

Prostorové analýzy v područí software

Jiří Horák

Institut geoinformatiky, HGF, VŠB-TU Ostrava, 17. listopadu 15,
708 33, Ostrava-Poruba, Česká republika
jiri.horak@vsb.cz

Abstrakt. Příspěvek se zamýšlí nad aplikací prostorových metod analýzy dat, který je stále více vázán a spojován s nabídkou programů. Pozitivní stránkou tohoto jevu je bezesporu dostupnost metod a jejich integrace s jinými nástroji (správa dat, prezentace dat, jiné analytické a modelovací funkce), za méně pozitivní lze považovat skutečnost, že tím dochází k selekci některých metod a jiné zůstávají pohřbeny v literatuře nebo jsou v lepším případě vázány na nepříliš známé a rozšířené systémy. Dalším problémem je košatost implementace těchto metod. Je uživatel schopen se dobře rozhodovat při nastavování a používání metod? Je komplexnost implementace metod vyvážena vhodnou podporou uživatele?

Klíčová slova: prostorové analýzy, software

Abstract. The paper deals with an application of spatial methods of data analysis, which is more bounded and linked to the SW offer. The positive aspects of this phenomenon are method's availability and integration with other tools (data management, data presentation, and other analytic and modelling functions). As a less positive reality we can understand the situation, when such process issues to selection of some methods and others stay burden in literature or they are included only in not known and spread systems. The next issue is the complexity and richness of method's implementation. Is a user able to make a good decision during setting and using of these methods? Is the complexity of method's implementation balanced by appropriate support of users?

Keywords: spatial analysis, software

1 Otázky místo úvodu

Většina myslivců, když si koupí novou pušku, zamíří nejdříve na stělnici a nejde rovnou na jelena. Musí si přece novou zbraň vyzkoušet, „nastřílet“, nastavit mířidla tak, aby výsledek v terénu byl optimální. V případě přesnosti střelby lze celkem dobře objektivně měřit, jak se daná zbraň chová.

Co však v případě software? Koupili jste nový SW? Vyzkoušeli ho? Ověřili jste si, že pracuje tak jak má? Proč si myslíte, že to autoři udělali zcela perfektně? Dostáváte snad na to garanci? Nebo jim věříte bezmezně? Čtli jste v licenčním ujednání, že uživatel souhlasí s tím, že výrobce nezodpovídá za žádné škody vzniklé užíváním SW?

2 Některé problémy spojené s užíváním software

Prostorové analýzy můžeme definovat jako soubor technik pro analýzu a modelování lokalizovaných objektů, kde výsledky analýz závisí na prostorovém uspořádání těchto objektů a jejich vlastností (Horák 2006). Takové velmi obecné vymezení ukazuje, že se jedná o velmi širokou sadu operací, které používáme v prostředí GIS k manipulaci s geodaty a které představují základní nástroje pro získání sekundárních (tj. odvozených) dat a informací potřebných v procesu rozhodování. Každý program podporující vytváření a realizaci GIS

obsahuje sadu funkcí, kterými je možné určitou množinu prostorových analýz provádět. Pochopitelně tato sada je nejmenší u jednoduchých programů zaměřených na prohlížení existujících geodat a mapových kompozic. Na druhou stranu se v poslední době silně **rozvíví vybavenost některých programů** analytickými a modelovacími funkcemi, objevují se specializované nízkonákladové programy se zajímavou nabídkou analytických operací. Těmi nízkonákladovými programy rozumíme běžně SW bez prodejního či udržovacího poplatku. Tato **zvýšená dostupnost** se samozřejmě odráží v rostoucím používání těchto funkcí a tak i zprostředkovaně k většímu využívání analytických operací, které do té doby nebyly běžným uživatelům známy a byly využívány často jen úzkou komunitou specialistů. Tento pozitivní jev však na druhou stranu vyvolává určité pochybnosti:

- **Jsou vždy vybírány a správně implementovány nejvhodnější funkce?**
- **Napomáhá daný programový produkt ke správnému používání (existence a propracovanost průvodců, helpu, jiné dokumentace a ukázek)?**
- **Je uživatel schopen správně použít implementované funkce?**

Stále více se objevují kritiky na stranu výrobců a uživatelů s tím, že jde o systémy typu black-box a uživatelé je slepě využívají.

S rostoucí nabídkou a dostupností také nutně vznikají otázky, v kterém programu potřebnou operaci provádět. Výrobci poskytují názvy funkcí a základní vysvětlení jejich činnosti, pro vysvětlení podstaty se zpravidla omezují na základní informace a odkaz na literaturu, podle které implementace proběhla a z pochopitelných důvodů nesdělují vysvětlení algoritmizace dané funkce. Pokud máme na výběr 2 obdobně pojmenované funkce, je jedno, kterou z nich použijeme? Liší se výsledky těchto operací? Nebo jsou rozdíly jen ve výkonu (např. v rychlosti provedení operace)? Na základě čeho si má uživatel vybrat? Má spoléhat na svého výrobce a dodavatele a použít funkci, která je v jeho standardně používaném prostředí k dispozici, a věřit, že rozdíly ve výsledku operace v jiném prostředí budou žádné nebo tak malé, že kvůli nim nestojí za to komplikovat zpracování výběrem jiného prostředí pro realizaci? Rostoucí interoperabilita snižuje přirozené bariéry mezi programy a prostředími a tak je přechod k jiné funkci či do jiného prostředí čím dál méně nákladný. Např. využití Web Processing Services (WPS) poskytuje možnost relativně snadno outsourcovat provedení příslušné operace mimo naše prostředí a využít specializované nabídky. Pokud se vrátíme k předchozí otázce, asi intuitivně tušíme, že rozdíly existují a že nikde není záruka, že by byly zanedbatelné. Výrobce samozřejmě musí volit kompromis mezi tím, co je snadno algoritmizovatelné, rychlé z hlediska provádění operace, dostupné a nejpřesnější. Situace se komplikuje tím, že jednotlivé metody nemusí mít stejné kvalitativní a kvantitativní ukazatele (přesnost, rychlost, spolehlivost atd.) pro různé předpokládané použití. Ve výsledku tedy nemusí být implementovaná funkce pro dané použití nejpřesnější, ale je třeba významně rychlejší.

Existuje řada článků, studií a studentských prací (taková porovnání patří k oblíbeným tématům studentských prací), které se zabývají srovnání 2 či více konkrétních operací realizovaných v různém prostředí či pro různé podmínky.

Přesto se domnívám, že existuje reálný deficit pro uživatele snadno dostupného srovnání a doporučení. Bohužel v tomto směru příliš nepomůže ani oblíbené hledání v Google či návštěva diskusních skupin.

Podívejme se na tyto problémy z hlediska jednoho typu operací – interpolace geodat.

3 Interpolace geodat

Interpolace je procedura odhadu neznámých hodnot ze známých (naměřených, zjištěných) hodnot v okolí. Zpravidla jde o tzv. lokální odhad, kdy odhadujeme hodnotu zájmové veličiny v bodě, kde nebyla primárně zjištěna nebo naměřena (Horák, 2008).

Interpolační procedury zahrnují řadu metod, kterými se provádí výpočet neznámé hodnoty např. metoda inverzních vzdáleností (IDW) a navazující Sheppardova metoda, polynomická regrese ve formě lokální polynomické interpolace, Fourierovy řady, klouzavé průměry, krigování (v celé řadě variant), radiální funkce a metoda minimální křivosti (využití splinových funkcí), metoda triangulace s lineární interpolací, ABOS (Approximation Based On Smoothing), metoda přirozeného souseda, bilineární interpolace, metody odhadu na základě Markovova náhodného pole, Bayesovo neparametrické vyhlazování. Někteří autoři (a programy) mezi interpolační metody řadí i tvorbu Thiessenových polygonů a trendové funkce. Je zřejmé, že se nabízí pestrá škála metod, navíc doprovázená řadou podtypů a možností nastavení parametrů, které ovlivňují výpočet např. řešení hraničního problému (hraniční korekce v Rockworks).

Není cílem tohoto článku vysvětlovat podstatu jednotlivých metod; jejich principy jsou vysvětleny např. v Hlásny (2007), Horák (2008), Cressie (1993), Burrough, McDonell (1998).

Jak si má uživatel vybrat vhodnou metodu?

V případě interpolace existují jistá **obecná doporučení**. Např. se uvádí, že kvalita interpolačních výsledků závisí na počtu a rozmístění známých bodů použitých pro výpočet a na tom, jak matematická funkce korektně modeluje sledovaný fenomén (u deterministických metod). Nejlepších výsledků se dosahuje, když se matematická funkce chová podobně jako sledovaný fenomén. Např. teploty vzduchu nad plochou oblastí budou vykazovat postupné změny, proto je žádoucí použít algoritmus, který vytváří spojité pole. Avšak pro výšková data v horském reliéfu by byl takový algoritmus nevhodný, protože by vedl k ořezání příkrých vrcholů a hran, tedy rysů, které by měly zůstat zachovány. Podobným obecným pravidlem je výběr metod podle přesnosti dodržení známých hodnot (rozdělené interpolačních metod na

exaktní a aproximační) či otázka, zda připouštíme ve výsledku vyšší a nižší hodnoty než je rozsah měření.

Vedle obecných doporučení existují i nástroje pro **hodnocení kvality výsledku**. V případě interpolace se používá:

- bumerangová metoda (crossvalidation, křížová validace)
- validace s kontrolním vzorkem
- validace s referenční funkcí (ta reprezentuje známý vývoj v ploše)

Ne všechny systémy je implementují, a pokud ano, jejich použití není jednoduché.

Křížová validace je již dnes implementována v řadě systémů, což je důležitým krokem vpřed, jak vyplývá např. ze srovnání interpolačních metod pro výšková data (Hofierka et al. 2005). Nicméně ani její použití není všelékem; není vhodná v případech malého počtu známých bodů a obecně používání několika málo statistických charakteristik (především střední chyby RMSE) pro celý soubor není vždy dostačující (viz příloha 1, kde jsou zobrazeny ukázky výsledku interpolace jednotlivými metodami pro vybranou situaci).

Je třeba upozornit, že porovnání výsledků v rámci 1 SW s využitím křížové validace nestačí a že je vhodné ověření, zda by parametry výsledku interpolace nebyly lepší v jiném prostředí, s jinými implementovanými procedurami. Je nutné v rámci rostoucí interoperability překonat bariéry mezi jednotlivými programy a provádět hodnocení výsledků napříč.

Bez nadsázky lze říci, že nejvíce teoreticky rozpracovanou interpolační metodou je krigování.

Krigování je základní geostatistickou metodou určování lokálního odhadu, která vychází z představy regionalizovaných proměnných, kde hodnota proměnné je funkcí polohy, a využití teorie náhodných funkcí. Tuto metodu představil G. Matheron v roce 1962 a od té doby se mohutně rozvíjí. Na několika příkladech je možné dokumentovat tento rozvoj.

A.G. Journal a CH. J. Huijbregths (1978) popisuje základní krigování, univerzální krigování, krigování globálního průměru, kokriging a základní nelineární techniky (disjunktivní krigování). Zde se rovněž již hovoří o nutnosti snížení velikosti řešení (soustavy rovnic) tím, že se zavádí jisté okolí. Rovněž se zde již hovoří o simulacích jako alternativě ke krigování a vysvětluje se princip podmíněné stochastické simulace, metody rotujících pásem, Gaussovské transformace a to včetně programového kódu v jazyce Fortran (program SIMUL pro výpočty sítě bodů ve 3D!).

A.G. Journal (1989) popisuje ve stručném úvodu do geostatistiky jednoduché (definované ale již Matheronem 1971), základní a univerzální krigování, kokriging, indikátorové krigování, pravděpodobnostní krigování, soft kriging a stochastické simulace.

Cressie (1993) uvádí jednoduché, základní, univerzální, lognormální a trans-Gaussovské krigování, kokriging, Bayesovo krigování, mediánem vyhlazované krigování (median-polish kriging) a simulační techniky, nelineární techniky (disjunktivní krigování), indikátorové

krigování, pravděpodobnostní krigování. Rozlišuje (stejně jako další autoři) varianty pro lokální a blokový odhad.

Burrough, McDonell (1998) se geostatistikou a krigováním zabývali jen částečně, nicméně kromě výše uvedených doporučují rovněž stratifikované krigování a vylepšený postup pro stochastickou podmíněnou simulaci.

Všimněme si u tohoto přehledu, kolik existuje různých metod (které mohou být navíc různě implementovány) a jak je většina z nich již stará. Např. stochastická simulace před rokem 1978, přesto se dosud minimálně rozšířila v praxi.

Srovnejme si, co v této oblasti nabízely či nabízejí konkrétní programy, a jak se nabídka vyvíjela. Bez nároku na úplný výčet uvedme snad nejběžnější programy, které se pro interpolace geodat používají.

V ArcView byly k dispozici pouze 2 základní interpolační metody (metoda inverzních vzdáleností a minimální křivosti) a krigování po doplnění extenze, což odpovídá nabídce jiných desktopových programů pro GIS. ArcGIS Spatial Analyst ve verzi 9.2 nabízí metodu inverzních vzdáleností, spline (regularizovaný, pod napětím) a krigování (základní a univerzální s možností nastavení nejvýznamnějších ukazatelů). Přesto specializované programy (např. Surfer, Isatis, GSTAT, modul Geostatistical Analyst) nabízejí zcela jiný rozsah interpolačních služeb ať již jde o metody nebo nastavování parametrů a kontrolu vstupních dat i výsledků zpracování, proto se budeme zabývat pouze jimi.

Geostatistical Analyst je implementován v ArcInfo od verze 8.0.2 v roce 2000. Jeho vývoj je spojen s příchodem Konstantina Krivoruchka do ESRI v roce 1998, když předtím pracoval jako ředitel GIS laboratoře na Sacharovově institutu radioekologie v Minsku a publikoval řadu prací o důsledcích Černobylské jaderné katastrofy.

Geostatistical Analyst již záhy nabízel poměrně pestrou skupinu interpolačních metod. Verze 9.2 poskytuje metodu inverzních vzdáleností, globální polynomická interpolace (popis trendu), lokální polynomická interpolace, radiální funkce (5 typů funkcí – thin-plate spline, spline pod napětím, úplně regularizovaný spline, multikvadratická funkce, inverzní multikvadratický spline) a krigování (jednoduché, základní a univerzální krigování, jednoduché nebo základní krigování s externím trendem, kokriging, indikátorové krigování, pravděpodobnostní krigování, disjunktivní krigování).

ArcGIS v. 9.3 má nabízet v nadstavbě Geostatistical Analyst nový nástroj – Gaussian Geostatistical Simulations. Zřejmě dovolí realizaci jak podmíněné tak i nepodmíněné stochastické simulace a vedle jednotlivých náhodných realizací pole zpracovat také souhrnné statistiky. K dispozici má být i tutorial (Jiravová, Dufek 2008).

GSTAT (vyvíjen 1996-2005) je volně šiřitelný program (www.gstat.org), který umožňuje studovat pole a provádět interpolaci ve 2D i 3D (Pebesma and Cees G. Wesseling, 1998). Nabízí následující interpolační metody - jednoduché krigování, základní krigování, univerzální krigování, stratifikované univerzální krigování, jednoduché kokrigování, základní kokrigování, univerzální kokrigování, standardizované kokrigování, krigování s externím driftem, blokové krigování, krigování trendu, nekorelovaný regresní odhad (trendu), monovariační i multivariační podmíněná i nepodmíněná multi-Gaussovská sekvenční simulace bodových hodnot nebo průměrů bloků a indikátorové nebo multiindikátorové sekvenční simulace.

Surfer 8 (Golden Software, <http://www.goldensoftware.com/products/surfer/surfer.shtml>) poskytuje interpolační metody: metoda inverzních vzdáleností, krigování (základní a univerzální), minimální křivosti, triangulace s lineární interpolací, Sheppardova metoda, radiální funkce a polynomická regrese (pro popis trendu), nejbližšího souseda (varianta Thiessenových polygonů), metodu přirozeného souseda, klouzavé průměry.

GRASS 6.3.0 (<http://grass.itc.it/>) obsahuje interpolační metody inverzních vzdáleností, lokální polynomická regrese, krigování a regularizovaný spline pod napětím (včetně křížové validace), rovněž i určení trendové plochy.

Tento neúplný přehled dokumentuje situaci, kdy jednoduše implementovatelné algoritmy jsou k dispozici u všech produktů (typicky metoda inverzních vzdáleností), více sofistikované metody jen u některých, často volně šiřitelných a vznikajících v univerzitním prostředí. I takový poměrně sofistikovaný nástroj jako je Geostatistical Analyst nemá např. interpolaci ve 3D.

Jak vyplývá z přehledu literatury, většina těchto metod byla popsána již poměrně dávno, některé však na svou implementaci v rozšířených programech teprve čekají.

Jaký to má vliv na uživatele? Které metody se skutečně používají, tedy které prostorové analýzy se v rámci interpolací používají?

Pro srovnání byla provedena rešerše v časopisech Geografický časopis a Geografie-Sborník ČGS. Bohužel mnoho příkladů zde nebylo nalezeno.

V Geografickém časopise (SAV) byl v roce 2006 publikován příspěvek, kde bylo přímo využito extenze Geostatistical Analyst pro interpolaci. Siska, P., Bryant Jr., V., Hung, I. (2006) podle velikosti střední chyby a s odkazem na doporučení ke Grassu použili úplný regularizovaný spline. V roce 2005 byl publikován příspěvek Hlásneho ke geostatistickému konceptu prostorové závislosti (Hlásny 2005) a využití při interpolaci, ale v předcházejících

ročníkách se tomuto problému příspěvky téměř nevěnovaly, výjimku tvoří roky 1997 a 1992 po 1 příspěvku. U některých příspěvků v jiných letech (v průměru ale přibližně 1 příspěvek ročně) byla nějaká metoda interpolace využita, nicméně autoři nepovažovali za potřebné ji dokumentovat.

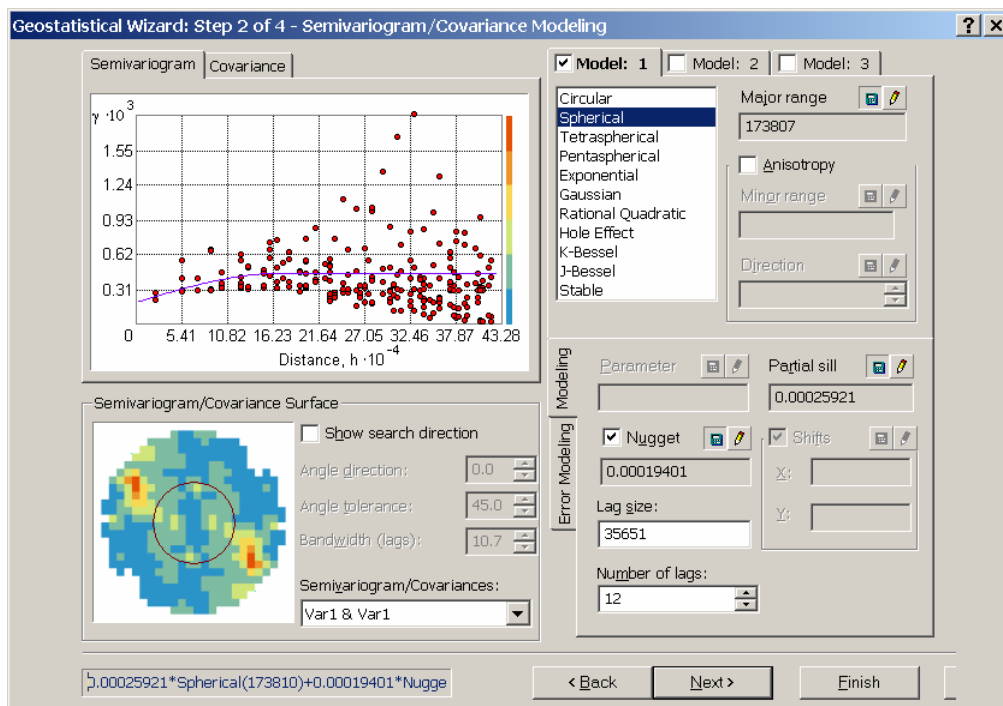
Geografie-Sborník ČGS byl studován jen pro období let 2004-2006. V 1 příspěvku v roce 2006 byla provedena interpolace DTM z vrstevnic v prostředí Grass. V roce 2005 bylo 3. číslo věnováno problematice českých jezer a prováděná batymetrická měření byla ve všech případech zpracována v programu Surfer s využitím krigování, avšak bez bližšího popisu, proč byla tato metoda vybrána.

Na těchto příkladech lze doložit, že se problematice správné volby interpolačních metod a nastavení parametrů věnuje ojediněle samostatný článek, avšak ve většině dalších článků není problému interpolace věnována dostatečná pozornost.

Je to opravdu tak důležité? Nestačí použít výchozí (default) nastavení u metody?

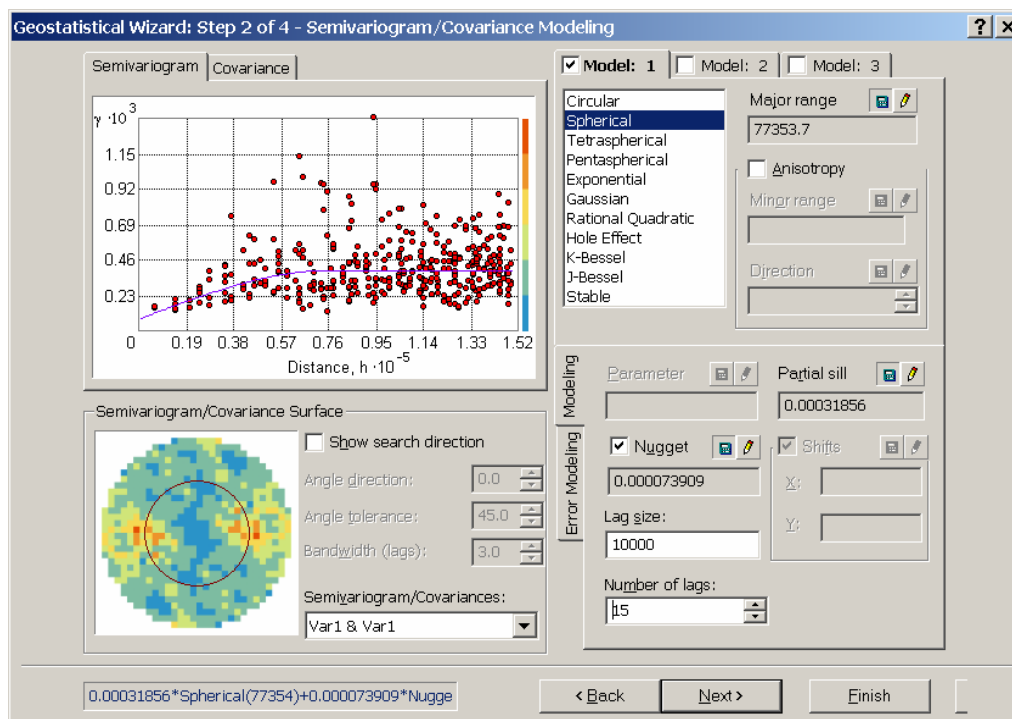
Opět 1 příklad, který ukazuje, jak může být výchozí nastavení nevhodné.

Standardní nastavení u ArcGIS Geostatistical Analyst při modelování strukturální funkce (před krigováním) určí na základě velikosti oblasti automaticky velikost základního kroku a počet kroků pro výpočet strukturální funkce. Na obr. 1 je vidět, že je základní krok nastaven na bezmála 36 km a počet kroků je 12. Některé známé hodnoty jsou však od sebe vzdáleny méně než 4 km. Při studiu pole jsou přitom pro nás nejdůležitější ty krátké vzdálenosti, kde hlavně studujeme vývoj pole. Zde se také určují významné parametry a následně odvozují váhy pro krigování.



Obr.1. Výchozí nastavení Geostatistical Wizard pro standardní dodávaný ukázkový příklad (měření ozónu v oblasti Kalifornie)

Daleko vhodnější je nastavit základní krok 10000 a počet kroků 15 (obr. 2).



Obr.2. Upravené výhodnější nastavení Geostatistical Wizard pro daný příklad

Z porovnání obrázků je patrné, jak se změnila hodnota zbytkového rozptylu a dosahu (charakterizující průběh prokládané modré křivky). V obr. 1 je zbytkový rozptyl 0,00019401 a

dosah 173, 8 km, kdežto na obrázku 2 již 0.000073909 (tedy 2,6 krát méně), dosah 77,4 km (tedy 2,2 krát méně). To jsou z hlediska interpretace zásadní rozdíly, zejména v případě dosahu, který charakterizuje, do jaké vzdálenosti se hodnoty vzájemně ovlivňují je rozdíl 77 nebo 174 km markantní.

Takováto změna nastavení však není uživateli nijak doporučena (není o něm zmínka ani v manuálech, pouze obecně, že je možné to změnit, ale ne již proč a co má sledovat).

Můžeme být s nabídkou dostupnou v SW spokojeni? Můžeme spoléhat na to, že to co je nám předkládáno, je to nejvhodnější pro náš případ použití?

Je zřejmé, že tyto problémy nejsou specifickou záležitostí geoinformatiky, ale že je třeba vycházet ze zkušeností s testováním SW obecně.

4 Obecně o problému kvality SW

American Society for Quality (ASQ) (<http://www.asq.org/>), která se zabývá certifikací kvality SW, uvádí 7 hlavních oblastí, které by měl popis kvality SW zahrnovat (http://en.wikipedia.org/wiki/Software_quality):

- General [quality] knowledge
- Software quality management
- Systems and software engineering processes
- Project management
- Software metrics and analysis
- Software verification and validation (V&V)
- Software configuration management

Kde v tomto schématu hledat sledování, měření a porovnávání kvality provádění operací?

Spolehlivost SW (SW reliability) se soustřeďuje na odolnost proti chybám a ovlivnění prostředím při běhu programu. Definice podle <http://www.businessdictionary.com/definition/software-reliability.html> říká, že spolehlivost SW je schopnost programu poskytovat zamýšlené funkce a operace v systémovém prostředí bez výskytu chyb či selhání systému.

Verifikace a validace programu je procesem kontroly, zda systém odpovídá zadaným požadavkům a zda vyhovuje zamýšlenému účelu. Verifikace se považuje za nižší úroveň kontroly, kde se sleduje pouze, zda byl výsledný produkt vytvořen podle zadání (tj. požadavků a připraveného návrhu, konceptu), tedy zda odpovídá původnímu plánu. Naproti tomu validace je procesem vyšší kontroly, protože ověřuje, zda produkt splňuje skutečné potřeby. Definice (Department of Defense Documentation of Verification, Validation &

Accreditation for Models and Simulations, *Missile Defense Agency, 2008*) říká, že validace je procesem, který určuje stupeň, do kterého model, simulace (nebo skupina modelů či simulací) a jejich asociovaná data jsou přesnou reprezentací reálného světa z hlediska zamýšleného použití.

Je zřejmé, že právě validace je klíčem k ověření, měření a popisu výsledku prováděných operací.

Validace i verifikace se ověřují testováním SW. Pestrá škála testovacích metod je uvedena na http://en.wikipedia.org/wiki/Functional_Testing, jako nejvhodnější se však jeví použití testovacích sad a příkladů k ověření, zda je výsledek uspokojivý (test cases).

Projekt BEST-GIS (Best Practice in Software Engineering and Methodologies for Developing GIS applications) vedený GISIG (<http://www.gisig.it/best-gis/>) se zaměřil na přípravu doporučení pro vývoj aplikací a příslušného uživatelského rozhraní s důrazem na aplikaci user-center design paradigmatu. V připravených návodech a doporučeních lze mimo jiné najít i několik doporučení, jak by SW měl provádět vybrané klíčové GIS operace, resp. jak by je měli uživatelé používat. Např. pro operaci buffering (tvorba obalové zóny) upozorňují, že uživatel musí dobře porozumět fyzikálnímu významu prováděné operace (např. prostředí musí být izotropní), aby získal smysluplný výsledek. Nebo pro tvorbu rastrového DTM z digitalizovaných vrstevnic doporučují, aby velikost pixelu (buňky rastru) byla menší než 1/3 vzdálenosti vrstevnic a teprve následně provádět převzorkování příliš jemného rastru do menšího rozlišení. Ani tento projekt však nesměřoval k posouzení výkonu a výsledku jednotlivých funkcí.

Americký program SAVER (System Assessment and Validation for Emergency Responders) založený U.S. Department of Homeland Security se zaměřuje mimo jiné na hodnocení a validaci výkonu produktů uvnitř systému, v rámci něhož bylo také navrženo a provedeno hodnocení GIS SW. Vybrané programy byly hodnoceny z hlediska 17 základních funkcí od atributového dotazování až po export vytvořené mapy do PDF s důrazem především na zátěžové testy, chování v síti a přístup k uživateli (<https://saver.fema.gov/actions/document.act.aspx?type=file&source=view&actionCode=submit&id=5173&file=SP-SPAWAR-2006-19-SMRY.pdf>).

Pro obecné testování škálovatelnosti a výkonu SW již existují určité datové sady. Problém benchmarkingu u SW je však zatím relativně málo studován a není zatím mnoho dokumentovaných indikátorů (podle <http://www.ijswis.org/cfp/Special-Issue-Scalability-cfp.htm>). Studují se rovněž metody syntetického vytváření takových testovacích datových sad.

5 Doporučení

Zdá se, že vývoj spěje do situace, kdy uživatel již naléhavě potřebuje vhodnější podporu svého rozhodování při volbě SW a metod prostorových analýz. Dokumentace je buď nedostatečná nebo naopak dostatečně podrobná, avšak příliš náročná pro většinu uživatelů. Uživatel by měl mít možnost, aby si sám ověřil vhodnost jednotlivých metod implementovaných v SW a také ověřil kvalitu jejich implementace (především chybovost a výkon). Řešením by mohly být vhodné testovací příklady (test cases), umístěné na volně dostupném serveru. Testovací příklady, určené pro jednotlivé prostorové operace, by měly obsahovat testovací data společně s popisem optimálního výsledku, který by měl na základě zkoumané operace vzniknout.

Testovací data mohou zahrnovat jak uměle připravené datové sady (např. kontinuální povrch definovaný matematickou funkcí, na kterém lze určit správný výsledek dané operace) nebo reálné datové sady, dokumentující konkrétní oblasti, konfigurace a problémy.

Uživatel by si vybral problém a tematickou oblast, která je nejbližší jeho požadavkům, stáhnul by si potřebná testovací data a provedl operaci v daném SW (případně opakoval a optimalizoval parametry, aby získal nejlepší, tedy nejpodobnější, výsledek) a získaný výsledek by porovnal s publikovaným optimálním výsledkem.

Zřejmě výběr jednotlivých tematických oblastí a jejich pokrytí testovacími příklady nebude vše postihující, ale vlastní testování by mělo napovědět, jak se daný systém chová v základních typových situacích (v případě interpolace např. situace typu vrstevnice, situace typu rovnoměrná síť, velmi nerovnoměrná síť, nehomogenity v území, s trendem atd.).

Tím by si uživatel ověřil schopnosti daného SW řešit situace, pro které se plánuje jeho využití.

Testovací data mohou být k dispozici v několika formátech, aby byly minimalizovány obtíže spojené s načítáním dat do jednotlivých programů.

Řada komerčních producentů dnes nabízí možnost vyzkoušení daného programu v časově omezené instalaci, což by pro sledovaný účel plně vyhovovalo.

Pro zajištění nezávislosti použitých příkladů by bylo vhodné ponechat systém otevřený jak pro doplňování dalších testovacích příkladů, které tam může vystavit každý přihlášený a ověřený uživatel, tak i ratingu a komentářů k daným příkladům. Právem administrátora by pak bylo vadné příklady odstranit.

6 Závěr

Tento příspěvek si klade za cíl vyvolat diskusi na téma validace implementace prostorových analýz v programech, rozšíření jednotlivých analýz, vhodnost jejich použití a spojené problémy.

Je nutné apelovat na výrobce a distributory, aby dodávaly k programům co nejpodrobnější dokumentaci a implementovali sofistikované průvodce, které zajistí, aby při využití náročnějších metod nedocházelo k zásadním chybám.

Je nutné apelovat na uživatele, aby kriticky přistupovali k výsledkům počítačové zpracování. Je třeba výsledky vždy ověřit a zajistit alespoň základní logickou kontrolu, aby se vyhnuli hrubým chybám.

Je nutné se zamyslet nad možností vybudovat vhodný systém na podporu uživatelů, který by mimo jiné mohl obsahovat testovací příklady, určené pro nezávislé otestování programu pro sledovaný účel.

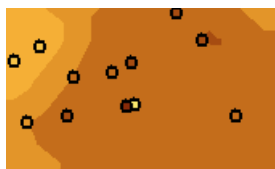
Dostupnost jednotlivých metod a programů bude dále narůstat, mimo jiné i v důsledku rozšiřování webových služeb (především WPS). Adekvátně tomuto vývoji by se měly vyvíjet i kontrolní a hodnotící mechanismy.

7 Literatura

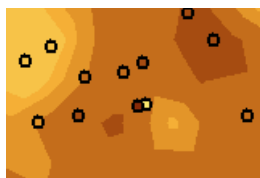
1. Burrough P.A., McDonell R.A. (1998): Principles of Geographical Information Systems. New York. Oxford University Press. ISBN 0-19--823365-5.
2. Cressie N. (1993): Statistics for Spatial Data. Wiley, 900 stran, ISBN 0-471-00255-0.
3. Hlásny, T.(2005) Geoštatistický koncept priestorovej závislosti pre geografické aplikácie. Geografický časopis. Vol. 57, no. 2, p. 97-116.
4. Hofierka J., Cebecauer T., Šúri M. (2005). Optimisation of Interpolation Parameters Using a Cross-validation. In Peckham R., Jordan G. (eds.) Best Practice in Digital Terrain Modelling: Development and Applications in a Policy Support Environment. European Commission Joint Research Centre, Ispra (VA).
5. Horák J. (2006): Prostorová analýza dat. Skripta VŠB-TUO. 149 stran. http://gislinb.vsb.cz/~hor10/PAD/Skripta/PAD_skriptaF.pdf.
6. Horák J. (2008): Zpracování dat v GIS. Skripta VŠB-TUO. 188 stran. <http://gislinb.vsb.cz/~hor10/ZpracovaniDatvGIS/GISANA21.pdf>.
7. Jiravová, Dufek (2008): ArcRevue 2/2008, 17.ročník, Praha, ISSN 1211-2135, str. 22-24.
8. Journal, A.G, Huijbregts P. (1978): Mining geostatistics. Academic Press, 600 stran.
9. Journal, A.G. (1989): Fundamentals of Geostatistics in five lessons. American Geophysical Union. 38 stran.
10. Pebesma E., Wesseling, C. (1998), Gstat: a program for geostatistical modelling, prediction and simulation. Computers & Geosciences Vol. 24, No. 1, pp. 17-31
11. Siska, P. Bryant Jr., V., Hung, I. (2006): Geospatial analysis of southern pine biome and pollen distribution patterns in Southeastern United States. Geografický časopis. Vol. 58, no. 4, p. 239-258.

8 Příloha 1

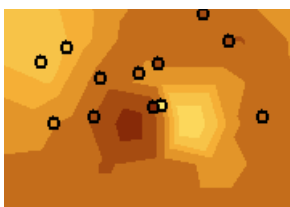
Ukázky výsledku interpolace jednotlivými metodami pro vybranou situaci 2 známých, avšak silně odlišných hodnot v těsné blízkosti. Příklad ukazuje standardní cvičná data měření ozónu v Kalifornii, kroužky reprezentují jednotlivá známá měření, tmavší odstín odpovídá vyšší hodnotě. Silně odlišné hodnoty jsou zhruba uprostřed obrázku. Stínované plochy odpovídají interpolovaným hodnotám.



Inverzní vzdálenosti (váha 2)



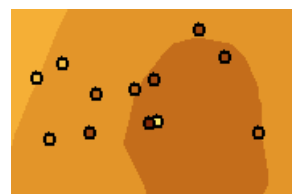
Inverzní vzdálenosti (váha 3.9 + nastavená anizotropie)



Radiální funkce – spline pod napětím



Radiální funkce – multikvadratická metoda



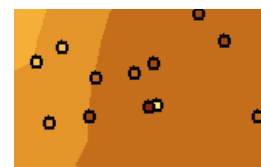
Lokální polynomy



Jednoduché krigování



Základní krigování



Základní krigování s externím trendem 2.řádu