
Izračunamo lahko stroške izbrane poti med dvema lokacijama.

Network Analyst

Razširitev omogoča analizo in načrtovanje poti, izdelavo potovalnih načrtov in upravljanja z voznim parkom. Omogoča delo z dinamičnimi modeli realnih pogojev, ki vplivajo na vozni park, kar vključuje tudi lastnosti cest, omejitve hitrosti, omejitve višine in gostote prometa, vse v odvisnosti od časa, ko izdelujemo analizo. Omogoča napredno usmerjanje in analizo omrežij [8]:

- analizo časa razvoza,
- upravljanje od točke do točke,
- načrt razvozne poti,
- definicijo servisnih območij,
- ugotavljanje najhitrejših in najkrajših poti,
- optimizacijo transporta.

Schematics

Omogoča generiranje, vizualizacijo in manipulacijo diagramov iz omrežja podatkov, ki prihajajo iz prostorske podatkovne baze oziroma kakršnihkoli podatkov, ki kažejo povezljivost.

Ima sledeče funkcionalnosti [8]:

- ustvari načrte iz kompleksnih mrež,
- preveri mrežne povezave,
- izvede nadzor kakovosti podatkovnega omrežja,
- vizualizira neprostorske povezane podatke,
- modelira socialne mreže,
- kreira diagrame.

Tracking Analyst

Razširitev je namenjena mapiranju objektov, ki se premikajo ali spreminjajo svoj status skozi čas. Beleži časovne podatke v realnem času in podpira omrežne povezave za GPS (*iz angl. Global Positioning System*). Pregledamo lahko pretekle podatke, izdelamo časovno zaporedje dogodkov ali podatkov, ugotovimo pojavljanje določenih vzorcev skozi čas, izdelamo diagram in sledimo spremembi podatkov skozi čas [8].

ArcScan

Razširitev zagotavlja orodja in ukaze, ki omogočajo, da skenirane posnetke kart spremenimo v vektorske sloje kot je shapefile podatkovni sloj. Postopkom preoblikovanja rastrskih podatkov v

vektorske pravimo vektorizacija. Vektorizacija se lahko opravi ročno ali na avtomatski način. ArcScan ima tudi orodja, ki omogočajo, da lahko izvedemo preprosto urejanje rasterskih plasti za vektorizacijo [8].

Survey Analyst

Razširitev ponuja zbirko orodij za obdelavo in analizo podatkov iz zemljiškega katastra. Omogoča nenehno izboljševanje natančnosti na podlagi obstoječih metodologij raziskovanja. Zmanjšuje čas, potreben za vzdrževanje katastrskih podatkov in omogoča sledenje zgodovini vseh sprememb [8].

Publisher

Orodje ponuja enostavno izmenjavo in distribucijo GIS zemljevidov in podatkov s komerkoli. ArcGIS Publisher pretvori zemljevide v format, ustrezen za objavljanje. Omogoča [8]:

- prikaz zemljevidov vsem uporabnikom,
- zaščito kart in podatkov pred neustrezno uporabo,
- kreiranje kart, ki ustrezajo potrebam uporabnikov,
- učinkovit in nadzorovan dostop do GIS podatkov,
- gradnjo prilagojenih pogledov za podatke s pomočjo orodja ArcReaderControl.

Na voljo so tudi zelo uporabne orodne vrstice, med katerimi so spodaj opisane le najpomembnejše.

Animation

Kreiranje animacij je omogočeno s spreminjanjem lastnosti slojev, kot je recimo transparentnost. Animacije so lahko zanimive za analizo skozi čas oziroma za predstavitev informacij skozi različne sloje [8].

Annotation

Razširitev omogoča dodajanje opomb in beležk. Tipično gre za dodajanje besedila, lahko pa tudi za dodajanje grafičnih oblik kot so puščice in okvirji [8].

ArcPad

Omogoča kopiranje podatkov za terensko uporabo v ArcPadu ter kopiranje sprememb nazaj v ArcMap [8].

Dimensioning

Orodna vrstica vsebuje vrsto kontrol in orodij, ki pomagajo ustvariti prave dimenzije s standardom urejevalnikom. Ustvarimo lahko horizontalno linearno, navpično linearno ali rotirano linearno razsežnosti kart [8].

Distributed geodatabase

Orodje omogoča podjetju, da razprši svoje podatke iz osrednjih strežnikov na regionalne ali lokalne urade, ki so lahko povezani ali nepovezani v skupno mrežno okolje. To se doseže z v

celotno ali delno lokalno kopijo dela podatkov iz glavne baze. Posodobitve sprememb je mogoče uskladiti med uradi s sinhronizacijo procesa [8].

Draw

Orodje se uporablja za dodajanje, odstranjevanje in spreminjanje grafičnih elementov na karti [8].

Geocoding

Prostorski objekti so v geografskem koordinatnem sistemu referencirani preko X in Y koordinat. Kljub temu pogosto koristimo informacije o krajih, ki so navedeni z naslovi. Preoblikovanje tovrstnih referenc v X in Y koordinati omogoča geokodiranje [8].

Georeferencing

Orodje omogoča georeferenciranje, ki je poravnavanje prostorskih podatkov (točk, poligonov) s slikovnimi kartami [8].

ArcMap vsebuje še druge praktične orodne vrstice, ki omogočajo še veliko funkcionalnosti, kot so orodne vrstice za prezentacijo podatkov, dodajanje napisov, dodajanje grafičnih elementov in efektov, verzioniranje, editiranje elementov v omrežju, ipd. [8].

1.6 Izmenjava podatkov iz ArcMap-a z zunanjimi orodji

ArcGIS Data Interoperability je razširitev v ArcGIS Desktop-u, ki omogoča preprosto uporabo in distribucijo podatkov v mnogo formatov. Uporaba načel ETL (iz angl. extract, transform, load) odpravlja ovire pri izmenjavi podatkov in zagotavlja točne prostorske podatke za uporabnike. Razširitev omogoča uporabo vseh standardnih GIS podatkov v ArcGIS Desktop okolju za kartiranje, vizualizacijo in analizo ne glede na format.

Orodje omogoča branje več kot 85 različnih formatov, vključujoč GML, XML, Autodesk DWG/DXF, MicroStation Design, MapInfo MID/MIF, TAB, Oracle in Oracle Spatial, Intergraph ter izvoz v več kot 50 različnih prostorskih formatov. Pretvorba med formati je avtomatska.

Po instalaciji ArcGIS Data Interoperability razširitve so tuji formati prepoznani in je z njimi omogočeno neposredno delo. Podatkovne datoteke, ki temeljijo na tujih formatih, postanejo vidne v ArcCatalog-u, kjer je omogočeno dodajati podatke kot sloje v ArcMap, vendar je omogočeno le branje oziroma vidnost teh slojev. Za urejanje tujih formatov je potrebno tuje formate pretvoriti v katerega izmed ArcMap formatov. To nam omogoča Data Interoperability orodna vrstica v ArcCatalogu, ki nudi možnost hitrega izvoza in uvoza podatkov.

2 TESTNI PODATKI

V praktičnem delu naloge sem izvedla analize na različnih vrstah podatkov. Podrobnejši opis podatkov sledi v poglavju Vizualizacija podatkov.

- Podatki o strukturi prebivalstva v Sloveniji

Podatki so shranjeni v Excelovem format (.xls). Vsebujejo informacije o številu žensk in moških v izbranih letih (1990, 1995, 2000, 2005, 2010) v posameznih statističnih regijah v Sloveniji. Podatki so prosto dostopni na spletnem portalu Statističnega urada Republike Slovenije. V nalogi so uporabljeni za ponazoritev, kako je možno prostorskim podatkom pridružiti podatke brez prostorskih atributov. Uporabila sem jih tudi za prikaz možnosti vizualizacije v ArcMap orodju.

- Podatki o statističnih regijah v Sloveniji

Podatki so shranjeni v ESRI shapefile formatu (.shp) (glej poglavje Vizualizacija podatkov). Podatkovni sloj vsebuje podatke o statističnih regijah v Sloveniji. Statistične regije so ena izmed teritorialnih ravni, za katere Statistični urad RS zbira in izkazuje statistične podatki [20x]. Podatke je možno pridobiti na Geodetski upravi Republike Slovenije. V nalogi so namenjeni ponazoritvi enostavnih možnosti vizualizacije z ArcMap-om.

- Podatki o koncentraciji ozona, meji ter mestih v Kaliforniji

Podatki so shranjeni v ESRI shapefile formatu (.shp) (glej poglavje Vizualizacija podatkov). Uporabljeni so trije podatkovni sloji. Podatkovni sloj OzonTocke vsebuje podatke o povprečni koncentraciji ozona na več lokacijah v Kaliforniji. Podatkovni sloj Meja vsebuje podatke o državni meji Kalifornije, podatkovni sloj Mesta pa vsebuje podatki o večjih mestih v Kaliforniji. Podatki so prosto dostopni na spletnih straneh podjetja ESRI. Na podatkih so izvedene statistične analize, ki jih omogoča razširitev Geostatistical Analyst, ter vizualizacija rezultatov analiz.

3 VIZUALIZACIJA PODATKOV

Ena izmed osnovnih funkcionalnosti, ki jih omogoča ArcMap, je vizualizacija podatkov. V prvem delu sem predstavila temeljne možnosti prikaza podatkov na karti. Izhajala sem iz podatkov o statističnih regijah, ki so shranjeni v formatu ESRI shapefile, ter podatkov o strukturi prebivalstva glede na spol in leto v posameznih statističnih regijah, shranjenih v Excelovem dokumentu.

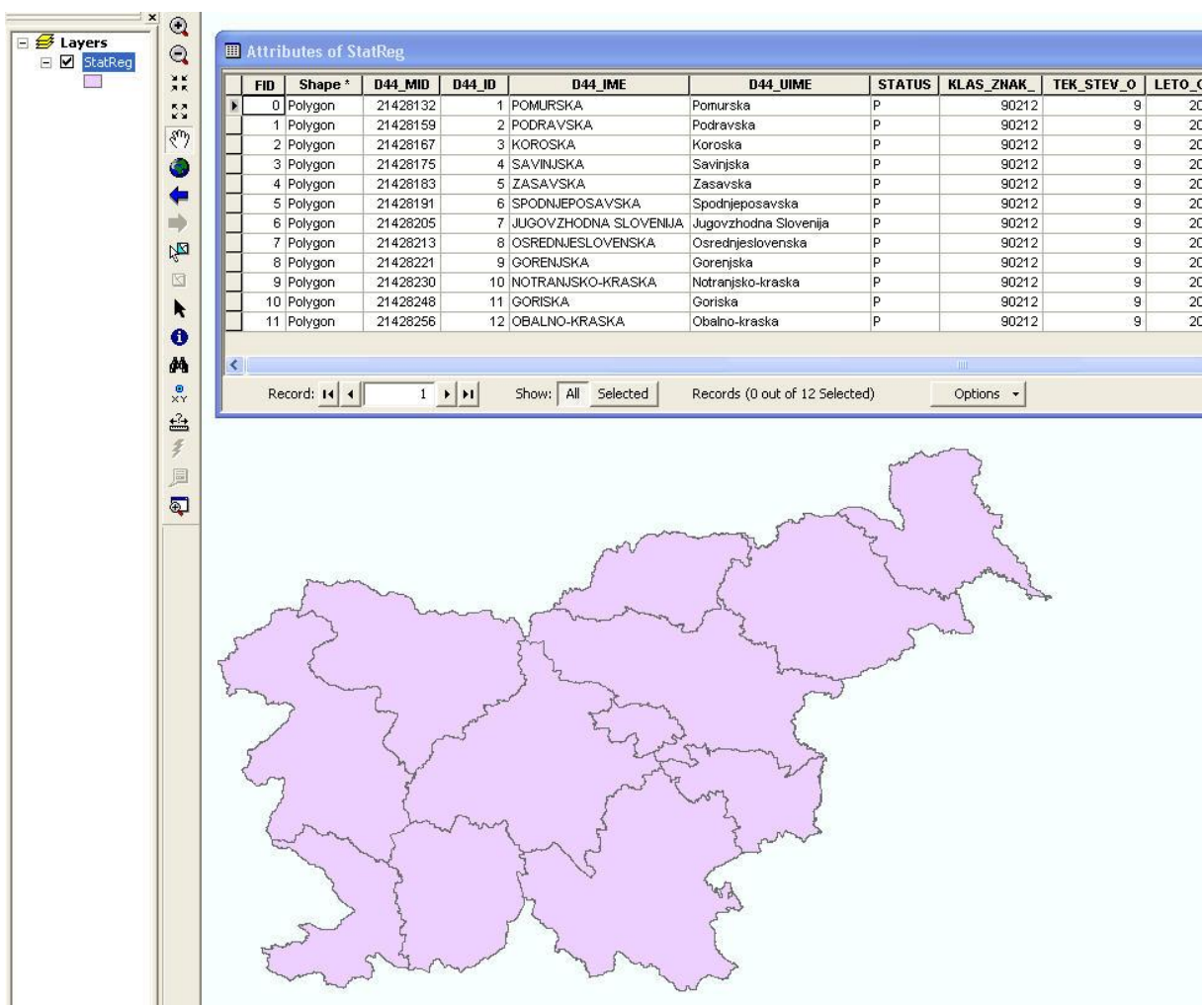
ESRI shapefile podatkovni sloj oziroma krajše shapefile podatkovni sloj je digitalni vektorski format za shranjevanje prostorske lokacije objektov ter drugih atributov. Shapefile podatkovni sloj shranjuje enostavne prostorske objekte kot so točke, poligoni in polilinije. Gre za tabelo, kjer so najpomembnejše informacije o prostorski lokaciji objektov. Te so shranjene v atributu »Shape«.

Točke imajo vrednost »Shape« atributa »Point«, ki vsebuje podatke o zemljepisni širini in dolžini objekta (X in Y koordinata). Kot primer si lahko predstavljamo lokacije vodnjakov kot točkovnih objektov na karti. Vrednost atributa »Shape« za prikaz poligonov je »Polygone«, ki vsebuje nekoliko več informacij (podatki o točkah ter linijah, ki tvorijo poligon). Vrednost »Polygone« bi uporabili za primer prikaza jezer na karti. Pogosta vrednost atributa »Shape« je tudi »Polyline«. Ta vsebuje podobne podatke kot vrednost »Polygone«, le da polilinije, ki jih sestavljajo točke in linije, ne tvorijo zaključene celote. S polilinjami bi naprimer ponazorili reke.

3.1 Priprava podatkov in združevanje prostorskih podatkov s podatki brez prostorskih atributov

Izhajala sem iz podatkov o statističnih regijah, ki so shranjeni v formatu ESRI shapefile, ter podatkov o strukturi prebivalstva glede na spol in leto v posameznih statističnih regijah, shranjenih v Excelovem dokumentu. Najprej si bomo ogledali, kako je možno podatke brez prostorske informacije v ArcMapu vezati na prostorske podatke ter jih prikazati na karti. Sledi pregled enostavnih načinov vizualizacije podatkov.

Podatki o statističnih regijah se nahajajo v shapefile podatkovnem sloju StatReg.shp. Poligonski shapefile podatkovni sloj vsebuje osnovne attribute kot so »FID« (polje je samodejno generirano in izpolnjeno z zaporednim številom vrstice), »Shape« (informacija o vrsti prostorskega objekta in njegovi lokaciji), ime statistične regije (D44_UIME) ter ostale attribute (omenimo še LENGTH, ki nosi informacijo o obsegu poligona ter AREA, ki predstavlja površino poligona).




Slika 6: Poligonski shapefile podatkovni sloj ter pripadajoča tabela atributov v ArcMapu (Vir: GURS, april 2010).

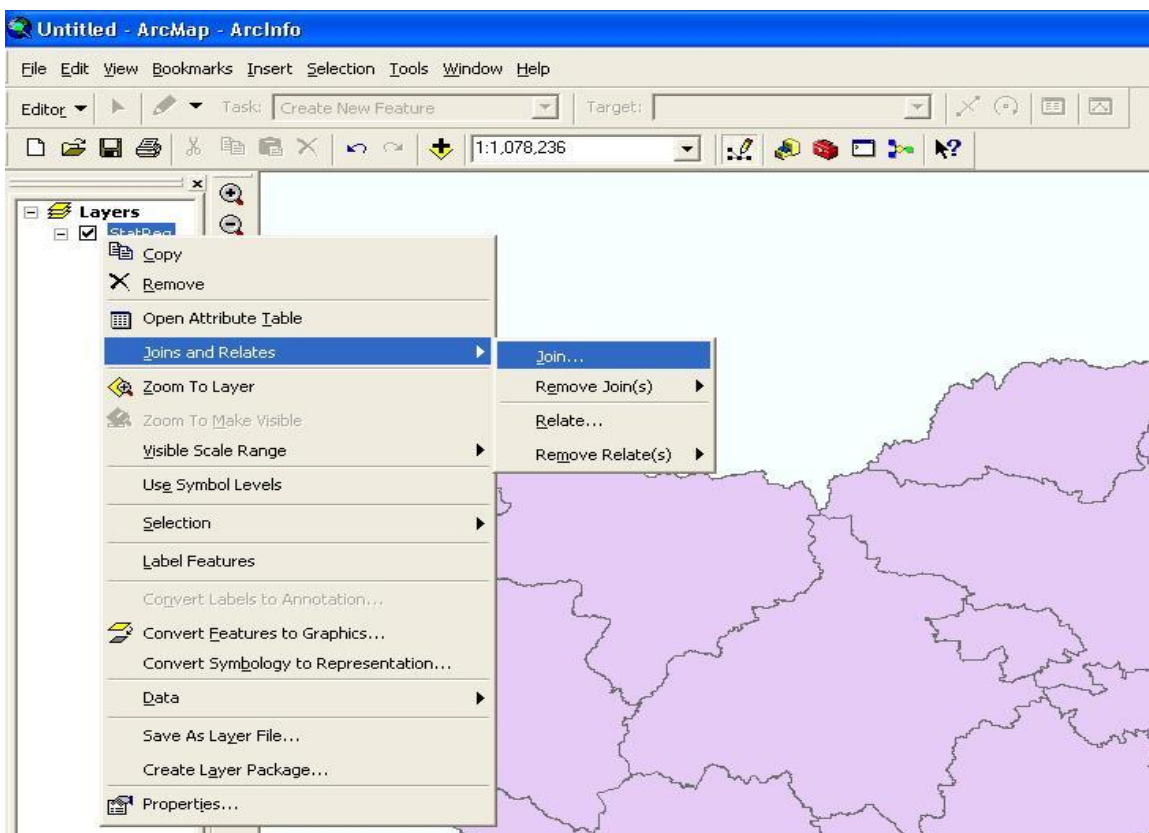
Podatki o prebivalstvu ter spolu v različnih letih po posameznih statističnih regijah se nahajajo v Excelovi datoteki. Tabela vsebuje naslednje attribute: statistične regije Slovenije (atribut ST_REG) ter število moškega prebivalstva v posameznih letih (atributi M_1990, M_1995, M_2000, M_2005 in M_2010) ter število ženskega prebivalstva v posameznih letih (atributi Z_1990, Z_1995, Z_2000, Z_2005 in Z_2010).

	A	B	C	D	E	F	G	H	I	J	K
1	ST_REG	M_1990	M_1995	M_2000	M_2005	M_2010	Z_1990	Z_1995	Z_2000	Z_2005	Z_2010
2	Pomurska	64089	63168	60585	59395	58198	67557	66699	64384	63322	61350
3	Podravska	159079	157860	155533	155656	159622	168770	167276	164072	163458	163721
4	Koroska	36966	36955	36866	36783	36551	37359	37402	37181	37056	36261
5	Savinjska	125110	125341	125679	126111	129967	132481	132247	130652	130641	130058
6	Zasavska	23214	22556	22518	22146	21893	24593	24141	23933	23401	22813
7	Jugovzhodna Slovenija	66440	67100	68152	68844	71612	68927	69454	69774	70251	70480
8	Osrednjeslovenska	230327	229496	235933	240645	259613	251988	252054	251856	256030	270033
9	Gorenjska	93452	94316	95803	96841	100247	99913	100491	100674	101501	102656
10	Notranjsko-kraska	24841	24226	24915	25281	26342	25595	25224	25629	25751	25875
11	Goriska	59057	58404	59438	59010	59588	61867	61424	61006	60612	59492
12	Obalno-kraska	49720	49633	50875	51789	55256	51834	52002	52453	53240	55156
13	Spodnje-posavska	36382	35779	34699	34701	34876	36816	36299	35145	35125	35316

Slika 7: Pripravljeni podatki o prebivalstvu po spolu v različnih letih po posameznih statističnih regijah (Vir: SURS, junij 2010).

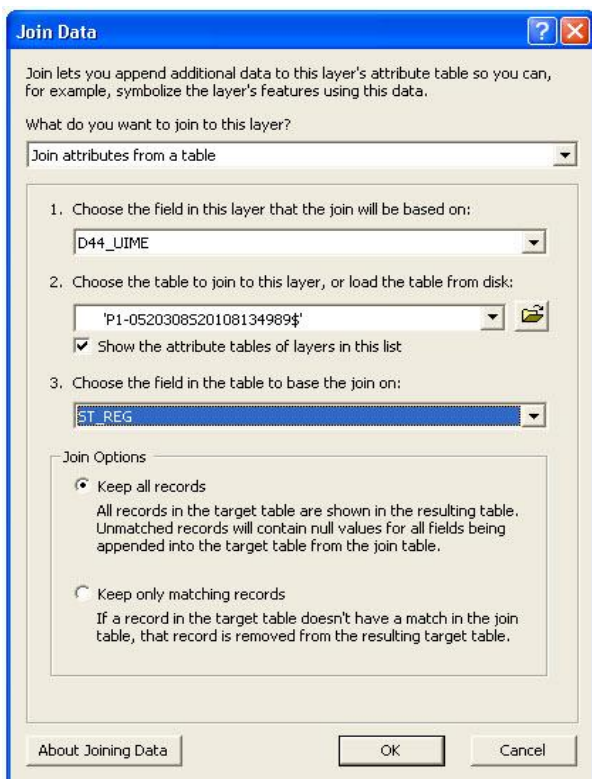
Podatke je potrebno pripraviti tako, da je mogoče izvesti združevanje. Vrednosti atributov, po katerih združujemo podatke, je potrebno razvrstiti v stolpce. V našem primeru so to imena statističnih regij. Urediti je potrebno tudi nazive statističnih regij. Šumnike, presledke ter ločila v imenih regij uredimo tako, da so enaka.

Združevanje tabel izvedemo v ArcMapu. Najprej dodamo shapefile. Kliknemo na ikono za dodajanje podatkov . V raziskovalcu poiščemo lokacijo datoteke StatReg.shp in jo dodamo. Dodani sloj se prikaže v ArcMapu. Dodati je potrebno še attribute o prebivalstvu po spolu in letih. Izvedemo združevanje podatkov sloja Statistične regije ter Excelove tabele preko stika (JOIN). V ArcMapu z desnim klikom na sloj StatReg odpremo možnosti, med katerimi izberemo »Joins and Relates« ter nato opcijo »Join«.



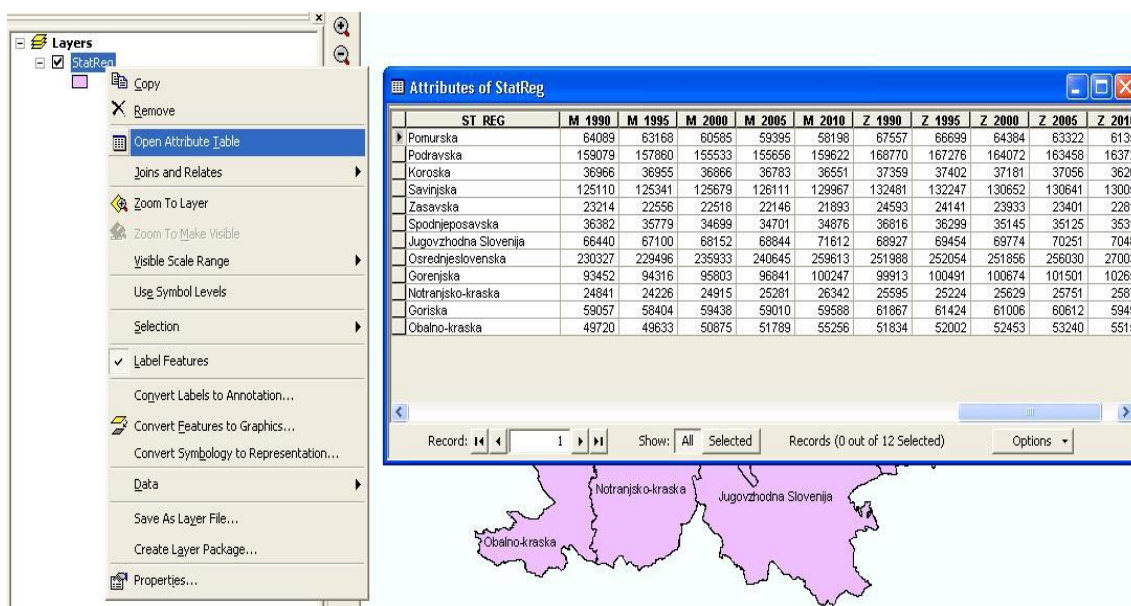
Slika 8: Opcija združevanja tabel v ArcMapu.

Odpre se okno, kjer določimo način združevanja podatkov. Izberemo možnost »Join attributes from table«. Kot atribut združevanja v StatReg.shp izberemo naziv statističnih regij (D44_UIME). Izberemo še lokacijo Excelove datoteke in atribut združevanja, torej naziv statističnih regij (ST_REG). Pustimo izbrano možnost, da naj se v kreirani tabeli obdržijo vsi zapisi iz obeh začetnih tabel ter potrdimo združevanje.



Slika 9: Združevanje tabel na podlagi naziva statističnih regij iz tabele StatReg.shp ter Excelove tabele.

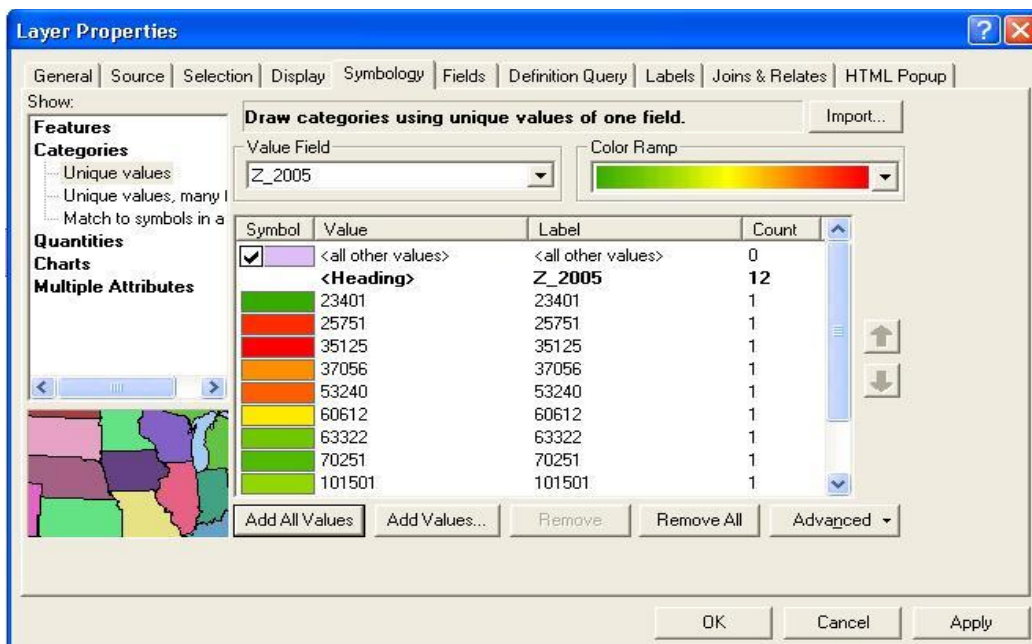
Rezultat združevanja preverimo v tabeli. Z desnim klikom na sloju RegStat izberemo med možnostmi »Open attribute table«, kjer so vidni atributi obeh tabel.



Slika 10: Tabela podatkovnega sloja StatReg po združevanju vsebuje attribute obeh tabel.

3.2 Možnosti vizualizacije

Sedaj lahko med možnostmi sloja izberemo opcijo »Properties«, ki nam odpre okno z lastnostmi sloja. Kliknemo zavihek »Symbology« ter izberemo možnost »Categories«. V polju »Value Field« izberemo atribut, ki ga želimo prikazati na karti (recimo žensko prebivalstvo v letu 2005 po statističnih regijah) ter potrdimo dodajanje vseh vrednosti s klikom na gumb »Add All Values«. V polju »Color Ramp« izberemo barvno lestvico, v kateri želimo prikazati posamezne kategorije podatkov.

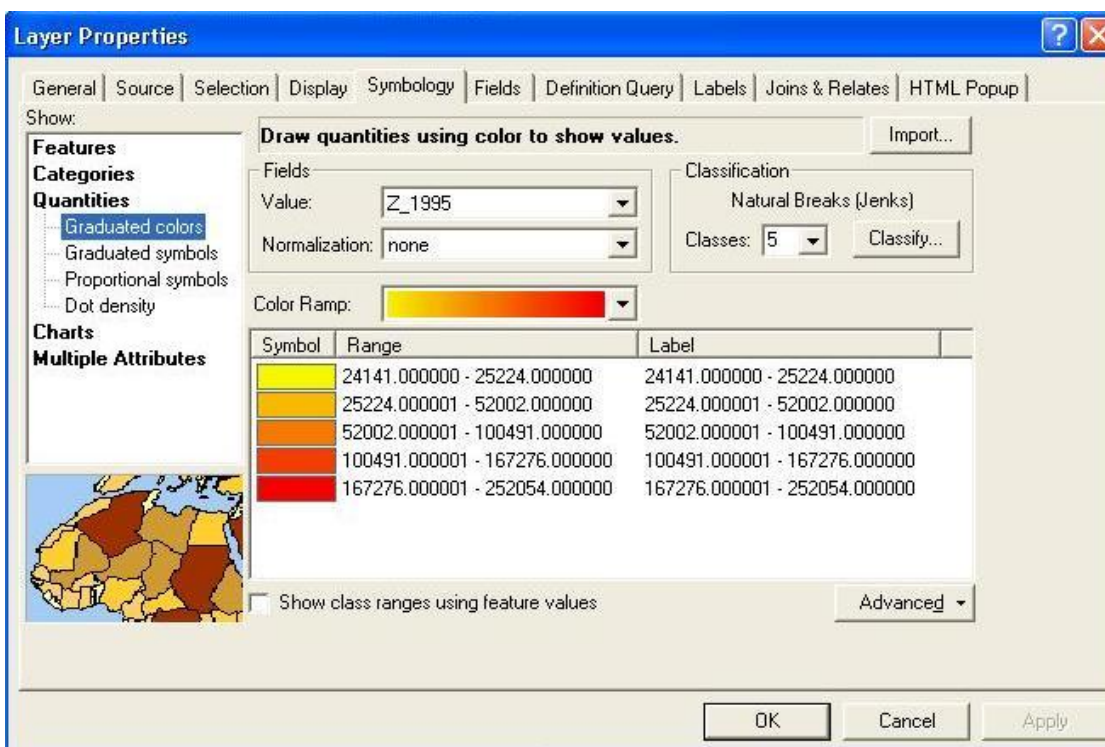


Slika 11: Nastavitve lastnosti sloja StatReg za prikaz kategorij števila žensk v posameznih statističnih regijah.



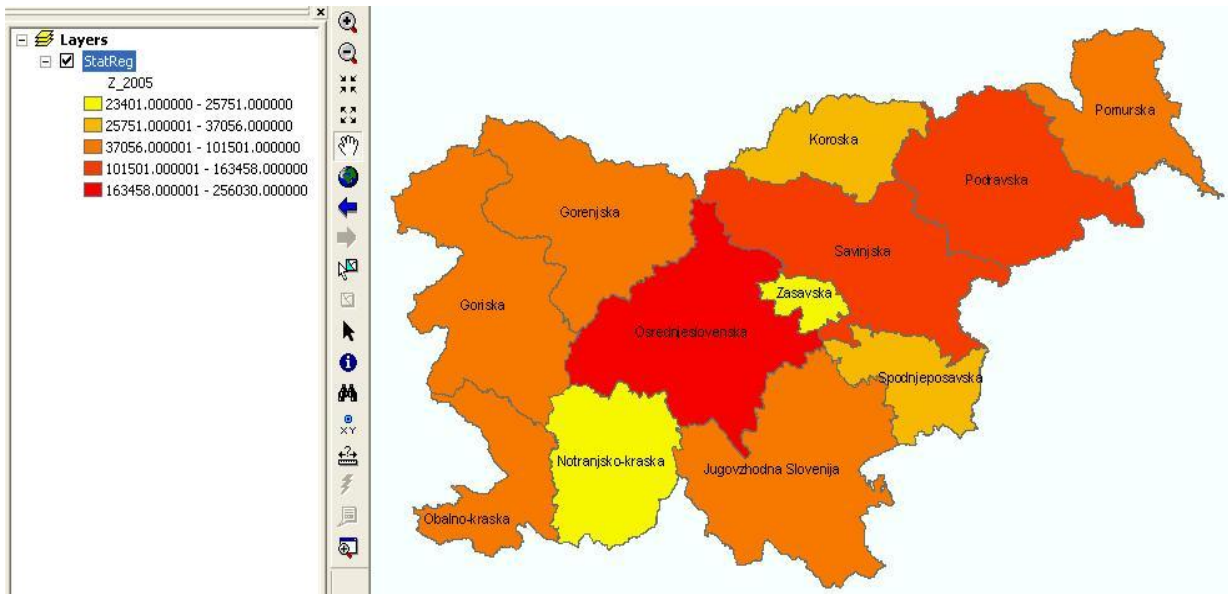
Slika 12: Kategorije števila žensk v posamezni statistični regiji.

Morda prikaz po kategorijah ni najustreznejši, saj je število kategorij relativno veliko (slika 13). Primernejši je prikaz, kjer so podatki urejeni v razrede, katerih število poljubno določimo. V zavihku »Symbology« izberemo možnost »Quantities«. V polju »Value Field« ponovno izberemo število žensk v letu 2005. V polju »Classes« določimo število razredov.



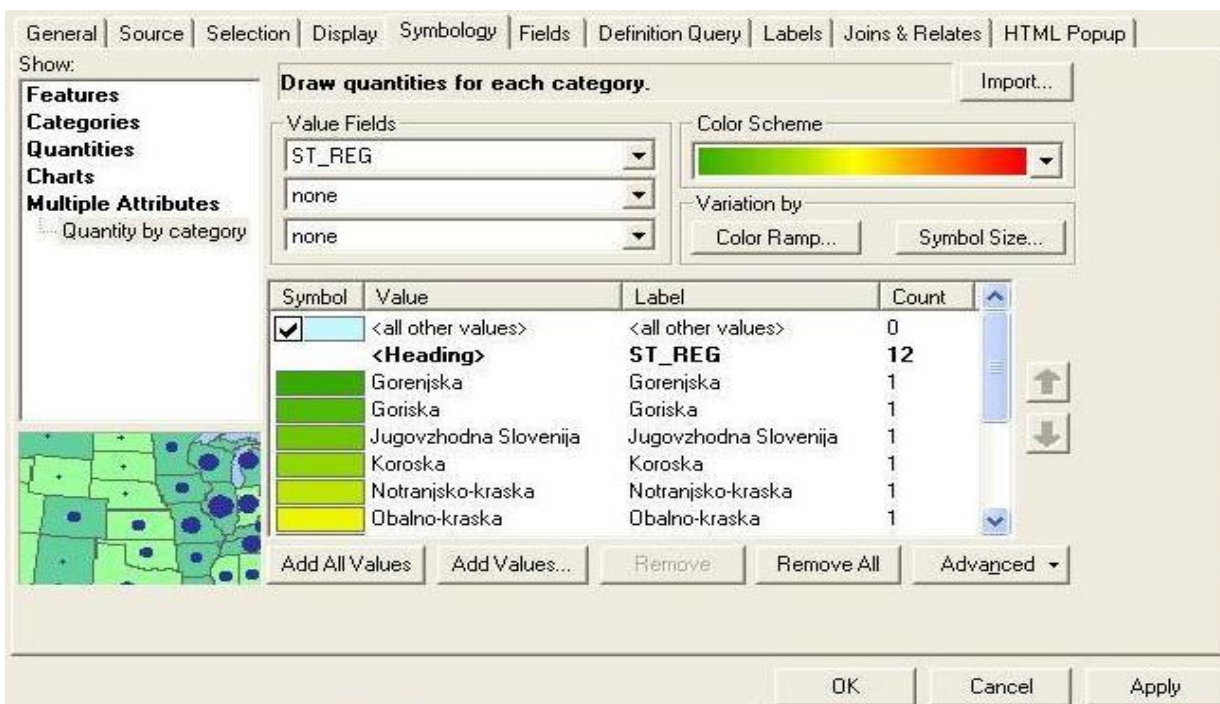
Slika 13: Nastavitve lastnosti sloja StatReg za prikaz ženskega prebivalstva, razvrščenega v razrede.

Vizualizacija daje jasnejši vpogled v podatke. Opazimo lahko, da je bilo število žensk v letu 2005 najmanjše v Zasavski in Notranjsko-kraški statistični regiji. Največ žensk je živelo v Osrednjeslovenski statistični regiji. Podobno število ženskega prebivalstva pa je bilo na Koroškem ter v Spodnjem Posavju (obe regiji pripadata razredu med 37056 ter 101501 prebivalci). Enakemu razredu pripadata tudi recimo Savinjska in Podravska regija.



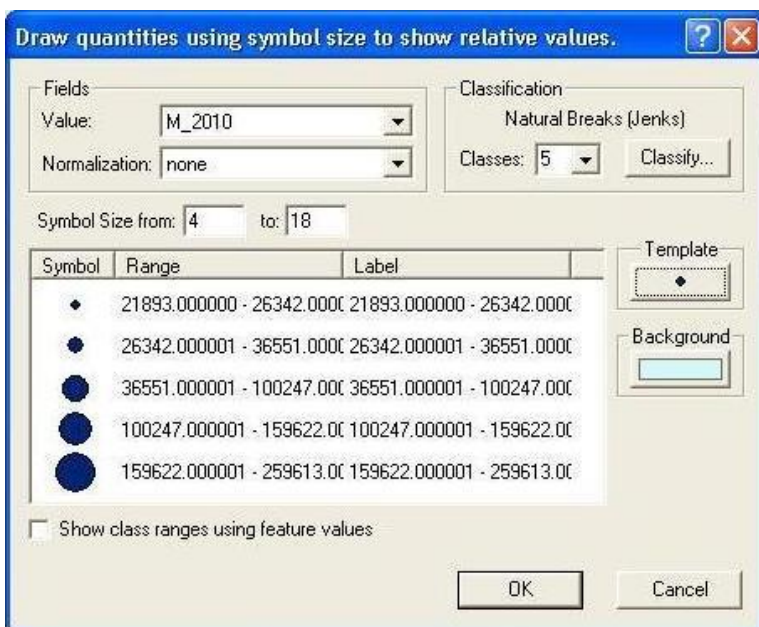
Slika 14: Razredi števila ženskega prebivalstva glede na statistične regije.

Podoben način vizualizacije atributov, ki daje več informacij, ker ponazori vrednost izbranega atributa z velikostjo izbranega lika, omogoča na zavihku »Symbology« možnost »Multiple attributes«.



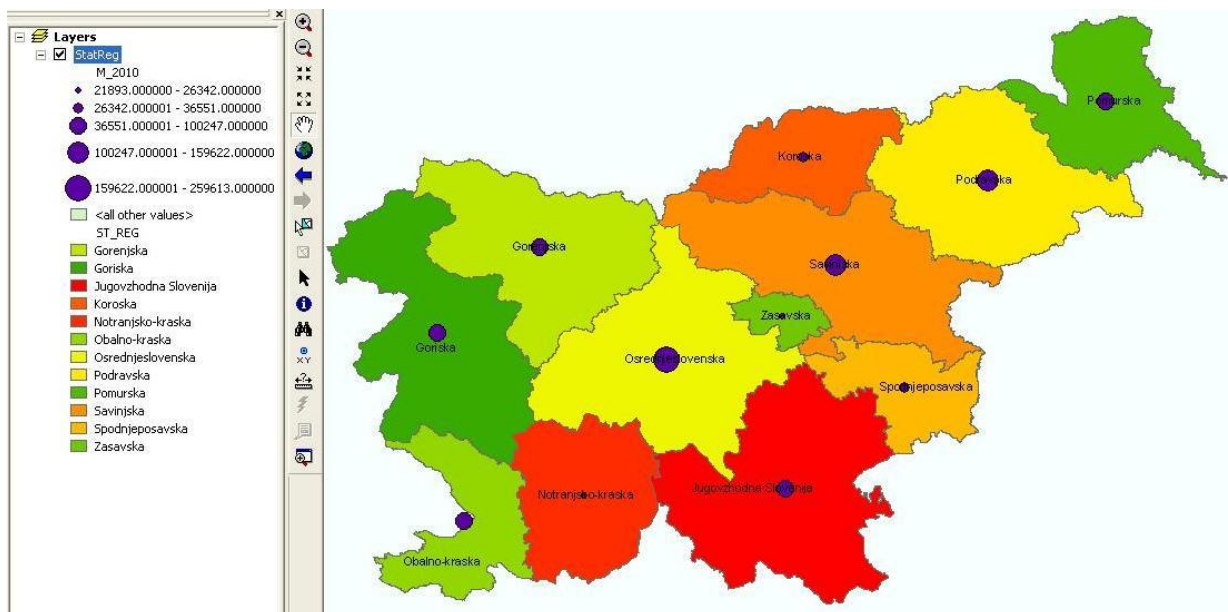
Slika 15: Možnost »Multiple attributes«.

Prikazati želimo število moškega prebivalstva v posameznih regijah v letu 2010, ponazorjeno z velikostjo izbranega lika. V polju »Value Field« izberemo statistične regije, pod možnostjo »Symbol Size« pa določimo atribut, katerega vrednosti nas zanimajo (M_2010) ter število razredov.



Slika 16: Nastavitve za prikaz relativnih vrednosti.

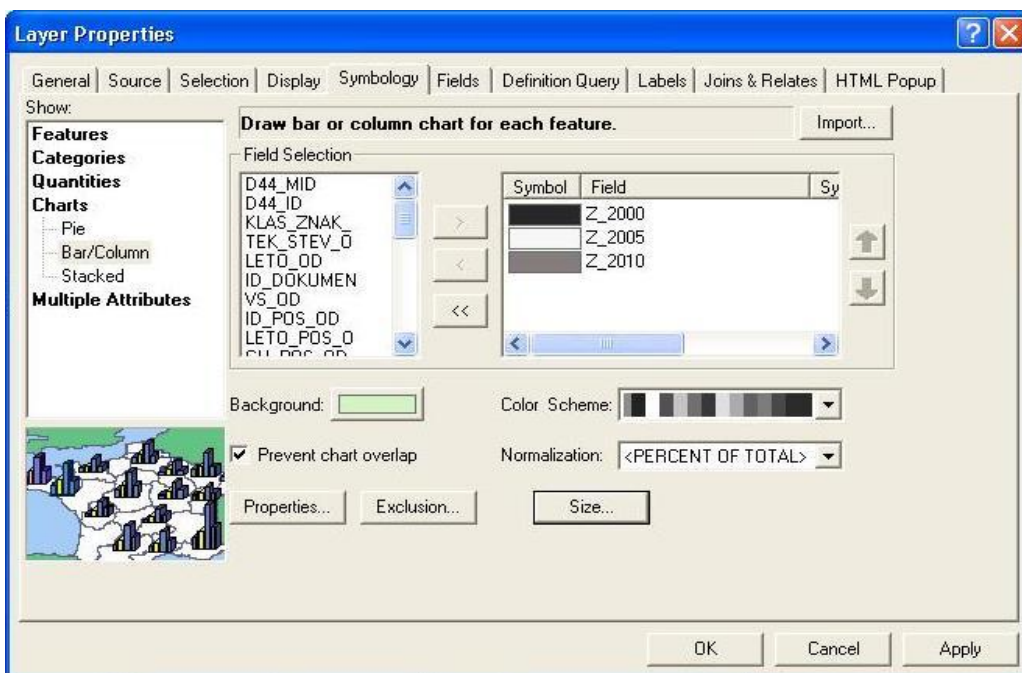
Velikost krogov odraža delež moškega prebivalstva. Največ moških živi v osrednji Sloveniji, najmanj pa v Notranjsko-kraški regiji.



Slika 17: Delež moškega prebivalstva, ponazorjen z velikostjo krogov.

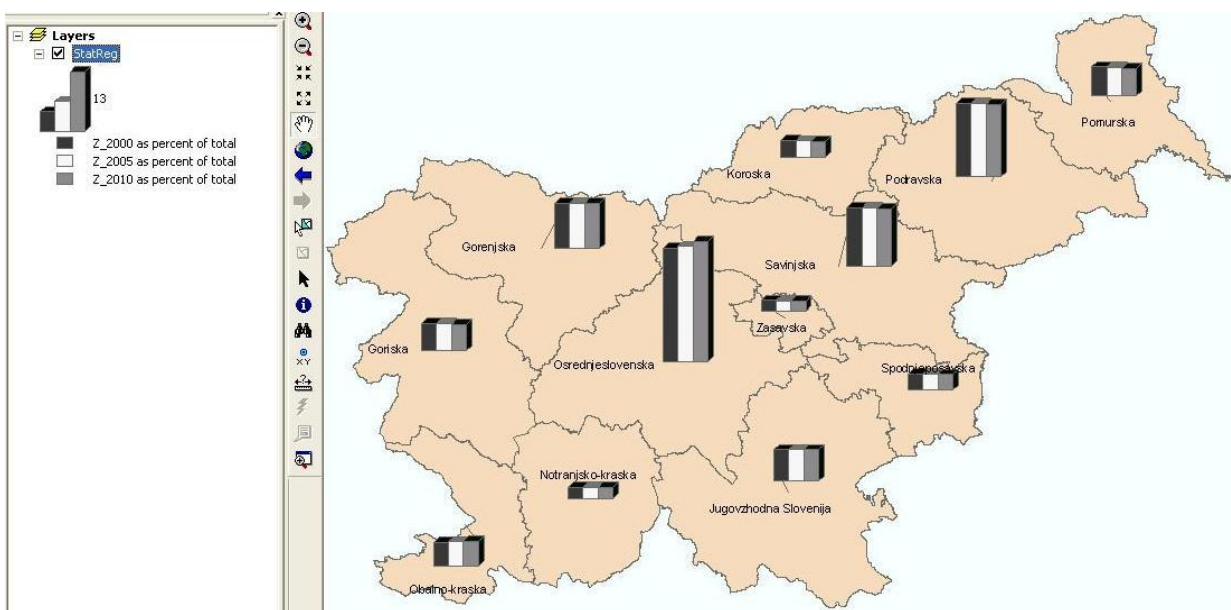
Ena od možnosti editiranja slojev je prikaz večjega števila atributov na karti, kar poveča vsebino informacije. Če nas zanima, kako se je število žensk v posameznih statističnih regijah spreminjalo skozi čas, lahko izberemo med lastnostmi sloja na zavihku »Symbology« možnost »Charts«. Določimo zelen način prikaza podatkov, kjer izberemo recimo »Bar/Column«. V polju »Normalization« izberemo opcijo »PERCENT OF TOTAL«. Gre za normalizacijo, pri kateri vrednost atributa prikažemo kot delež od vsote vseh vrednosti po vseh izbranih letih (2000, 2005, 2010) [2]:

$$\text{delež atributa} = \frac{\text{vrednost atributa v določenem letu}}{\text{vsota vrednosti vseh atributov v vseh letih}} \quad (1)$$



Slika 18: Nastavitve lastnosti sloja StatReg za prikaz relativnega števila žensk po posameznih statističnih regijah v letih 200, 2005 in 2010.

Podatki na karti nudijo lažjo predstavo o relativnem številu žensk v posameznih statističnih regijah. Delež žensk v vseh izbranih letih je največji v Osrednjeslovenski regiji, najmanjši pa v Zasavski ter Notranjsko-kraški regiji. Na karti so razvidne tudi primerjave znotraj posamezne regije. Tako je naprimer v Osrednjeslovenski regiji delež žensk glede na izbrana leta v porastu, saj delež žensk v letih med 2000 in 2010 narašča.



Slika 19: Relativno število žensk v letih 2000, 2005 in 2010 v posameznih statističnih regijah.

4 GEOSTATISTIKA IN RAZŠIRITVENI MODUL GEOSTATISTICAL ANALYST

Razširitev omogoča statistično analizo prostorskih podatkov, kreiranje celotne površine na podlagi vrednosti spremenljivk na merjenih lokacijah, oblikovanje prostorskih trendov in primerjavo različnih statističnih modelov.

Statistične metode prostorske interpolacije temeljijo na statistični analizi prostorskih podatkov in statističnem modeliranju porazdelitve spremenljivk v prostoru. V splošnem z objektivno prostorsko interpolacijo na osnovi prostorsko neenakomerno razporejenih izmerjenih podatkov izračunamo vrednosti želenih spremenljivk na lokacijah, kjer ni meritev [5]. Na podlagi točkastih meritev napovedujemo vrednosti za neznane lokacije znotraj območja preučevanja. Interpolacija je spreminjanje točkovnih informacij v zvezna površja.

Področje statistične interpolacije se največkrat imenuje geostatistika, metoda prostorske interpolacije pa Kriging. Interpolirane vrednosti za izbrane lokacije, kjer ni meritev, izračunamo kot linearno kombinacijo vrednosti meritve v izbrani okolici točke in jih imenujemo interpolirane vrednosti [18].

4.1 Predstavitev podatkov

Za predstavitev možnosti razširitvenega modula Geostatistical Analyst bomo uporabili:

- shapefile podatkovni sloj Meja, v katerem je izrisana meja Kalifornije,
- shapefile podatkovni sloj OzonTocke s točkami meritev ozona,
Sloj OzonTocke vsebuje podatke o povprečni koncentraciji ozona v letu 1996, merjeni v delcih na milijon. Meritve so se izvajale dnevno v 8 urnih razmakih. V našem primeru bomo kot kritično mejo določili maksimalno povprečno vrednost ozona 0.12 delcev na milijon.
- shapefile podatkovni sloj Mesta z lokacijami večjih mest v Kaliforniji.

Strukturiran proces kreiranja površine povprečne koncentracije ozona v zraku je sledeč:

Prikaz podatkov

Dodajanje slojev in njihov prikaz v ArcMap-u.

Pregled podatkov

Raziskovanje statističnih lastnosti podatkov.

Izbira modela

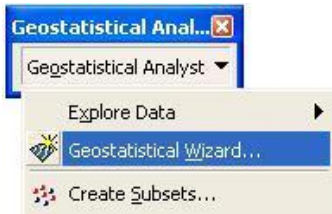
Izbiranje ustreznega modela za kreiranje površine.

Ovrednotenje modela

Določimo lahko, kako dobro je model napovedal neznane vrednosti.

4.2 Kreiranje površine povprečne koncentracije ozona

V ArcMap-u najprej omogočimo vidnost razširitve. Za začetek si bomo ogledali, kako kreirati površino koncentracije ozona s pomočjo čarovnika. Najprej kliknemo na orodno vrstico »Geostatistical Analyst«, nato pa izberemo možnost »Geostatistical Wizard«.



Slika 20: Možnost »Geostatistical Wizard« v orodni vrstici »Geostatistical Analyst«.

Uporabili bomo točkovno oziroma navadno Kriging metodo, zato si jo bomo podrobneje ogledali. Navadni Kriging temelji na predpostavki, da lokalne aritmetične sredine niso nujno enake aritmetični sredini populacije. Zato pri ocenjevanju vrednosti na neznani lokaciji Kriging izhaja le iz vrednosti atributa na sosednjih točkah [13].

Na lokaciji, kjer ne poznamo vrednosti atributa, skuša Kriging oceniti vrednost po formuli:

$$Z^*(u) = \sum_{i=1}^n \lambda_i * Z(u_i) \quad (2)$$

u predstavlja lokacijo z neznanom vrednostjo atributa, na kateri je ocenjena vrednost atributa enaka $Z^*(u)$. n je število sosednjih točk, na podlagi katerih določamo ocenjeno vrednost, λ_i pa predstavlja utež za posamezno sosednjo točko. Navadni Kriging predpostavlja, da je vsota vseh uteži enaka 1.

$$\sum \lambda_i = 1 \quad (3)$$

Povzamemo lahko, da je ocena neznane vrednosti atributa na točki u uteženo povprečje znanih vrednosti sosednjih točk.

Ocenjena vrednost atributa se verjetno razlikuje od dejanske vrednosti na tej lokaciji. Tej razliki pravimo napaka ocene.

Razpršenost ocenjene vrednosti se imenuje varianca ocene:

$$s_z^2 = \frac{\sum_{i=1}^n (Z^*(u) - Z^*(p))_i^2}{n} \quad (4)$$

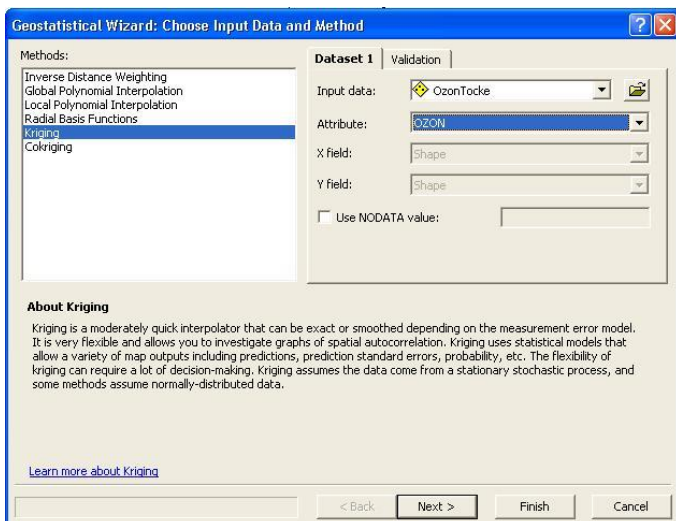
Standardna napaka ocene pa je koren variance ocene:

$$s_z = \sqrt{s_z^2} \quad (5)$$

Ocena in standardna napaka ocene sta odvisni od izbora uteži. Kriging skuša najti optimalne uteži, pri katerih je standardna napaka ocene minimalna.

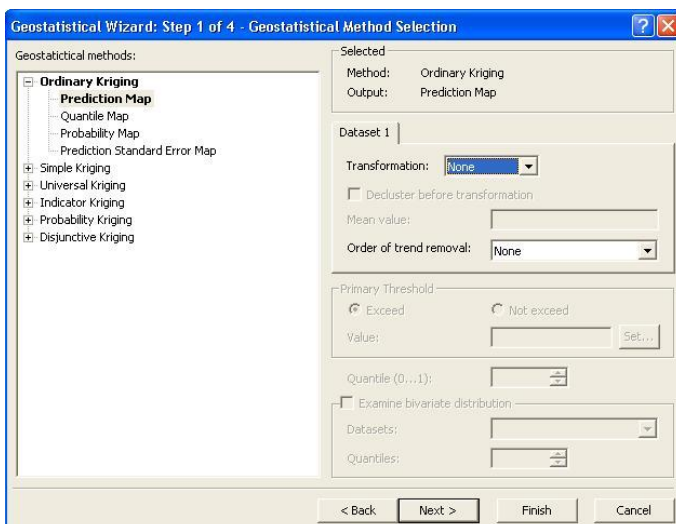
Matematično-statistično ozadje Kriginga je dokaj zapleteno, vendar Geostatistical Analyst omogoča tako imenovano "black box" analizo, pri kateri samega postopka, da pridemo do končnega rezultata, ni potrebno poznati. Takšne analize so na prvi pogled precej privlačne, saj se nam ni potrebno mučiti z zahtevnimi enačbami in funkcijami, vendar moramo biti zelo pazljivi pri uporabi in interpretaciji rezultatov [19].

V polju »Input Data« izberemo sloj OzonTocke. V polju »Attribute« izberemo OZON. Med metodami izberemo »Kriging«.



Slika 21: Izbrani vhodni podatki ter metoda v čarovniku Geostatistical Analyst.

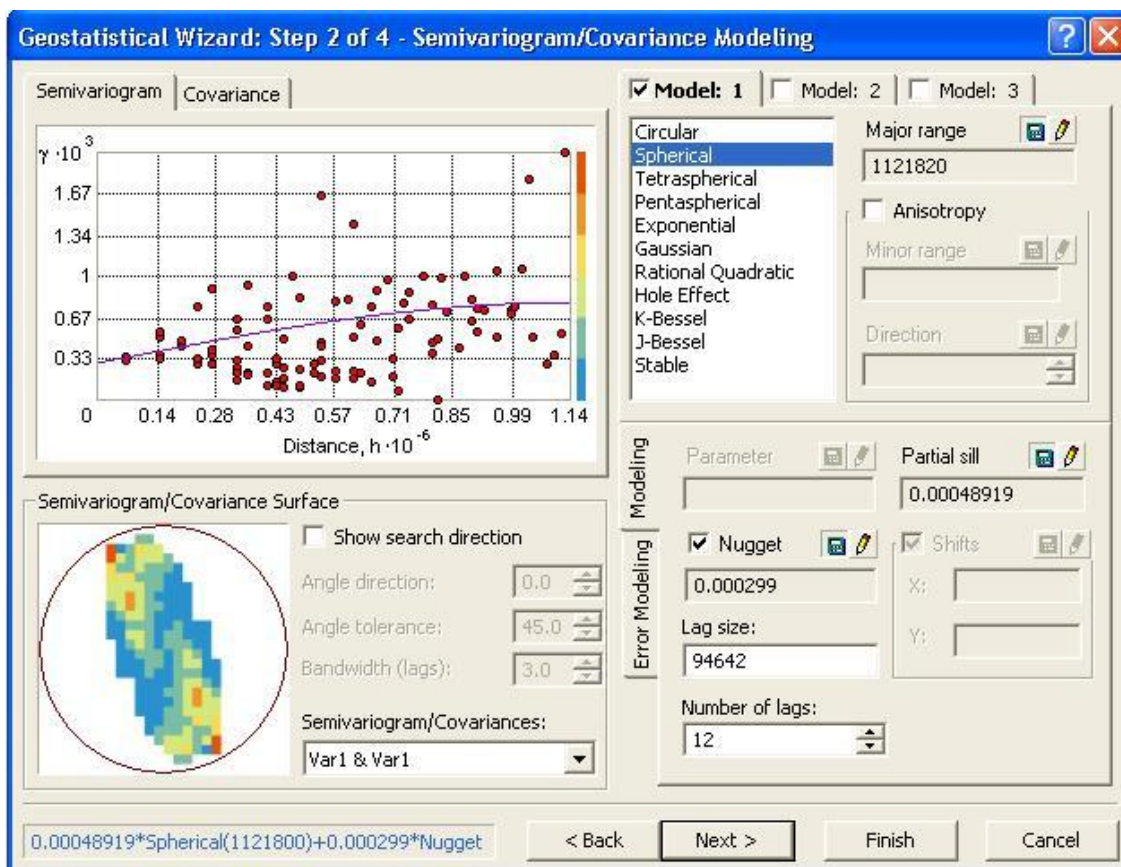
Pustimo izbrani opciji »Ordinary Kriging« in »Prediction Map« ter nadaljujemo.



Slika 22: Izbrana metoda v čarovniku Geostatistical Analyst.

Kriging temelji na predpostavki, da je prostorska variabilnost pojava, ki jo predstavljajo Z vrednosti, statistično homogena na celotni površini. Uporaba Kriginga je zato vprašljiva, če se proučevani pojav v prostoru razporeja izrazito nezvezno, s hitrimi spremembami, saj končni rezultat zato ne bo odražal realnega stanja [19].

V naslednjem oknu imamo možnosti raziskovanja prostorskih povezav med merjenimi točkami. Semivariogram deluje na principu, da so objekti, ki so si bližje, bolj podobni kot objekti, ki so med sabo bolj oddaljeni in preveri to predpostavko. Kaže razliko kvadratov vrednosti med vsakim parom točk na različnih razdaljah. Semivariogram je graf, ki opisuje lastnosti regionalizirane spremenljivke. Uporabljamo ga za določanje uteži, ki jih bomo upoštevali v posamezni točki z znano vrednostjo atributa, za oceno vrednosti v novi točki.

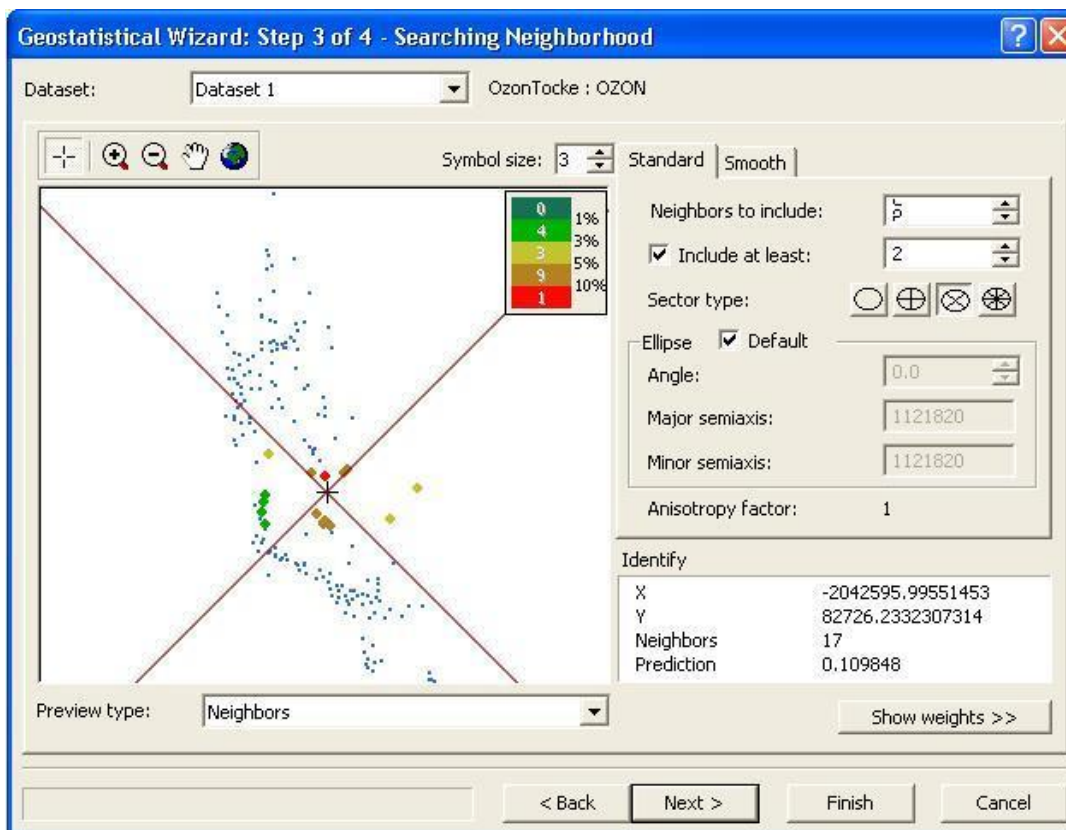


Slika 23: Semivariogram.

Iz slike je razvidno, da semivarianca narašča z oddaljenostjo med paroma meritv. Bližje kot so meritve, manjša je variabilnost vrednosti in boljša je napoved. Razpon x osi je opredeljen z največjo razdaljo med dvema znanima meritvama.

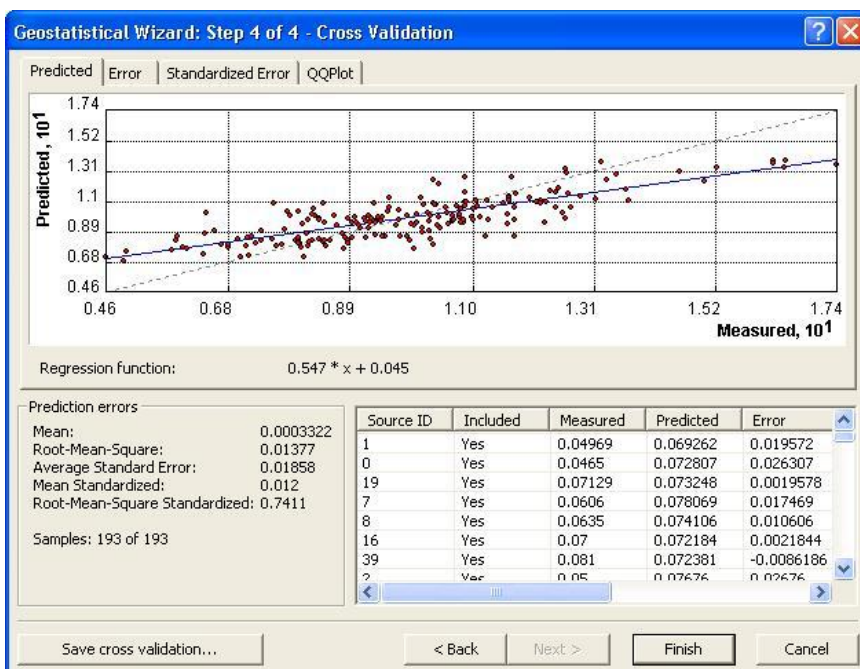
V naslednjem oknu križec označuje lokacijo brez izmerjene vrednosti ozona. Za napovedovanje vrednosti na lokacijah, kjer nimamo podatkov o meritvah, se uporabijo vrednosti lokaci, kjer so bile izvedene meritve. Izberemo lahko število sosednjih točk, na podlagi katerih se bodo računale

vrednosti ozona na neizmerjenih lokacijah. Rdeče točke bodo pri izračunu imele večjo težo kot zelene točke, saj so bližje lokaciji, za katero izračunavamo vrednost.



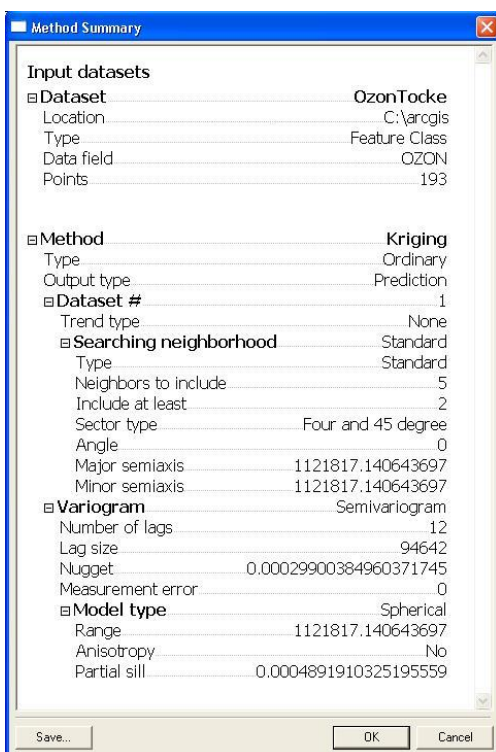
Slika 24: Vnos števila meritev, ki jih bomo upoštevali pri ocenjevanju vrednosti ozona na neizmerjenih lokacijah.

Na naslednjem zavihku imamo informacijo o tem, kako dobro model oceni vrednosti na neznanih lokacijah. Če kreiramo več površin z uporabo različnih modelov, nam podatki na naslednjem oknu omogočajo, da primerjamo, kateri model daje natančnejše ocene vrednosti ozona. Manjše kot so standardizirane napake, boljši in veljavnejši je model.



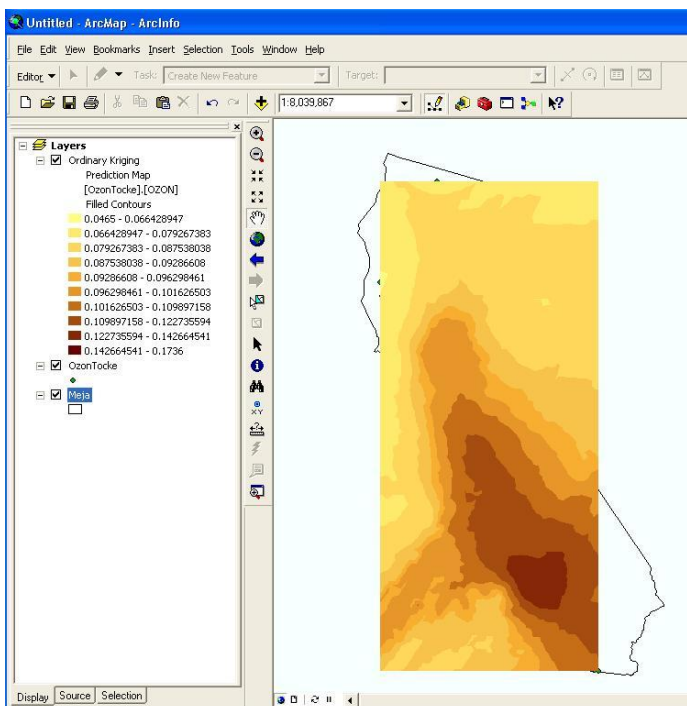
Slika 25: Napake pri ocenjevanju površine koncentracije ozona.

Okno »Output Layer Information« povzema informacije o metodi in parametrih, ki so bili uporabljeni za kreiranje površine.



Slika 26: Povzetek uporabljene metode in parametrov.

Karta z vrednostmi koncentracije ozona se pojavi kot prvi sloj.

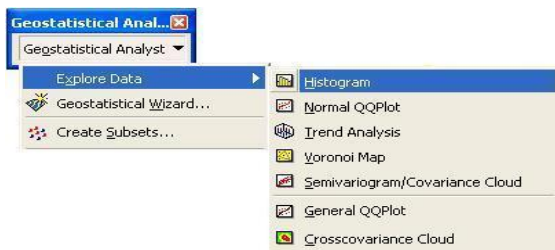


Slika 27: Kreirana površina koncentracije ozona.

4.3 Porazdelitev podatkov

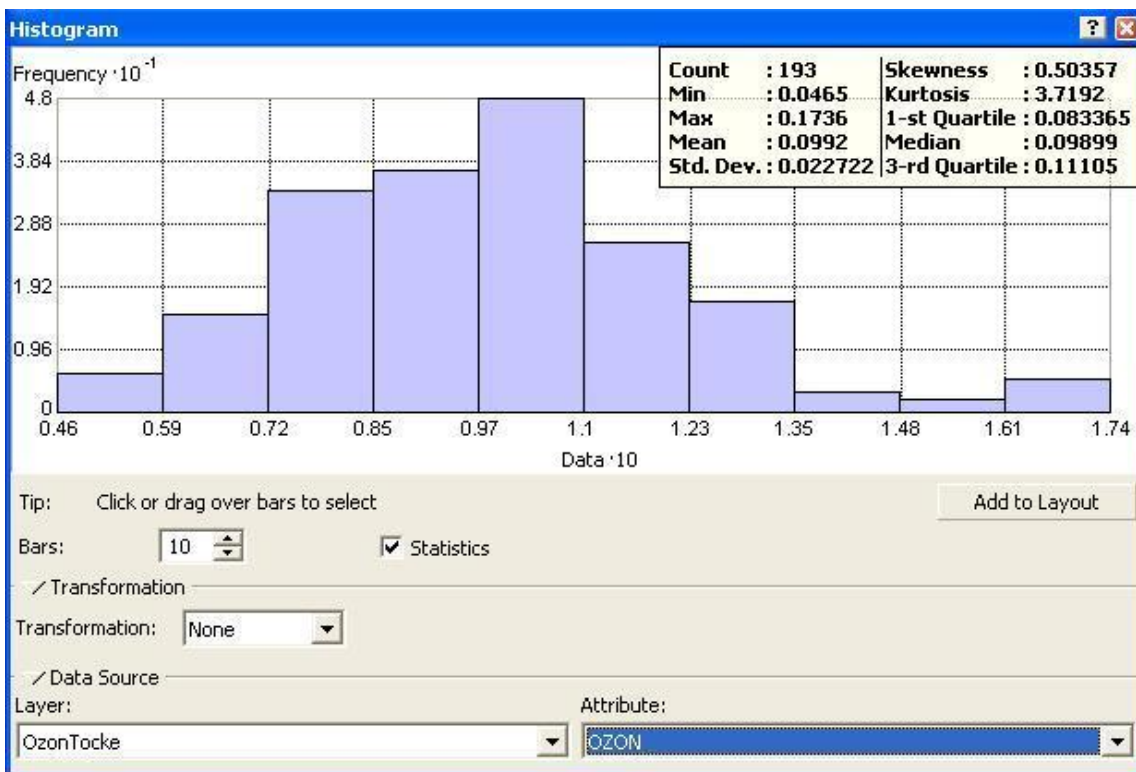
Interpolacijske metode za generiranje površine, ki smo jih uporabili v čarovniku, dajo najboljše rezultate, kadar je distribucija podatkov normalno porazdeljena. Predpostavljamo normalno porazdelitve je bistveno za množico statističnih izračunov, saj velja, da se vzorec, ki je izvzet iz celotne populacije, porazdeljuje približno po normalni krivulji tudi, če vrednosti vseh enot populacije niso porazdeljene normalno. Če porazdelitev odstopa od normalne, je potrebno podatke transformirati v normalno porazdelitev. S histogramom raziščemo porazdelitev posameznega atributa.

Kliknemo možnost »Explore Data« in izberemo »Histogram«.



Slika 28: Izbira opcije »Histogram«.

V polju »Layer« izberemo sloj OzonTocke. V polju »Attribute« pa izberemo OZON.

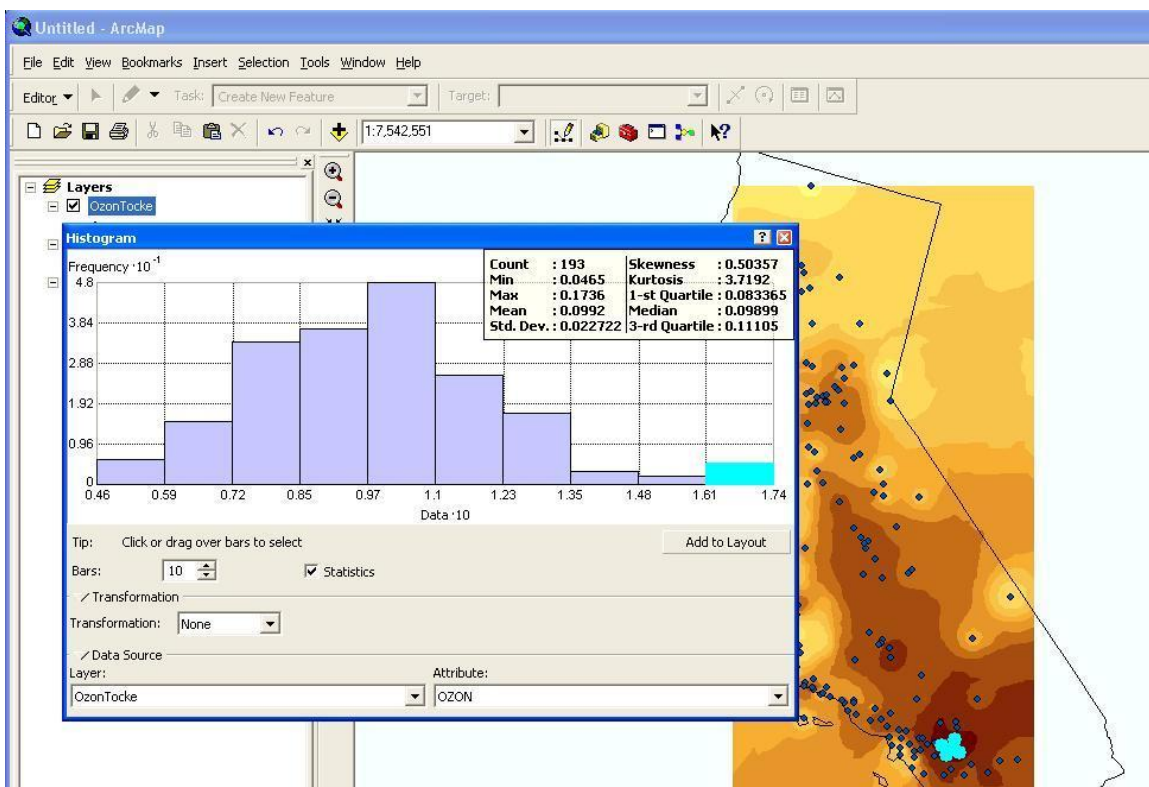


Slika 29: Histogram za atribut koncentracije ozona.

Histogram vsebuje deset razredov. Relativni delež podatkov je ponazorjen z višino posameznega stolpca. Pomembne lastnosti histograma so aritmetična sredina, razpršitev in simetrija. Za hitro preverjanje normalnosti porazdelitve lahko uporabimo primerjavo vrednosti aritmetične sredine in mediane. Mediana je srednja vrednost zaporedja števil, ki razdeli po velikosti razvrščena števila na dve enaki polovici po številu elementov. Če sta vrednost aritmetične sredine in mediane podobni, kot v našem primeru, je izpolnjen eden od pogojev za normalno porazdelitev podatkov.

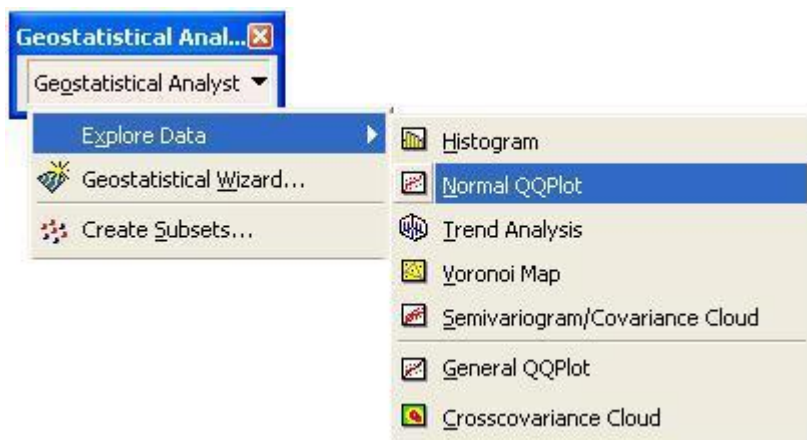
Iz histograma je razvidno, da je porazdelitev po obliki unimodalna, saj ima en modus (en vrh). Modus je vrednost, ki se najbolj pogosto pojavlja v množici vrednosti.

Porazdelitev je tudi simetrična. Nekoliko izstopa desni del porazdelitve, vendar vsebuje razmeroma majhen del točk z visoko koncentracijo ozona. Če kliknemo na stolpec z vrednostmi od 0.161 do 0.174 delcev na milijon, se obarvajo točke v bližini Los Angelesa. Gre za točke iz iste lokacije, in njihovo število je razmeroma majhno, zaradi česar lahko zaključimo, da je porazdelitev simetrična. Simetričnost porazdelitve ter unimodalnost sta še dva pogoja, ki morata biti izpolnjena, da lahko povzamemo, da so podatki porazdeljeni normalno.



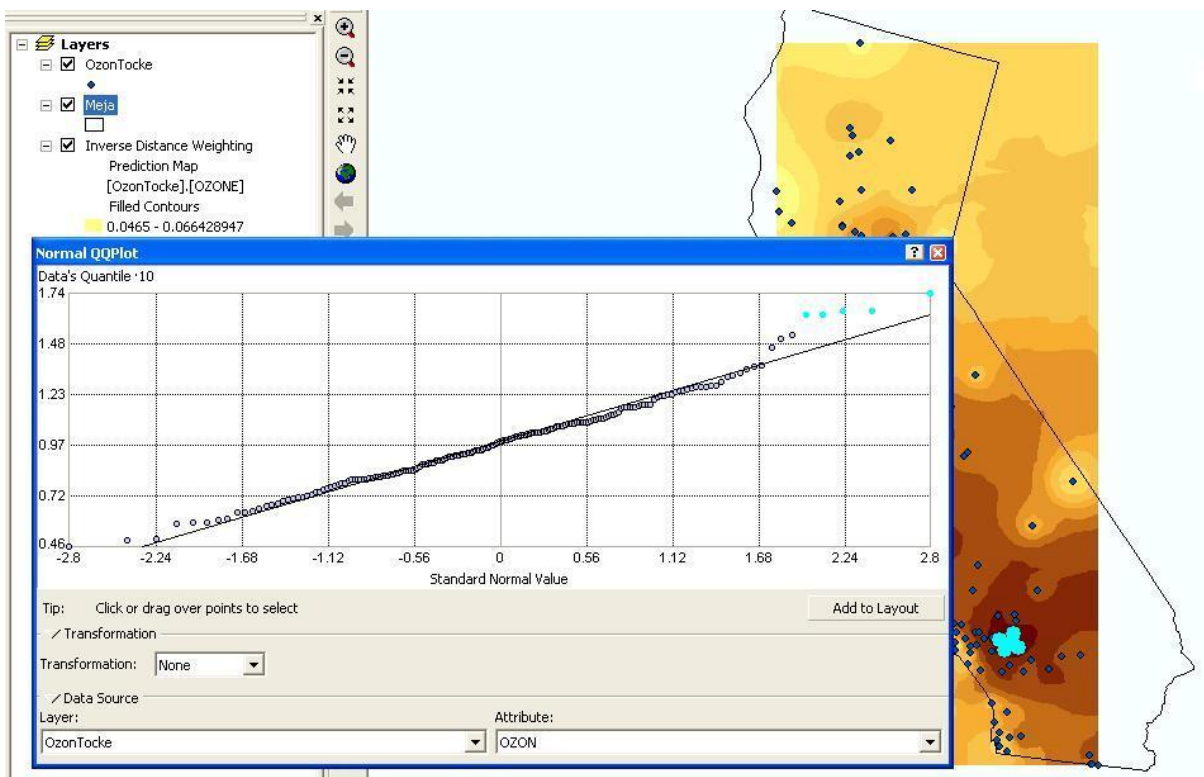
Slika 30: Lokacija točk z visoko koncentracijo ozona.

Grafični način preverjanja normalnosti porazdelitve predstavlja možnost QQPlot. V ArcMapu izberemo »Explore Data« ter »Normal QQPlot«.



Slika 31: Možnost »Normal QQPlot«.

V polju »Layer« izberemo sloj OzonTocke, v polju »Attribute« pa izberemo OZON.



Slika 32: Izbira sloja OzonTocke in atributa OZON.

QQPlot predstavlja dodaten način preverjanja, ali so naši podatki porazdeljeni normalno. Grafična metoda primerja kvantile dveh porazdelitev. Kvantil je vrednost, ki deli število podatkov na enake dele. Po številu delov, na katere delimo po nekem atributu urejene podatke, poznamo različne kvantile, ki jih označujemo kot p-kvantile. Kvantil 0,5 je enak mediani in razdeli podatke na dva enaka dela.

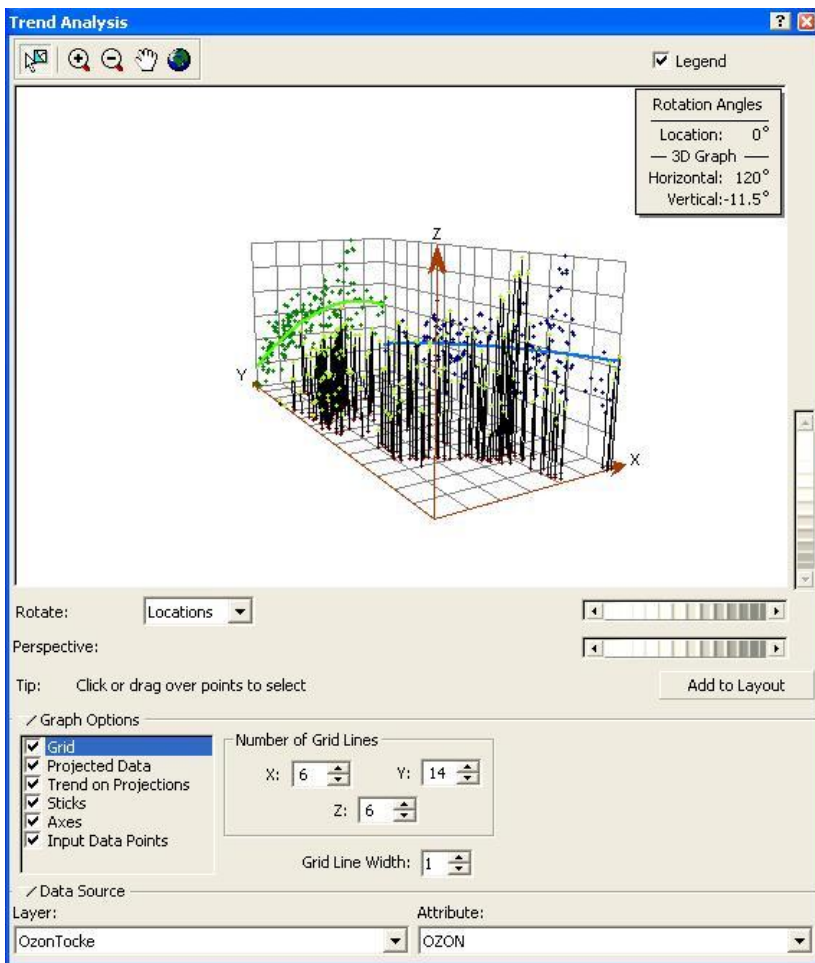
Za dve identični porazdelitvi bi se v grafu izrisala ravna črta. V našem primeru se naša porazdelitev podatkov primerja z normalno porazdelitvijo. Bolj ravna kot je črta naših podatkov, večja je možnost, da so podatki normalno porazdeljeni. Naši podatki v majhni meri odstopajo od ravne črte. Glavna odstopanja se tako kot v histogramu pojavljajo pri visoki koncentraciji ozona.

V primeru, ko bi porazdelitev podatkov v histogramu ali QQ grafu v večji meri odstopala od normalne porazdelitve, bi bilo podatke potrebno transformirati pred uporabo nadaljnjih interpolacijskih tehnik.

4.4 Identifikacija trendov

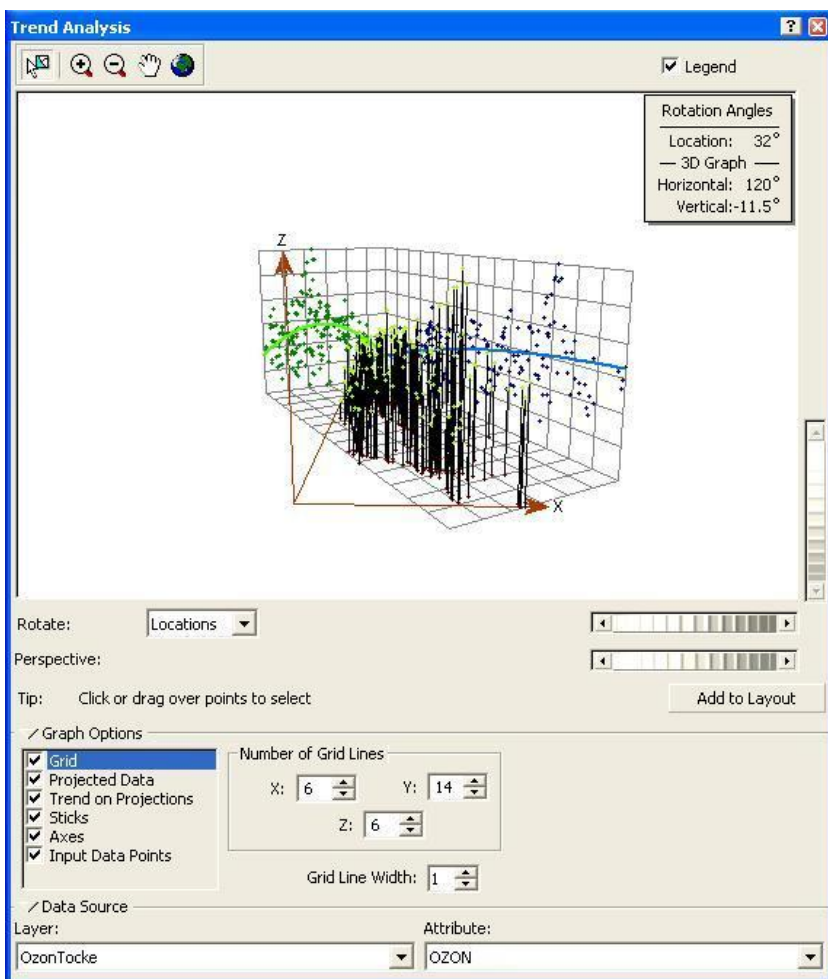
Geostatistical Analyst omogoča tudi identifikacijo glavnih trendov v podatkih. V ArcMapu izberemo »Explore Data« ter »Trend Analysis«.

V polju »Layer« določimo sloj OzonTocke, v polju »Attribute« pa izberemo OZON.



Slika 33: Izbira sloja in atributa za analizo trendov ter globalni trendi.

Vsaka vertikalna črta predstavlja lokacijo in višino posamezne točke, ki odraža povprečno koncentracijo ozona. Točke so projicirane v oseh vzhod-zahod in sever-jug. Točkam je narisana najbolj prilegajoča črta. Če je le ta vodoravna, potem točke ne odražajo posebnega trenda. Zelena črta se začne z nizkimi vrednostmi, ki proti vzhodu naraščajo.



Slika 34: Rotiranje prikaza trendov.

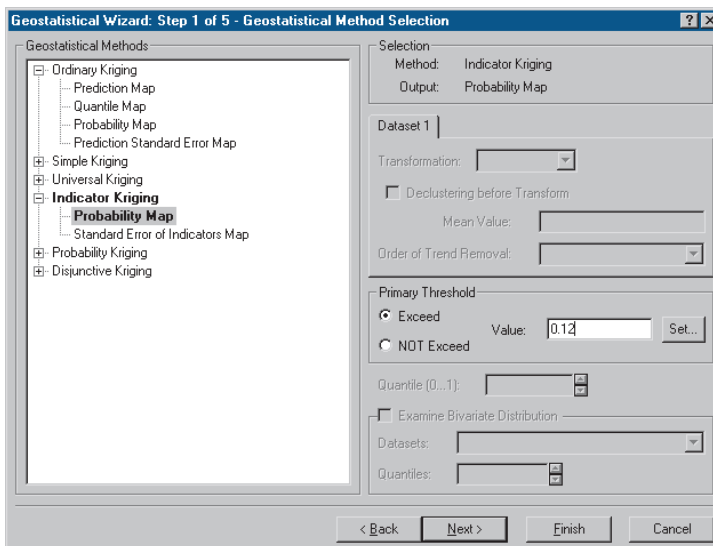
S pomočjo spremembe rotacije na 30° lahko vidimo, da obstaja trend v obliki navzdol obrnjene črke U. Za opis trenda bi izbrali polinom druge stopnje. Opazimo še, da je trend v smeri južnozahodno-severovzhodno. Trend verjetno odraža dejstvo, da je onesnaženje nizko ob obali, v smeri proti gostejši populaciji znotraj države narašča, proti goram z upadom prebivalstva pa ponovno pada.

4.5 Karta verjetnosti, kjer vrednost atributa presega kritično mejo

Razširitev omogoča tudi prikaz verjetnosti, da bo koncentracija ozona presegla izbrano kritično vrednost. Če želimo naprimer določiti območja verjetnosti, kjer povprečna vrednost ozona preseže 0.12 delcev na milijon, lahko uporabimo več različnih tehnik, ogledali pa si bomo indikatorsko Kriging tehniko. Ta pretvori podatke v vrednost 0, če je vrednost ozona manjša od kritične vrednosti, ter v vrednost 1, če je vrednost ozona večja od kritične vrednosti.

Ponovimo nekaj korakov iz prvega dela, kjer smo uporabili čarovnika. Kliknemo na orodno vrstico »Geostatistical Analyst«, nato pa izberemo možnost »Geostatistical Wizard«. V polju

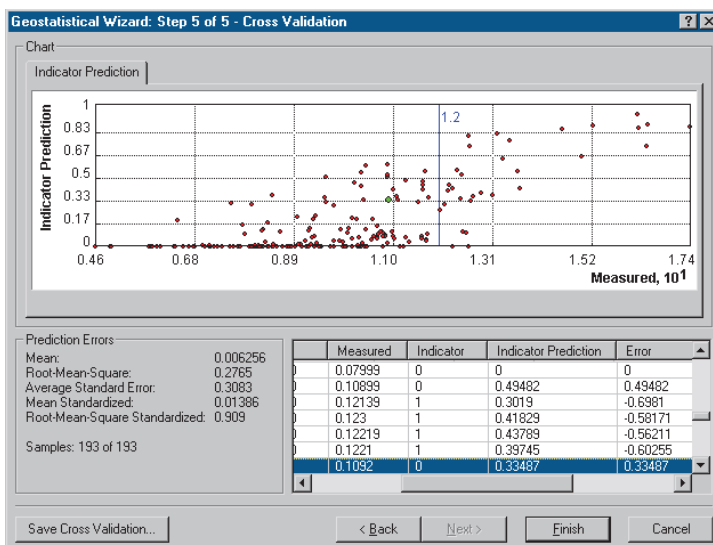
»Input Data« izberemo sloj OzonTocke. V polju »Attribute« izberemo OZON. Med metodami izberemo »Kriging«. Izberemo opciji »Indicator Kriging« in »Probability Map« ter nastavimo mejno vrednost na 0,12.



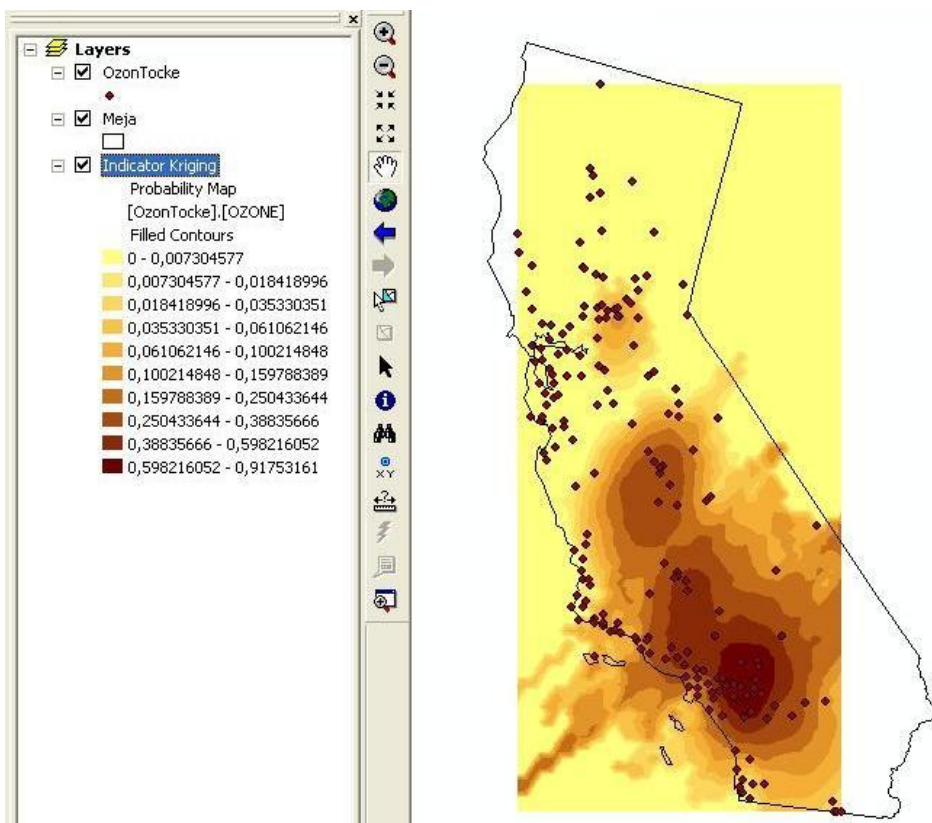
Slika 35: Nastavitve mejne vrednosti za prikaz karte verjetnosti.

Na oknu »Semivariogram/Covariance Modeling« ter »Searching Neighborhood« pustimo izvirne nastavitve ter kliknemo gumb za nadaljevanje.

Modra linija predstavlja mejno vrednost 0,12 delcev na milijon. Na levi strani so točke z vrednostjo, manjšo od mejne, ki imajo transformirano vrednost 0. Na desni strani pa so točke z vrednostjo, ki je večja od mejne, in imajo transformirano vrednost 1. V tabeli izberemo poljubno vrstico z indikator vrednostjo 0. Točka se obarva v grafu in je na levi strani mejne linije.



Slika 36: Graf z mejno vrednostjo koncentracije ozona ter točkami meritev.



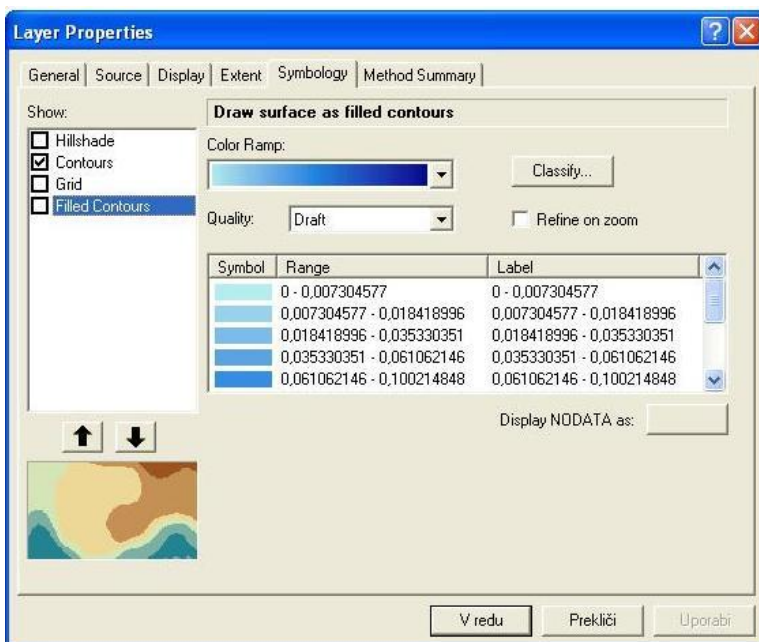
Slika 37: Karta z območji verjetnosti, s katero povprečna vrednost ozona v letu 1996 presega kritično vrednost 0,12 delcev na milijon.

S temno rjavo bravo so prikazana območja, kjer je verjetnost, da povprečna koncentracija ozona presega 0,12 delcev na milijon, večja od 59,8%.

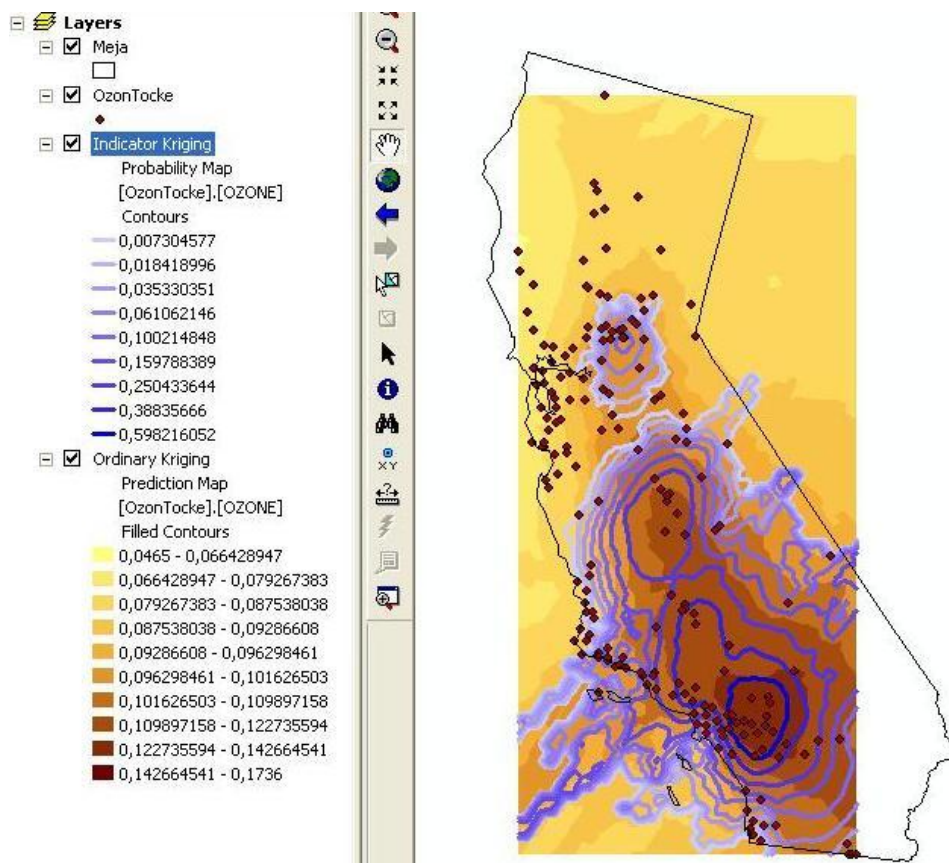
4.6 Končna karta s površino povprečne koncentracije ozona in območji preseganja kritične meje ozona

Skupaj želimo prikazati površino koncentracije ozona ter karto verjetnosti preseganja kritične meje ozona. Nivoje verjetnosti prikažemo kot črte na kreirani površini povprečne koncentracije ozona.

Z desnim klikom na sloju »Kriging layer« izberemo »Properties« ter kliknemo na zavihek »Symbolology«. Odstranimo kljukico pri možnosti »Filled Contours« ter izberemo opcijo »Contours«. V polju »Color Ramp« izberemo poljubno barvno označevanje.

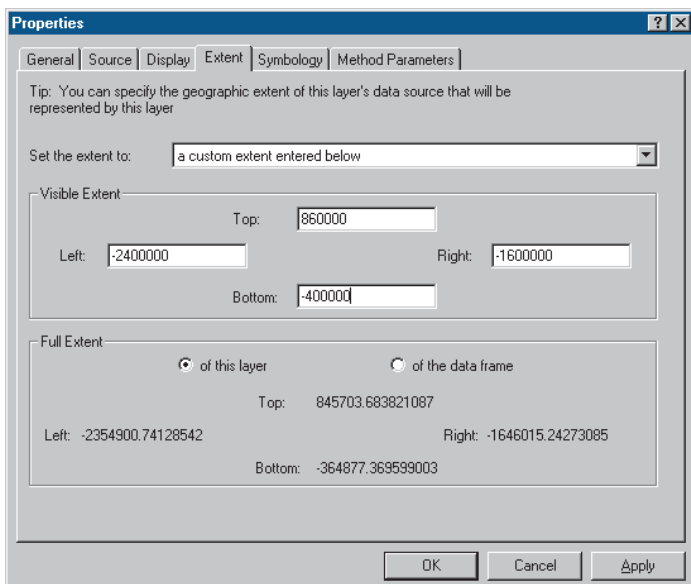


Slika 38: Nastavitve za prikaz karte verjetnosti.



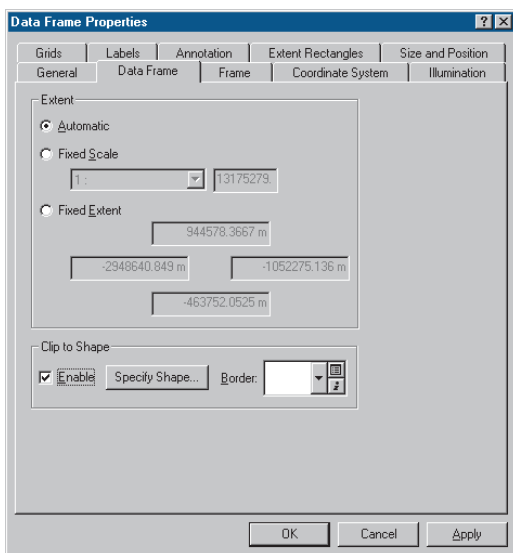
Slika 39: Kreirana površina in verjetnost presejanja mejne vrednosti ozona.

Karti se ne nahajata znotraj mej države. Za prilagoditev obeh kart meji države na obeh slojih pod lastnostmi kliknemo zavihek »Extent«. V polju »Set the extent to« nastavimo možnost »select a custom extent entered below« ter vnesemo želene vrednosti za prikaz obsega karte.



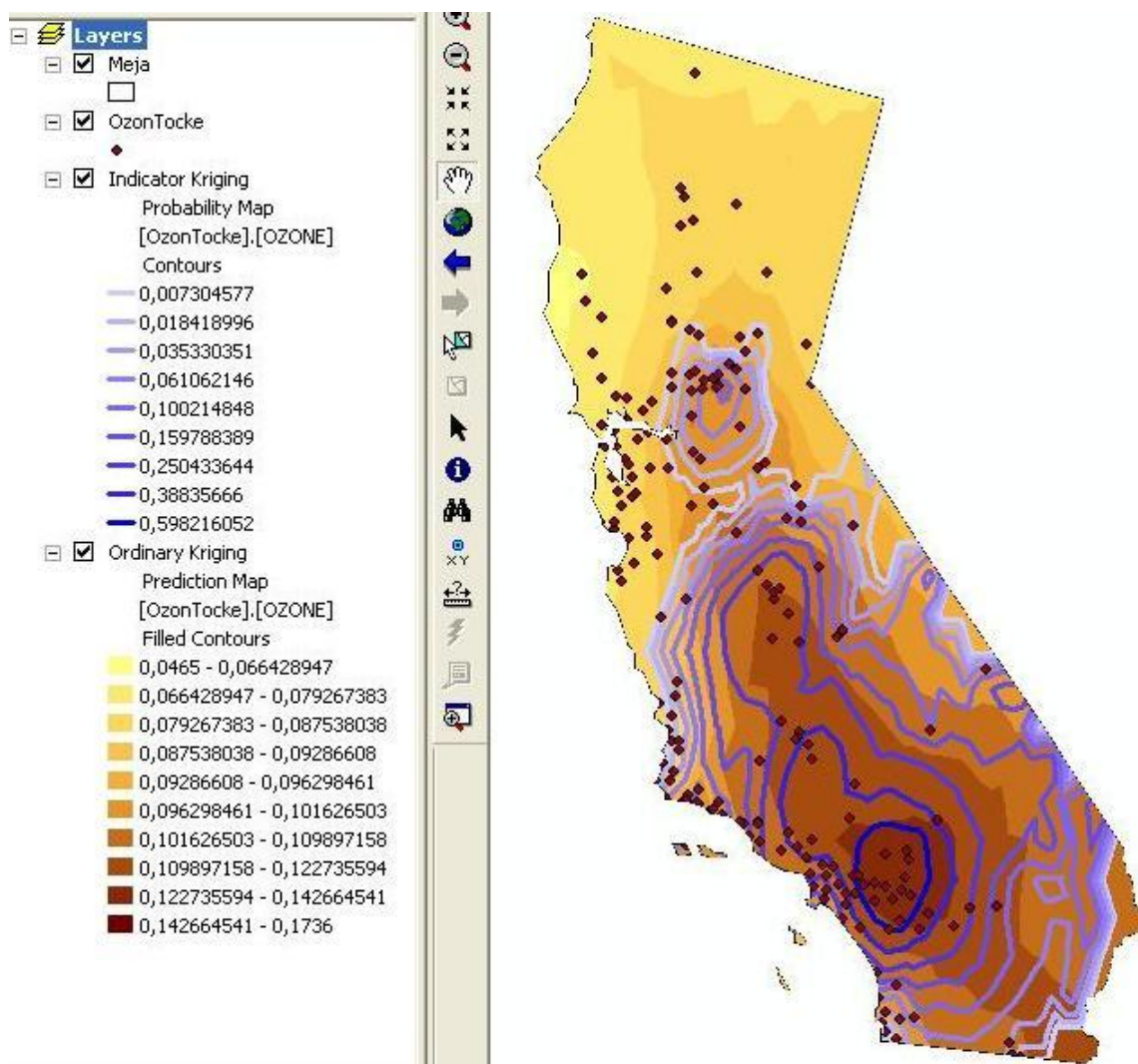
Slika 40: Nastavitve za prikaz karte.

Z desnim klikom na »Layers« se nam odpre okno, kjer izberemo možnost »Properties«. Kliknemo zavihek »Data Frame« ter označimo polje »Enable Clip«. Kliknemo gumb »Specify Shape« ter na novem dialogu izberemo možnost »Outline of Features«. V polju »Layer« izberemo sloj Meja ter potrdimo izbiro.



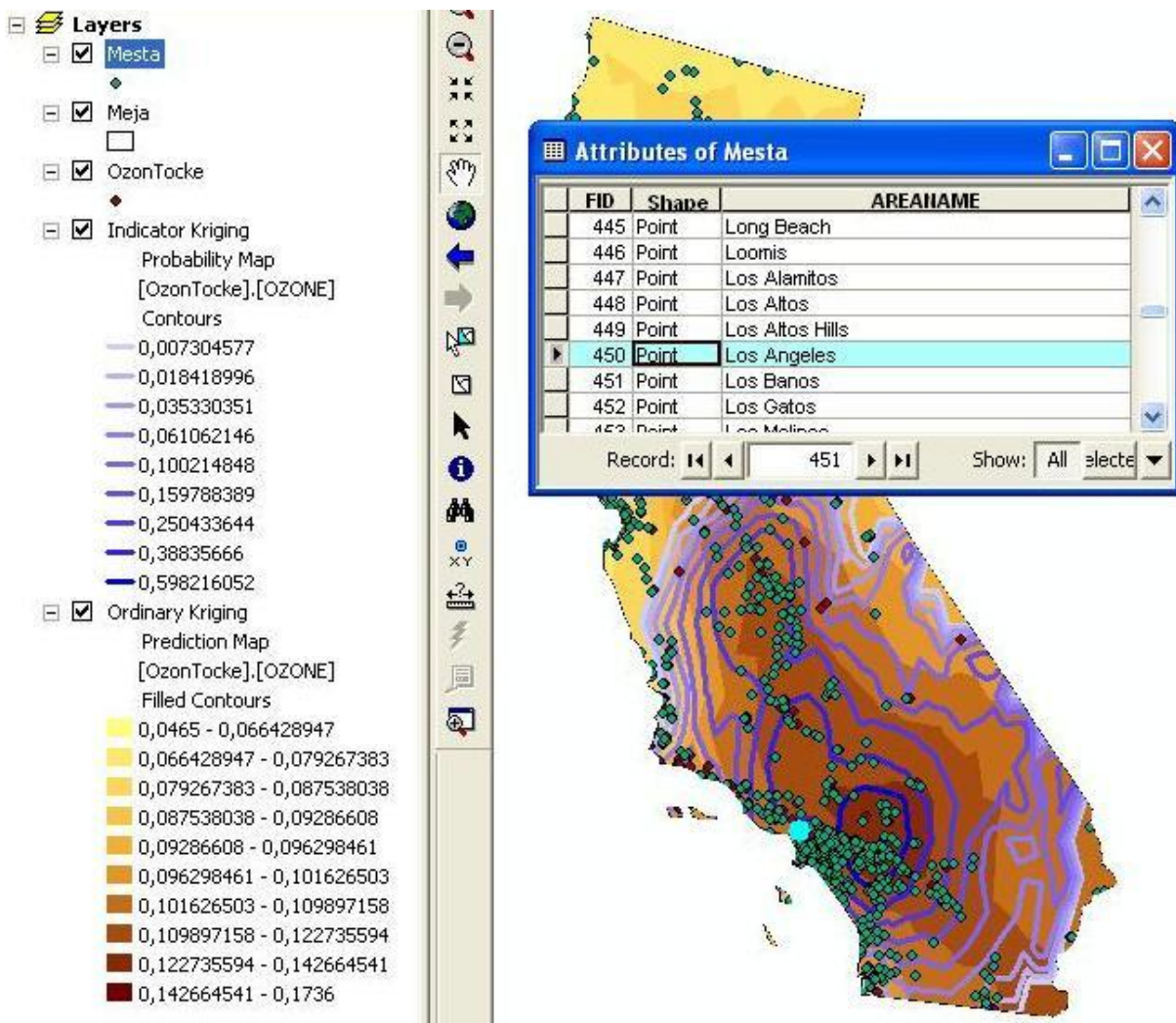
Slika 41: Izbira mej Kalifornije kot oblike, ki se ji naj prilagodi kreirana površina.

Končna karta izgleda takole.



Slika 42: Kreirana površina povprečne koncentracije ozona in nivoji verjetnosti presejanja mejne koncentracije ozona v mejah države.

Za boljšo predstavo dodamo še sloj Mesta. Kliknemo gumb »Add Data« ter dodamo sloj. Z desnim klikom na sloj Mesta med možnostmi izberemo »Open Attribute Table«. V tabeli poiščemo vrednost atributa »AREA_NAME« Los Angeles, ki se obarva na karti.



Slika 43: Iskanje lokacij na karti preko tabele atributov.

5 ZAKLJUČEK

Namen diplomske naloge je bil predstaviti, kakšne so zmožnosti vizualizacije ter statistične analize prostorskih podatkov v orodju ArcMap. Raziskovanja sem se lotila tako, da sem se najprej podrobneje spoznala z aplikacijo, nato pa sem se osredotočila na pripravo podatkov ter delo s praktičnimi primeri.

Ugotovila sem, da je ArcMap z vidika vizualizacije ter geostatistike vsestransko uporabna aplikacija, ki ima zraven še mnogo drugih funkcionalnosti za delo s prostorskimi podatki. Vizualizacija prostorskih podatkov na kartah daje uporabniku vsebinsko bogatejše informacije, enako velja tudi za vizualizacijo rezultatov geostatističnih analiz na prostorskih podatkih. Aplikacija ponuja mnogo različnih možnosti geostatističnih analiz, odvisno od lastnosti porazdelitve podatkov, na katerih se analize izvajajo.

Posebej bi poudarila, da je ArcMap lahko izjemno uporabno orodje v rokah večjega uporabnika. Priprava podatkov ter izbira in poznavanje ustreznih analiz namreč zahtevajo določen nivo predznanja. Vendar je v orodju samem ter na internetu veliko informacij ter vodičev, s pomočjo katerih uporabnik v kratkem času pridobi potreben nivo znanja in izkušnje. Za pripravo podatkov v ustrezni obliki, primerni za izvajanje analiz, je potrebno osnovno znanje oziroma nekaj izkušenj z orodjem ArcMap. Na hitro lahko kaj več o tem izvemo iz uporabniške pomoči ArcGIS Desktop Help, ki se namesti v okviru ArcGIS Desktop orodja. V največjo pomoč pa so zagotovo praktični primeri in vodiči, ki so na voljo na spletu. Ob pregledu večjega števila konkretnih primerov dobimo idejo, kako je potrebno organizirati podatke, da je možno izvesti želene analize. Za izvajanje statističnih analiz je potrebno predznanje s področja statistike. Zaželjeno je razumevanje osnovnih statističnih pojmov ter statističnih postopkov. Za razširitev, ki sem jo uporabila sama, obstaja veliko napisanega tudi v uporabniški pomoči v ArcMap orodju. Vendar je že za razumevanje te pomoči zaželjeno statistično znanje, ki ga lahko uporabnik dobi iz literature, kjer so predstavljeni osnovni statistični pojmi, ter preko konkretnih primerov, ki so dostopni na spletu.

V realnem življenju pri sprejemanju prostorskih odločitev izhajamo iz realnega problema, ki ga želimo rešiti s pomočjo ArcMap aplikacije. Na tem mestu sem se soočala s težavami, da je problemskih situacij sicer veliko, vendar je zelo težko pridobiti ustrezne podatke, na podlagi katerih bi lahko poiskala rešitve. Ugotovila sem, da je za reševanje prostorskih problemov izjemnega pomena dostop do ustreznih podatkov. Kot namen diplomske naloge sem nato opredelila možnosti vizualizacije ter statističnih analiz. Izhajala sem iz funkcionalnosti ArcMap-a ter pridobila podatke, ki so ustrezali zahtevi, da je možno z njimi ponazoriti zmožnosti vizualizacije ter statistike.

V diplomski nalogi sem raziskala možnosti vizualizacije ter geostatistika v orodju ArcMap. V nadaljnjih raziskavah bi lahko raziskali, kakšne rešitve na tem področju ponujajo druga orodja za delo s prostorskimi podatki.

6 KAZALO SLIK

Slika 1: Geografski informacijski sistem po tehnološki opredelitvi.	4
Slika 2: Primerjava rastrske in vektorske ponazoritve točkovnih, linijskih in območnih objektov.	6
Slika 3: Primerjava funkcij rastrskega in vektorskega podatkovnega modela.	6
Slika 4: ArcGIS programska oprema.	8
Slika 5: Družina produktov ArcGIS Desktop.	9
Slika 6: Poligonski shapefile podatkovni sloj ter pripadajoča tabela atributov v ArcMapu (Vir: GURS, april 2010).	21
Slika 7: Pripravljeni podatki o prebivalstvu po spolu v različnih letih po posameznih statističnih regijah (Vir: SURS, junij 2010).	22
Slika 8: Opcija združevanja tabel v ArcMapu.	23
Slika 9: Združevanje tabel na podlagi naziva statističnih regij iz tabele StatReg.shp ter Excelove tabele.	24
Slika 10: Tabela podatkovnega sloja StatReg po združevanju vsebuje attribute obeh tabel.	24
Slika 11: Nastavitve lastnosti sloja StatReg za prikaz kategorij števila žensk v posameznih statističnih regijah.	25
Slika 12: Kategorije števila žensk v posamezni statistični regiji.	26
Slika 13: Nastavitve lastnosti sloja StatReg za prikaz ženskega prebivalstva, razvrščenega v razrede.	26
Slika 14: Razredi števila ženskega prebivalstva glede na statistične regije.	27
Slika 15: Možnost »Multiple attributes«.	28
Slika 16: Nastavitve za prikaz relativnih vrednosti.	28
Slika 17: Delež moškega prebivalstva, ponazorjen z velikostjo krogov.	29
Slika 18: Nastavitve lastnosti sloja StatReg za prikaz relativnega števila žensk po posameznih statističnih regijah v letih 200, 2005 in 2010.	30
Slika 19: Relativno število žensk v letih 2000, 2005 in 2010 v posameznih statističnih regijah.	30
Slika 20: Možnost »Geostatistical Wizard« v orodni vrstici »Geostatistical Analyst«.	32
Slika 21: Izbrani vhodni podatki ter metoda v čarovniku Geostatistical Analyst.	33
Slika 22: Izbrana metoda v čarovniku Geostatistical Analyst.	33
Slika 23: Semivariogram.	34
Slika 24: Vnos števila meritev, ki jih bomo upoštevali pri ocenjevanju vrednosti ozona na neizmerjenih lokacijah.	35
Slika 25: Napake pri ocenjevanju površine koncentracije ozona.	36
Slika 26: Povzetek uporabljene metode in parametrov.	36
Slika 27: Kreirana površina koncentracije ozona.	37
Slika 28: Izbira opcije »Histogram«.	37
Slika 29: Histogram za atribut koncentracije ozona.	38
Slika 30: Lokacija točk z visoko koncentracijo ozona.	39
Slika 31: Možnost »Normal QQPlot«.	39
Slika 32: Izbira sloja OzonTocke in atributa OZON.	40
Slika 33: Izbira sloja in atributa za analizo trendov ter globalni trendi.	41

Slika 34: Rotiranje prikaza trendov.	42
Slika 35: Nastavitve mejne vrednosti za prikaz karte verjetnosti.	43
Slika 36: Graf z mejno vrednostjo koncentracije ozona ter točkami meritev.	43
Slika 37: Karta z območji verjetnosti, s katero povprečna vrednost ozona v letu 1996 presega kritično vrednost 0,12 delcev na milijon.	44
Slika 38: Nastavitve za prikaz karte verjetnosti.	45
Slika 39: Kreirana površina in verjetnost preseganja mejne vrednosti ozona.	45
Slika 40: Nastavitve za prikaz karte.	46
Slika 41: Izbira mej Kalifornije kot oblike, ki se ji naj prilagodi kreirana površina.	46
Slika 42: Kreirana površina povprečne koncentracije ozona in nivoji verjetnosti preseganja mejne koncentracije ozona v mejah države.	47
Slika 43: Iskanje lokacij na karti preko tabele atributov.	48

7 LITERATURA

- [1] P.A., Burrough, R.A., McDonnell, *Principles of Geographical Information Systems*, New York: Oxford University Press, 1998, pogl. 2.
- [2] G. Dailey, "Normalizing Census Data Using ArcMap", *ArcUser*, str.52-53, 2006.
- [3] S., Drobne, T., Podobnikar, *Osnovni pojmi v geografskih informacijskih sistemih*, Univerza v Ljubljani: Fakulteta za gradbeništvo za geodezijo, 1999, pogl. 1.
- [4] P., Goovaerts, "Geostatistics for Natural Resource Evaluation", *Oxford Univeristy Press*, str. 369-376, 1997.
- [5] D., Kastelec, *Objektivna prostorska interpolacija meteoroloških spremenljivk in njihovo kartranje : doktorska disertacija*, Ljubljana: Fakulteta za matematiko in fiziko, 2001, pogl. 2.
- [6] J., Požun, *Vizualna reprezentacija kompleksnih informacij: Primer aplikacije newsmap*, Univerza v Ljubljani: Fakulteta za družbene vede, 2008, pogl. 2.
- [7] P., Tan, M., Steinbach, V., Kumar, *Introduction to data mining*, Addison-Wesley, 2006, pogl. 2.
- [8] ArcGIS Desktop Help.
- [9] (2010) ArcGIS. Dostopno na: <http://www.esri.com/software/arcgis/index.html>
- [10] (2010) ArcGIS. Dostopno na: <http://www.esri.com/software/arcgis/arcgisserver/index.html>
- [11] (2010) ArcGIS. Dostopno na: <http://www.esri.com/software/arcgis/arcgismobile/index.html>
- [12] (2010) ArcGIS. Dostopno na: http://www.esri.com/products/index.html#developer_tools_panel
- [13] (2010) Department of civil and enviromental engineering. Dostopno na: <http://www.cee.vt.edu/ewr/environmental/teach/smprimer/kriging/kriging.html>
- [14] (2010) Dostopno na: http://www.scdhec.gov/presentations/ESRI_Conference_08/tws/workshops/tw_114.pdf
- [15] (2010) Geostatical Analyst. Dostopno na: http://www.usgsquads.com/downloads/factsheets/ESRI_geostatistical_analyst.pdf

- [16] (2010) Geografski informacijski sistemi. Dostopno na:
<http://www.km.fgg.uni-lj.si/predmeti/TIUS/data/GIS/TIUS-%20GIS%20in%20PA.pdf>
- [17] (2010) NCGIA. Dostopno na: <http://www.ncgia.ucsb.edu/giscc/units/u002/>
- [18] (2010) Osnove geografskih informacijskih sistemov. Dostopno na:
<http://users.volja.net/damijanZ/GIS/gis.html>
- [19] (2010) Rešitve za terenski zajem podatkov. Dostopno na:
<http://www.geoservis.si/uporabno/gis/gis.htm>
- [20] (2010) Wikipedija. Dostopno na:
http://sl.wikipedia.org/wiki/Statisti%C4%8Dne_regije_Slovenije
- [21] (2010) Wikipedija. Dostopno na: <http://sl.wikipedia.org/wiki/Geografija>