

Prostorové modelování a základy geostatistiky



Petr Dobrovolný

Obsah přednášky

1. Prostorové a kartografické modelování
2. Kartografický model a základní členění modelů
3. Mapová algebra -modelování s rastrovými daty
4. Úvod do geostatistiky
5. Statistický popis prostorového uspořádání objektů (bodů, linií, ploch)
6. Přehled metod interpolace

Základní literatura

BORROUGH, P.A., McDONNELL, R.,A (1988):
Principles of Geographical Information Systems.
Oxford University Press, Oxford, 333s.

DeMERS, N.M. (2002): GIS Modelling in Raster.
Wiley & Sons. New York, 203 s.

TOMLIN, D. (1990): Geographic Information
Systems and Cartographic Modelling. Prentice
Hall, New Jersey.

LEE, J., WONG, D.W.S. (2001): Statistical
Analysis with ArcView GIS, J. Willey & Sons,
New York, 192 s.

Kartografické modelování

Metodologie zpracování prostorových dat, která
používá map jako proměnných a prostorových
operací jako konstruktorů (operátorů)
v algebraických výrazech.

Modelování není ovládnání SW (znalost práce s textovým
editorem nezaručuje znalost psaní dobrých románů).

Význam modelování v GIS

Jednotlivé
úrovně
abstrakce
reality



Kartografické modelování je základním nástrojem
analýzy v GIS. Nástroji GIS je vytvářen model
reálného světa. Nad tímto modelem je vytvářen další
(formální) model, který popisuje, předpovídá,
simuluje, syntetizuje či analyzuje procesy a jevy –
kartografický model.

Modelování je proces, který vyžaduje porozumění
a vyjasnění následujících kroků:

1. Co má být výstupem modelu (co modeluji – chci
popsat daný stav, nebo něco předpovědět, nebo
najít vhodné řešení z několika možných)
2. Jakých datových zdrojů bude zapotřebí - vrstvy
a vazby mezi nimi
3. Způsob prezentace reálného světa (prostorové
informace) v modelu – jakým datovým modelem
budou objekty a vztahy mezi nimi
prezentovány.
4. Jakých nástrojů použijeme

Základní dělení kartografických modelů

V závislosti na zvolených kritériích existuje několik způsobů klasifikací modelů v GIS, neexistuje jednotná terminologie, mezi dvěma kategoriemi modelů často neexistuje striktní hranice ale postupný přechod.

3 základní kritéria třídění (DeMers, 2002):

1. Dělení modelů na základě účelu použití
2. Dělení modelů na základě použité metodologie a technik
3. Dělení modelů na základě použité logiky

1. Dělení modelů podle účelu použití

1.1 Modely deskriptivní - účelem je popsat současné procesy a funkce složek v systému (krajiny, dopravního systému, ekosystému, ...)

1.2 Modely preskriptivní - účelem je predikovat možné budoucí stavy systému.

Deskriptivní modely

• Jedná se o modely **pasivní**. Hlavním úkolem je popis jednotlivých složek studovaného území, jejich současného stavu. Odpovídají tedy především na otázku: Co je tady? (támhle).

• Snaží se kvantifikovat současný stav – např. geometrické vztahy objektů na mapě či mapách. Od jednoduchého měření (vzdálenost, plocha, obvod, ...) až po komplexní integrující charakteristiky (tvar, izolovanost, struktura, hierarchie).

• Tyto modely umožňují především zjišťovat **strukturní vztahy** (pattern), porovnávat je mezi dvěma mapami či územími.

• Deskriptivní modely mohou být charakterizovány také jako **syntetické**. Postihují vztahy mezi jednotlivými vrstvami

• **Dekonstruktivní modely** – pro testování sensitivity či významnosti jednotlivých složek v modelu lze jednu každou složku z modelu odstranit a studovat její vliv na fungování celku.

Preskriptivní modely

• Lze je charakterizovat jako **aktivní**. Řeší odpověď na otázku „Co by mělo být?“ – typický příklad: optimální lokalizace objektu.

• Možná předpověď vzniká na základě důkladné deskripce

• Dvoji charakter výstupu z modelu:

- A. nejlepší (jedno) řešení na základě stanovených kritérií
- B. variantní řešení (několik) které vyhovují zadaným kritériím.

• Umožňují analyzovat **dynamiku** procesů.

Dále je lze dělit na modely

- holistické – modelují proces jako celek
- atomistické- model „vrstev“ – čtenější

2. Dělení modelů podle použité metodologie

2.1 Modely stochastické – založené na pravděpodobnostní statistice (např. regresní model)

2.2 Deterministické – založené jasně definovaných vztazích příčiny a následku (např. odtokový model, model predikce znečištění, modely erozní - USLE)

3. Dělení modelů podle použité logiky

3.1 Induktivní metody – konstrukce obecného modelu z jednotlivých tématik, běžně využívají empirických vztahů.

3.2 Deduktivní metody – od obecných pravidel ke specifickému, algoritmické

Kartografické modelování představuje typický analytický nástroj GIS.

Problémy spojené s využíváním analytických nástrojů:

- Uživatel velmi často porovnává neporovnatelné (porovnávání nominálních a ordinálních dat). Příklad: Numerické hodnoty reprezentující nominální kategorie (např. landuse) jsou děleny či násobeny ordinálními, intervalovými či poměrovými daty a ve výsledku nedávají žádný význam.
- Zaměňuje příčinu a následek
- Nepokouší se testovat alternativní způsoby
- Nedostatečná znalost o podstatě, způsobu sběru či způsobu organizace dat v databázi

Základní druhy prostorových údajů podle použité škály hodnot:

NOMINÁLNÍ - jména, kvalitativní, disjunktní (nepřekrývají se), lze s nimi provádět pouze některé logické operace (porovnávání, existence či neexistence)

ORDINÁLNÍ - údaje, jež lze seřadit podle určitého kritéria. Je známé jejich pořadí, nikoli však rozdíl (rozdíl mezi dvěma určitými kategoriemi nemusí být stejný jako rozdíl mezi jinými dvěma kategoriemi).

INTERVALOVÁ - lze je také odečítat, zjišťovat rozdíly, mají pevnou stupnici. Obsahují nulu, která je "uměle" vytvořena v dané stupnici a nachází se "uprostřed".

POMĚROVÁ - umožňují provádět i operaci dělení. Nula zde vyjadřuje neexistenci dat. Jsou jednostranně omezena.

Proč modelovat s rastrovými daty?

- Je řada věcí, které lze modelovat vhodněji s použitím vektorových dat (síťové analýzy) a naopak (modelování spojitých polí – povrchů, difúzní modely)
- Používané algoritmy na řešení problémů v prostředí „vektorovém“ a „rastrovém“ se často značně liší.
- Důvody historické (průhlednost a jednoduchost prvních nástrojů) a pragmatické (cena a dostupnost)
- Doména environmentálního modelování (na rozdíl od modelů technických - doména CAD systémů).

Proč modelovat s rastrovými daty?

- Větší šíře rastrových nástrojů
- Možnost prezentace diskretních i spojitých prostorových informací
- Možnost definování vztahů sousednosti
- Využitelnost principů mapové algebry pro kartografické modelování, jednoduchost řady algoritmů
- Výrazně rostoucí podíl prostorových dat získávaných metodou DPZ a jejich rastrová povaha
- Kompatibilita s obdobnými datovými zdroji

Výhody

- Jednoduchý sběr dat
- Vhodné pro kombinování více vrstev – mapová algebra, snadné překrývání
- Jednoduchost operací, jednoduchost datové struktury (matice)
- Vhodné pro analýzy souvislých povrchů
- Rychlé polohové dotazování

Nevýhody

- Velká paměťová náročnost
- Omezená geometrická přesnost daná rozlišením (velikostí buňky)
- Nižší vizuální kvalita
- Nevhodnost pro síťové analýzy

Rastrový datový model

- viz. základní přednáška

Přesnost a kvalitu zobrazení reality za pomoci rastrového modelu ovlivňují tyto faktory:

- Rozměr buňky („prostorové rozlišení“ v DPZ)
- Způsob přiřazení hodnoty atributu buňce
- Počet úrovní použitých pro zaznamenání hodnot atributu („radiometrické rozlišení“ v DPZ)

Ad 1) Menší rozměr buňky znamená přesnější geometrii, avšak geometricky narůstající paměťový prostor, (vzrůst počtu tzv. smíšených pixelů v DPZ)

Ad 2) Hodnoty atributu v buňce mohou být určeny několika způsoby:

- Bodová hodnota zaznamenaná kdekoli v buňce
- Aritmetický průměr několika bodových měření
- Vážený aritmetický průměr (váha – např. plocha či počet bodů s určitou hodnotou)
- Minimální či maximální hodnota atributu v ploše
- Hodnota s nejvyšší vahou
- Výsledek interpolačního algoritmu

Ad 3) Používané hloubky rozlišení (počet stavů použitých k popisu změn v chování atributu):
Binární rastr (0,1), integer, floating point (real)

Zdroje rastrových dat:

- primární (obrazová data DPZ)
- sekundární
 - metody interpolace bodových měření (GPS)
 - metody rasterizace vektorových dat
 - skenování analogových dat

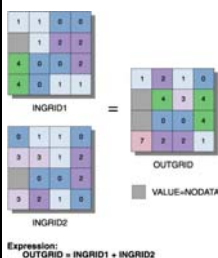
Prezentace základních druhů objektů v rastrovém datovém modelu

- body
- linie
- polygony
- povrchy

Mapová algebra

Soubor metod analýzy prostorových dat uchovávaných v rastrovém datovém modelu. Používá map jako proměnných a prostorových operací jako operátorů v algebraických výrazech.

$$\text{Nová_Mapa} = f(\text{Vstupní_Mapa1}, \text{Vstupní_Mapa2}, \dots)$$



Rastr je analogií matice a MA je analogií maticového počtu.

Rastr využívá modelu vrstev. Jednotlivé vrstvy (vstupní rastry, gridy) jsou pomocí operátorů, funkcí, výrazů či iterací kombinovány do výstupního rastru (většinou).

Mapová algebra – obecný princip

- MA tvoří podstatnou část prostorových analýz v GIS a je také základem řady metod zpracování obrazu v DPZ
- Jsou definovány obecné principy MA, neexistuje norma, nejednotná terminologie a nejednotná syntaxe v závislosti na implementaci v konkrétním systému.
- Řada funkčně shodných prostorových operací vystupuje pod jiným označením v různých systémech).
- Základní myšlenky formuloval Tomlin (1990).

Mapová Algebra v prostředí ArcView

Extenze Spatial Analyst, GRID object

GRID – podtřída třídy Object. Je to objekt, který v ArcView uchovává data v rastrovém formátu. Vhodný pro prezentaci spojitě prostorové informace. Skládá se z buněk (cell).

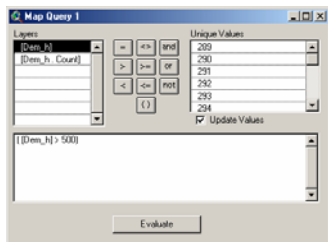
Nově vytvořený grid je soubor dočasný. Je na uživateli, zda ho uloží trvale na disk

Fyzicky je na disku Grid uložen jako složka se jménem, které zadal uživatel. K přenosu jinam je třeba mít i složku s názvem INFO.

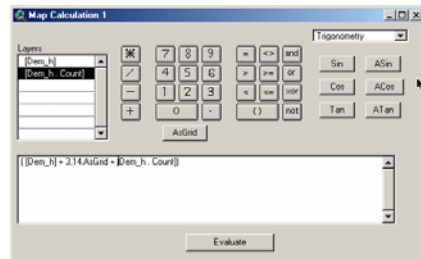
Možné formy použití MA (uživatelská rozhraní)

1. **Prostorové dotazování (Map query)** – výsledkem je tzv. bitová mapa nominálních hodnot (0 – nesplňuje, 1 – splňuje podmínku prostorového dotazu)
2. **Mapový kalkulátor** - Sestavování výrazů mapové algebry pomocí funkcí a operátorů
3. **Skriptovací jazyk AVENUE** - Sekvence výrazů, algoritmus, program
4. **Grafické modelování** - (Model Builder, PCI Modeller), nástroje využívající makrojazyka a grafických prvků vývojových diagramů

Prostorové dotazování (Map query)



Mapový kalkulátor – základní nástroj mapové algebry

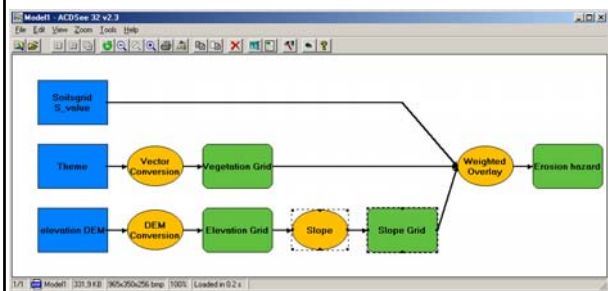


AVENUE Script

```

theView = av.GetProject.FindDoc("View1")
theSrcName = Grid.MakeSrcName( "c:\uran\grid\dem_h" )
if (theSrcName = NIL) then
  MsgBox.Error("Chybí jméno", "Chyba")
  return NIL
end
theGrid = Grid.Make(theSrcName)
theGTheme = GTheme.Make( theGrid )
theView.AddTheme( theGTheme )
theGTheme.SetVisible(true)
theView.Invalidate
  
```

Grafické modelování - Model Builder



Požadavky na vstupní datové soubory

I nejjednodušší model často využívá více datových vrstev majících různý původ, různou projekci, velikost buňky atd.

Pokud používaný modelovací nástroj neřeší sám problém unifikace vstupních dat (např. sčítat dva různé rastry s různým rozsahem či s různou velikostí buňky), je nutné ji provést předem. Tento úkol může zahrnovat dvě činnosti:

1. změnu projekce
2. převzorkování (unifikaci velikosti buňky a výpočet nových hodnot každé buňky)

Ad 1 – musí být známy převodní vztahy (transformační rovnice – např. TRANSMAP)

Ad 2 – nejpoužívanější způsoby převzorkování:

- interpolace metodou nejbližšího souseda
- bilineární interpolace
- kubická konvoluce



Základní způsoby převzorkování

a. Nejbližší soused – jediný možný způsob převzorkování použitelný na nominální (kategorická, kvalitativní) data, jeho použití na data kvantitativní nedává nejlepší výsledky. Je založen na prostém posouvání hodnot do pozice nejbližší buňky, ne na interpolaci

b. Bilineární interpolace – hodnota nové buňky je vypočtena jako vážený průměr čtyř nejbližších buněk původních

c. Kubická konvoluce – hodnota nové buňky je vypočtena jako vážený průměr šestnácti nejbližších buněk původních

Praktické poznámky k předzpracování dat

- Pokud systém dokáže spojit dva rozdílné rastry – má někde nastaven defaultní způsob převzorkování.
- Jak je řešen problém okrajů rastru při převzorkování?
- Je zachován typ dat? (integer vs. Floating point)
- Převzorkováním se mění statistické charakteristiky, nezachovává se minimální a maximální hodnota v gridu

Základní nástroje (objekty) pro manipulaci s prostorovými daty v mapové algebře

Operátory: aritmetické, relační, booleovské, kombinatorní, logické, akumulací, přiřazovací

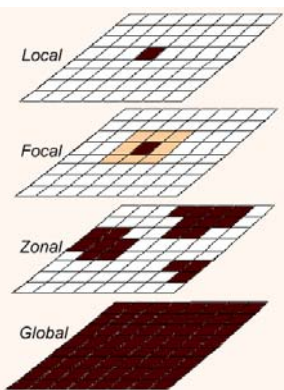
Funkce: aritmetické, logické, trigonometrické, logaritmické

Třídy funkcí v mapové algebře:

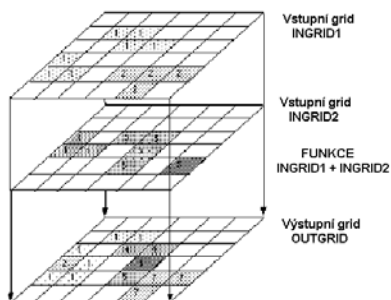
- lokální
- fokální
- zonální
- globální

Základní třídy funkcí (operací) v mapové algebře

- lokální (point)
- fokální (neighborhood)
- zonální (area)
- globální (map)



LOKÁLNÍ OPERACE



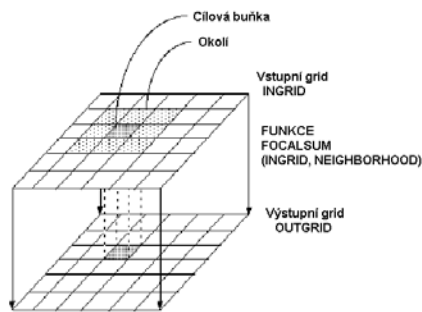
LOKÁLNÍ (BODOVÉ) OPERACE

Nová hodnota buňky je počítána pouze z jedné buňky určitou matematickou operací či jako kombinace několika odpovídajících buněk z jiných vrstev.

Dělení podle počtu vstupních gridů

- Unární operace (vytvoření, přidání konstanty)
- Binární operace (porovnávání)
- Operace s více vstupy (hledání min, max, ...)

FOKÁLNÍ OPEARACE



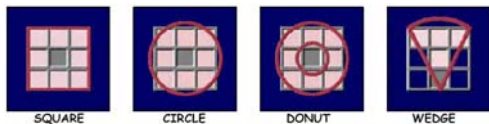
FOKÁLNÍ OPEARACE (neighborhood analysis, filtrace)

- Výstupní hodnota buňky je počítána jako jistá operace s touto buňkou a s buňkami jejího okolí.
- Velikost a tvar okolí lze definovat. Velikost je udávána v jednotkách použitého zobrazení nebo v počtu buněk (je udávána lichým číslem).
- Problém okrajů zpracovávaného gridu (zmenšení o polovinu okénka -1, replikace vstupních hodnot či replikace výsledků)
- Fokální funkce mají velký význam pro zpracování obrazu (filtrace obrazu). Použití v kartografickém modelování: úprava spojitých polí (shlazování DEM)

Fokální funkce

- „fokální“ statistika
- nízkofrekvenční filtrace
- vysokofrekvenční filtrace
- způsoby definování filtrovacího okna a jeho tvary
- filtrace pod maskou (podle zadaného atributu)

- Typy okolí:
- pravoúhelník
 - kružnice
 - mezikruží
 - výseč
 - nepravidelné okolí



Fokální statistika

0	5	10
5	20	10
20	10	30

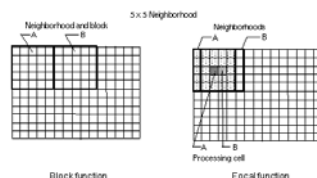
SUM	110
AVERAGE	12,22
MAXIMUM	30
MINIMUM	0
MEDIAN	10
MAJORITY	10
MINORITY	30
DIVERSITY	5
RANGE	30

Příklady využití fokálních funkcí:

- simulování šíření požáru na napě landuse
- odstranění „šumu“
- úprava výsledků klasifikace (potlačení malých plošek)

BLOKOVÉ OPERACE

Jistou variantou fokálních funkcí jsou tzv. blokové funkce. Mají tyto odlišnosti:

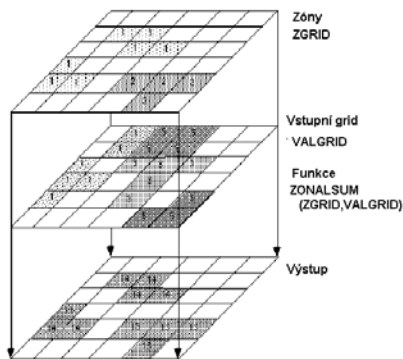


- Na rozdíl od předchozího případu je bloková statistika a blokové funkce počítána ne pro každou buňku, ale pro blok.
- U blokových funkcí se okolí nepřekrývá ale pokládá se na grid jako „dlaždice“
- Výsledná hodnota je přiřazena k bloku definovanému jako minimální pravoúhelník, do kterého se vejde ve funkci použité okolí

ZONÁLNÍ OPERACE

- Vedle jednoho či několika vstupních gridů je v zonálních operacích použit také grid, který definuje tzv. zóny - plochy se stejnou hodnotou gridu.
- Zóna je použita na místo okolí ve fokálních funkcích, jinak je princip stejný.
- Výstupní grid je funkcí hodnot buněk všech vstupních gridů a také jejich asociace s jinými buňkami dané zóny.
- Zónou může být kategorie landuse, parcela, ...

ZONÁLNÍ OPERACE



Základní skupiny zonálních operací

- **ZONÁLNÍ STATISTIKA** - základní průměrová statistika (průměr, sd, max, min, diverzita, modus, medián, ...)
- **ZONÁLNÍ GEOMETRIE** - počítají se geometrické atributy zóny v rámci vstupního gridu - plocha, obvod (perimetr), centroid
- **KONTINGENČNÍ TABULKY (CROSSTABULATION)** - matice ukazující možné kombinace buněk dvou vstupních gridů.

(Zóny prvního gridu tvoří řádky matice, zóny druhého vstupního gridu potom sloupce matice.
Např. - dvě mapy landuse z různého období - matice ukazuje, která kategorie se změnila na jinou kategorii, nebo - landuse a nadmořská výška - zastoupení tříd landuse v určitých výškových stupních).

Podle použité metody může být výstupem zonálních funkcí:

1. tabulka

2. Histogram (pro každou zónu)

Ekvivalenty v menu Spatial Analyst:

- Tabulate Areas
- Histogram by Zones

Příklady použití zonálních funkcí:

- Výpočet ploch jednotlivých kategorií landuse v rámci každého povodí (= zóny)
- Identifikování ohrožených (atd.) parcel (= zón) v rámci zpracovaného území
- Vyplnění lokálních depresí („bezodtokých oblastí“) ve spojitém povrchu (např. DEM)

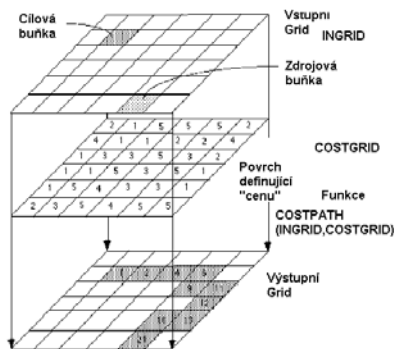
GLOBÁLNÍ OPERACE

Nová mapa vznikne transformací, jejíž rovnice byla sestavena na základě hodnot všech buněk rastru (např. polynomická transformace, definování obalových zón, Thiessenovy polygony, ...)

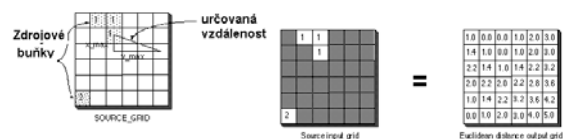
Oblasti využití

- Hledání optimální trasy
- Morfometrická analýza
- Hydrologické modelování

GLOBÁLNÍ OPERACE



Výpočet euklidovské vzdálenosti:



Na vstupu je grid, který obsahuje jednu nebo několik buněk nenulových - tzv. zdrojové buňky, od kterých se počítá např. vzdálenost.

Určování příslušnosti (alokace)

