



Česká zemědělská univerzita v Praze

**Fakulta životního
prostředí**

Program Gorint

Dokumentace
Uživatelská příručka

Petr Bašta

Praha, 2011

Tato dokumentace včetně programového vybavení byla vypracována
v rámci grantového projektu

*Návrh koncepce řešení krizové situace vyvolané výskytem sucha a nedostatkem vody
na území České republiky*

řešeného v rámci

**Programu bezpečnostního výzkumu České republiky v letech 2010–2015
(BV II/2-VS)**

Hlavní řešitel za Výzkumný ústav vodohospodářský T. G. M. v.v.i.:
Ing. Ladislav Kašpárek, CSc.

Hlavní řešitel za Českou zemědělskou univerzitu v Praze:
prof. Ing. Pavel Pech, CSc.

Obsah

1	Úvod	2
2	Teoretický základ	2
2.1	Úvod do problematiky	2
2.2	Prostorové interpolace	3
2.3	Geostatistické interpolace	4
2.3.1	Semivariogram	4
2.3.2	Kriging	6
2.3.3	Simple kriging	7
2.4	Vážené průměrování metodou inverzních vzdáleností (IDW)	8
3	Instalace programu	9
4	Použití programu	10
4.1	Vstupní data	10
4.2	Přehled parametrů	10
4.3	Přehled hlavních funkcí, jejich argumentů a způsobu zadávání	13
5	Případová studie – ukázka aplikace	15
5.1	Vstupní kód programu	15
5.2	Výstupy programu	16

1 Úvod

Program slouží jako **Gridově ORientovaný INterpolátor**, který umožňuje prostorové interpolace povrchů (např. nadmořské výšky) ze zadaných bodů obsahujících informaci o měřené hodnotě analyzované veličiny. Takovou veličinou může být kromě nadmořské výšky např. koncentrace těžkých kovů v půdě, půdní hydraulická vodivost, úhrn srážek, výška sněhové pokrývky nebo jakákoli další prostorově orientovaná veličina. Program nabízí geostatistické metody interpolace – ordinary kriging a universal kriging – a deterministickou metodu váženého průměrování pomocí inverzních vzdáleností, IDW. Výstupem programu **Gorint** je rastr (tvořený ortogonálním gridem), který je možno vizualizovat nebo dále zpracovávat v prostředí GIS.

2 Teoretický základ

2.1 Úvod do problematiky

Digitální model terénu (DTM) představuje numerickou reprezentaci terénu, avšak lze jej definovat různými způsoby. [Miller and LaFlamme \(1958\)](#) jej definují jako statistickou reprezentaci spojitého povrchu země prostřednictvím velkého počtu vybraných bodů se známými souřadnicemi X , Y , Z v libovolné souřadnicové soustavě. Hodnoty DTM jsou potom funkcí veličiny, kterou daný model popisuje [El-Sheimy et al. \(2005\)](#). [Moore et al. \(1991\)](#) charakterizují DTM jako uspořádané pole číselných hodnot, které kvantifikují sledovanou charakteristiku terénu v libovolných bodech geografického povrchu. DTM lze charakterizovat také jako matematický (nebo digitální) model, v němž jedna či více matematických funkcí reprezentuje povrch terénu na základě měřených bodových dat; tyto matematické funkce značíme jako funkce interpolační ([Li et al., 2005](#)). Sledovaná veličina – např. výška z – je funkcí polohy. Z pojmu digitální model terénu se vyčleňuje termín digitální elevační model (DEM). Jde o speciální případ DTM, kdy číselné hodnoty představují výšky daných bodů ([Moore et al., 1991](#)). Výška je funkcí geografické polohy ([O’Callaghan and Mark, 1984](#)). DEM je základem pro veškeré analýzy topografických charakteristik na povodí. V souvislosti s DTM a DEM je dále zmiňován také termín digitální analýza terénu (DTA), jež zahrnuje veškeré procesy, které kvantitativně popisují terén prostřednictvím DTM či DEM ([Hengl et al., 2003](#)).

Prostřednictvím digitální analýzy terénu závisí na DTM mnoho aplikací na bázi geografických informačních systémů (GIS), např. analýzy půdy, intenzity slunečního záření, mapování hloubky podzemní vody či mocnosti sněhové pokrývky a samozřejmě hydrologické modelování ([Moore et al., 1991](#)). Výsledky těchto aplikací jsou tzv. charakteristiky (atributy) terénu. Mezi primární terénní charakteristiky patří především přímé deriváty nadmořské výšky (sklonitost terénu, křivost, expozice apod.), mezi sekundární řadíme např. albedo.

K reprezentaci povrchu terénu se běžně užívají tři metody: isolinie, trojúhelníková nepravidelná síť (TIN) a nejčastěji využívané gridové struktury. Grid představuje maticovou strukturu, která implicitně definuje topologické vztahy mezi datovými body ([Fairfield and Leymarie, 1991](#)). Pravidelný grid je složen z dvourozměrné matice hodnot výšek v bodech, jejichž vzdálenosti mezi sebou jsou v obou směrech x a y konstantní ([El-Sheimy et al., 2005](#)). Kvalita a přesnost takového modelu je silně závislá na hustotě bodů; z hlediska snadné algoritmizace, příznivé doby trvání výpočtů a snadné implementace pro aplikace environmentálního modelování je však jeho využití výhodné ([Collins and Moon, 1981](#)).

Kvalita DTM závisí především na dvou faktorech: (1) na kvalitě měřených dat vstupujících do modelu a (2) na použité metodě tvorby DTM z těchto dat ([Hengl et al., 2003](#)). Kvalitu mě-

řených dat lze ovlivnit výběrem metody měření; optimální metoda se volí dle povahy modelu a účelu jeho aplikace (Moore et al., 1991; Heritage et al., 2009; Chaplot et al., 2006). Zatímco pro nejpodrobnější mapování velkých měřítek se využívá laser scan, pro běžné modely terénu postačí tachymetrická měření laserovými teodolity či referenčními GPS s centimetrovou přesností a pro mapování rozsáhlých ploch malých měřítek se využívá stereoskopická fotogrammetrie, satelitní snímky nebo digitalizace vrstevnicových map (Chaplot et al., 2006; Hengl et al., 2003). Od zvolené metody měření se odvíjí hustota zaznamenaných bodů modelu, na čemž plně závisí další aplikace – např. nízká hustota bodů DTM může vést k nadhodnocení sekundárních terénních charakteristik (Quinn et al., 1991) a podhodnocení gradientu sklonitosti terénu (Wilson and Gallant, 2000; Fairfield and Leymarie, 1991). Další, neméně významný vliv na výsledný model terénu, má také prostorová struktura vstupních referenčních bodů (Chaplot et al., 2006; Heritage et al., 2009).

Stejně tak záleží na způsobu tvorby DTM. Tvorba DTM je proces predikce hodnot sledované veličiny v predikovaném bodě z hodnot změřených v referenčních (vzorových) bodech situovaných ve specifikovaném okolí predikovaného bodu; tento proces je označován jako interpolace (El-Sheimy et al., 2005). Existuje řada interpolačních technik. Mnoho dosavadních studií se zabývá srovnáváním různých technik a výběrem té optimální. Někteří autoři – např. Burrough and McDonnell (1998) a Wilson and Gallant (2000) ukazují, že mezi existujícími interpolačními technikami podávají lepší výsledky geostatistické metody (kriging). Totéž dokazují ve své studii rovněž Sarangi et al. (2006), kteří srovnávají geostatistické metody s metodami deterministickými (IDW, spline) pro účely odhadu prostorové variability odtokové odezvy. Jiné práce ale dokazují, že naopak deterministické metody poskytují lepší výsledky, např. Mitášová et al. (2005).

2.2 Prostorové interpolace

Sledovaná veličina – např. výška z – je funkcí polohy, $z = f(x, y)$ (Li et al., 2005).

Za prostorové interpolace je označována skupina metod, které slouží k vytváření spojitých povrchů (polí) z bodových měření. Naprostá většina interpolačních postupů je založena na principu prostorové autokorelace – tedy na předpokladu, že hodnoty predikované veličiny v lokalitách blízkých si budou více podobné než hodnoty v lokalitách vzdálených (Ježek and kolektiv, 2008).

Předpoklady úspěšné prostorové interpolace jsou dle Ježek and kolektiv (2008): (i) existence dostatečně reprezentativního vzorku měřených dat, (ii) teoretické i empirické znalosti o povaze prostorové diferenciace studovaného jevu, (iii) vhodné vlastnosti měřené veličiny a vhodný typ dat (ordinální, intervalová, poměrová), (iv) znalost podstaty použitelných interpolačních metod, (v) znalost způsobu výběru nejvhodnější metody.

V případě kvalitních vstupních dat dává většina interpolačních technik podobné výsledky (Burrough and McDonnell, 1998).

Deterministické metody lze využít v případech, kdy existuje dostatek informací o prostorovém chování studovaného jevu, takže je možné tento jev popsat matematickou funkcí. Tyto metody umožňují i extrapolaci za hranice vymezené měřenými body. Naproti tomu stochastické modely zahrnují koncept náhodnosti za předpokladu, že hodnoty interpolovaného povrchu z daného měřeného vzorku jsou jen jednou z nekonečného množství možných variant. Do skupiny těchto metod patří např. metody krigingu či analýza trendu (Ježek and kolektiv, 2008; El-Sheimy et al., 2005).

2.3 Geostatistické interpolace

Geostatistickou metodu kriging vystavěl francouzský matematik Georges Matheron na te-
zích Daniela Gerharduse Kriega, po němž ji také roku 1962 nazval. Kriging je založen na
skutečnosti, že interpolovaný povrch lze lépe vystihnout stochastickou funkcí než shlazující
matematickou funkcí. Jde o metodu geostatistickou.

Obecným konceptem této metody je prostorová závislost, autokorelace jevů. Vztaženo ke
geografickému povrchu, míra podobnosti výšek dvou bodů klesá s rostoucí vzdáleností těchto
bodů. Prostorová korelovanost veličiny je charakterizována varigramem, který se stanoví na
základě naměřených dat a předpokladů o chování veličiny. Kriging pak poskytuje nejlepší
neustranný lineární odhad interpolované veličiny (Best Unbiased Linear Estimator = BLUE)
(Goovaerts, 1997). Je vhodný pro interpolování proměnných, které se v prostoru mění s jistou
kontinuitou, ale nelze je popsat jednoduchou shlazující funkcí některého z globálních interpo-
látorů (Ježek and kolektiv, 2008).

2.3.1 Semivariogram

Experimentální semivariogram $\gamma(\mathbf{h})$ determinuje průměrnou odlišnost mezi páry bodů vzdá-
lených o vektor \mathbf{h} . Konkrétně jde o strukturální funkci, která popisuje závislost průměrné
kvadratické difference hodnot prostorové proměnné veličiny Z na vzdálenosti h . Hodnoty se-
mivariogramu jsou definovány jako polovina rozptylu mezi hodnotami z každého páru bodů
(Goovaerts, 1997):

$$\gamma(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} [z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})]^2 \quad (1)$$

Jestliže bude velikost posunu \mathbf{h} malá, lze očekávat vysokou míru podobnosti mezi oběma
body v každém páru a naopak.

Platí-li stacionarita modelu, (tj. neměnná střední hodnota s posunem o vektor \mathbf{h}), vztah
mezi semivariogramem a kovariancí lze vyjádřit takto (Goovaerts, 1997):

$$\gamma(\mathbf{h}) = C(0) - C(\mathbf{h}) \quad (2)$$

pokud kovariance existuje.

Průběh semivariogramu obvykle začíná z počátku souřadné soustavy a stoupá až k prahové
hodnotě, za kterou už zůstává neměnný. K detailnímu popisu semivariogramu se rozlišují
následující prvky:

- Lag \mathbf{h} – vzdálenost h mezi jednotlivými páry bodů (vektor posunutí \mathbf{h}); vynáší se na
horizontální osu, hodnoty rozptylu (resp. semivariance) se vynáší na vertikální osu;
- Sill c neboli práh – maximální hodnota semivariance, je rovna rozptylu zpracovávaných
dat. Se zvětšující se vzdáleností bodů v páru se už semivariance nemění. Celková hodnota
prahu je součtem zbytkového rozptylu (nugget c_0 , viz níže) a tzv. částečného prahu
(partial sill c_p)
- Range a neboli dosah – kritická vzdálenost bodů v páru, při které semivariance dosahuje
prahu. Dosah definuje pro predikovaný bod velikost okolí, které je nutné uvažovat při
interpolaci hodnoty v tomto bodě;

- Nugget variance c_0 neboli zbytkový rozptyl – hodnota semivariance pro $\mathbf{h} = 0$. Teoretický semivariogram prochází počátkem souřadné soustavy, ale v experimentálních semivariogramech často dochází k nugget efektu.

Protože uvedený empirický semivariogram (1) je citlivý na vliv odlehlých bodů v prostoru, Cressie and Hawkins (1980) navrhli další, robustnější modely experimentálního semivariogramu. Následující dva jsou obsaženy v programu **Gorint**:

$$\gamma(\mathbf{h}) = \frac{\left(\frac{1}{N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} |z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})|^{\frac{1}{2}} \right)^4}{2 \left(0,457 + \frac{0,494}{N(\mathbf{h})} \right)} \quad (3)$$

Druhý vztah je založen na mediánu:

$$\gamma(\mathbf{h}) = \frac{\left(\text{med} \{ |z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})|^{\frac{1}{2}} : \mathbf{h} \in N(\mathbf{h}) \} \right)^4}{2 \cdot 0,457} \quad (4)$$

Teoretickým semivariogramem se pak rozumí jednoduchá matematická funkce, která prokládá experimentální semivariogram. Variogram by měl být obecně trojrozměrný pro data v prostoru nebo dvojrozměrný pro data v ploše. S ohledem na malý počet dat se ovšem daná veličina, je-li to možno, považuje za izotropní a používá se jednorozměrný variogram, kde se vektor \mathbf{h} považuje za skalár h (všesměrný variogram) (Ježek, 2002). Rozlišujeme řadu teoretických variogramů, **Gorint** má implementovány následující čtyři:

(i) exponenciální:

$$\gamma(\mathbf{h}) = c_0 + c_p \left(1 - e^{-\frac{3\mathbf{h}}{a}} \right) \quad (5)$$

(ii) sférický:

$$\gamma(\mathbf{h}) = \begin{cases} c_0 + c_p \left(\frac{3\mathbf{h}}{2a} \right) & \text{pokud } \mathbf{h} \leq a \\ c_0 + c_p & \text{pokud } \mathbf{h} > a \end{cases} \quad (6)$$

(iii) gaussovský:

$$\gamma(\mathbf{h}) = c_0 + c_p \left(1 - e^{-\frac{3\mathbf{h}^2}{a^2}} \right) \quad (7)$$

(iv) mocninný:

$$\gamma(\mathbf{h}) = c_0 + a\mathbf{h}^p \quad \text{pro } 0 < p < 2 \quad (8)$$

Je-li hodnota exponentu $p = 1$, jde o lineární semivariogram.

Matematický popis uvedených typů teoretických semivariogramů uvádí např. Goovaerts (1997).

2.3.2 Kriging

Interpolovaný povrch tvoří tři složky:

- obecný trend (drift) – deterministická složka; trukturální komponenta s konstantním průměrem či trendem;
- regionalizovaná proměnná – stochastická složka; kolísání průběhu funkce (sníženiny či vyvýšeniny), jehož podstatu lze vyjádřit určitou matematickou funkcí jako v případě trendu, ale které vyjadřuje určitou prostorovou korelaci;
- náhodný šum – náhodná, prostorově nezávislá složka; gaussovský šum s nulovou střední hodnotou a nenulovým rozptylem.

Všechny tři složky jsou analyzovány separovaně. První složka je odhadována za pomoci obecné trendové funkce, druhá (prostorově korelované kolísání) je analyzována metodou variogramu.

Obecná rovnice krigingu pro nestranný lineární regresní odhad neznámé hodnoty $\hat{Z}(\mathbf{x})$ v místě (bodě) \mathbf{x} je (Goovaerts, 1997):

$$\hat{Z}(\mathbf{x}) - m(\mathbf{x}) = \sum_{i=1}^{n(\mathbf{x})} w_i(\mathbf{x}) [Z(\mathbf{x}_i) - m(\mathbf{x}_i)] \quad (9)$$

$Z(\mathbf{x}_i)$ jsou známé výšky v měřených bodech \mathbf{x}_i ; w_i^{SK} jsou váhy, $m(\mathbf{x})$ a $m(\mathbf{x}_i)$ jsou předpokládané hodnoty náhodné veličiny $Z(\mathbf{x})$ a $Z(\mathbf{x}_i)$, trend.

Chyba modelu je definována jako náhodná proměnná $\hat{Z}_{SK}(\mathbf{x}) - Z(\mathbf{x})$. Rozptyl této chyby je pak definován (Goovaerts, 1997) takto:

$$\sigma_E^2(\mathbf{x}) = \text{Var} \left\{ \hat{Z}_{SK}(\mathbf{x}) - Z(\mathbf{x}) \right\} \quad (10)$$

Kriging poskytuje nestrannost odhadu, proto je rozptyl chyby odhadu σ_E^2 minimalizován tak, aby platilo:

$$\text{E} \left\{ \hat{Z}_{SK}(\mathbf{x}) - Z(\mathbf{x}) \right\} = 0 \quad (11)$$

Náhodnou funkci $Z(\mathbf{x})$ lze rozdělit na reziduální komponentu $R(\mathbf{x})$ a trendovou komponentu $m(\mathbf{x})$ (Goovaerts, 1997),

$$Z(\mathbf{x}) = R(\mathbf{x}) + m(\mathbf{x}) \quad (12)$$

Reziduální komponenta představuje *stacionární náhodnou funkci* s nulovou střední hodnotou a kovariancí $C_R(\mathbf{h})$:

$$\text{E}\{R(\mathbf{x})\} = 0 \quad (13)$$

$$\text{Cov}\{R(\mathbf{x}), R(\mathbf{x} + \mathbf{h})\} = \text{Cov}\{R(\mathbf{x}) \cdot R(\mathbf{x} + \mathbf{h})\} = C_R(\mathbf{h}) \quad (14)$$

Předpokládaná střední hodnota náhodné funkce $Z(\mathbf{x})$ v lokalitě \mathbf{x} je pak hodnotou trendu:

$$\text{E}\{Z(\mathbf{x})\} = m(\mathbf{x}) \quad (15)$$

Na základě typu trendu rozlišujeme tři varianty krigingu (Goovaerts, 1997):

- simple kriging (SK) – trend je známý a konstantní, $m = \text{konst.}$;
- ordinary kriging (OK) – trend je konstantní, ale neznámý;
- universal kriging (UK) – trendová komponenta je lineární kombinací matematických funkcí závislých na poloze.

2.3.3 Simple kriging

Modelování trendové komponenty $m(\mathbf{x})$ jako známé stacionární střední hodnoty m umožňuje rovnici (9) psát jako lineární kombinaci $(n(\mathbf{x}) + 1)$ vstupních komponent: $n(\mathbf{x})$ náhodných proměnných $Z(\mathbf{x}_i)$ a střední hodnotu m (Goovaerts, 1997):

$$\hat{Z}_{SK}(\mathbf{x}) = \sum_{i=1}^{n(\mathbf{x})} w_i^{SK}(\mathbf{x}) [Z(\mathbf{x}_i) - m] + m \quad (16)$$

$$= \sum_{i=1}^{n(\mathbf{x})} w_i^{SK}(\mathbf{x}) Z(\mathbf{x}_i) + \left[1 - \sum_{i=1}^{n(\mathbf{x})} w_i^{SK}(\mathbf{x}) \right] m \quad (17)$$

$\hat{Z}_{SK}(\mathbf{x})$ je odhad výšky v predikovaném bodu X ; $Z(\mathbf{x}_i)$ jsou známé výšky v měřených bodech \mathbf{x}_i ; w_i^{SK} jsou váhy.

$n(\mathbf{x})$ vah w_i^{SK} se určí minimalizováním rozptylu $\sigma_E^2(\mathbf{x})$ chyby odhadu podle vztahu (10) a pod podmínkou (11).

Chybu odhadu $\hat{Z}_{SK}(\mathbf{x}) - Z(\mathbf{x})$ lze vyjádřit jako lineární kombinaci $(n(\mathbf{x}) + 1)$ reziduálních náhodných proměnných $R(\mathbf{x})$ a jedné proměnné $R(\mathbf{x})$ (Goovaerts, 1997):

$$\hat{Z}_{SK}(\mathbf{x}) - Z(\mathbf{x}) = [\hat{Z}_{SK}(\mathbf{x}) - m] - [Z(\mathbf{x}) - m] \quad (18)$$

$$= \sum_{i=1}^{n(\mathbf{x})} w_i^{SK}(\mathbf{x}) R(\mathbf{x}_i) - R(\mathbf{x}) \quad (19)$$

$$= \hat{R}_{SK}(\mathbf{x}) - R(\mathbf{x}) \quad (20)$$

kde $R(\mathbf{x}_i) = Z(\mathbf{x}_i) - m$ a $R(\mathbf{x}) = Z(\mathbf{x}) - m$. Rozptyl chyb odhadu lze potom vyjádřit:

- pomocí kovariance, viz Goovaerts (1997),

$$\sigma_E^2(\mathbf{x}) = \text{Var} \{ \hat{R}_{SK}(\mathbf{x}) \} + \text{Var} \{ R(\mathbf{x}) \} - 2\text{Cov} \{ \hat{R}_{SK}(\mathbf{x}), R(\mathbf{x}) \} \quad (21)$$

$$= \sum_{i=1}^{n(\mathbf{x})} \sum_{j=1}^{n(\mathbf{x})} w_i^{SK}(\mathbf{x}) w_j^{SK}(\mathbf{x}) C_R(\mathbf{x}_i - \mathbf{x}_j) + C_R(0) - 2 \sum_{i=1}^{n(\mathbf{x})} w_i^{SK}(\mathbf{x}) C_R(\mathbf{x}_i - \mathbf{x}) \quad (22)$$

$$= Q(w_i^{SK}(\mathbf{x}), i = 1, \dots, n(\mathbf{x})) \quad (23)$$

- pomocí semivariogramu, viz Ježek (2002),

$$\sigma_E^2(\mathbf{x}) = 2 \sum_{i=1}^{n(\mathbf{x})} w_i^{SK} \gamma(\mathbf{x}_i - \mathbf{x}) - \sum_{i=1}^{n(\mathbf{x})} \sum_{j=1}^{n(\mathbf{x})} w_i^{SK}(\mathbf{x}) w_j^{SK}(\mathbf{x}) \gamma(\mathbf{x}_i - \mathbf{x}_j) \quad (24)$$

kde $\gamma(\mathbf{x}_i - \mathbf{x}_j)$ je hodnota semivariance určená ze semivariogramu pro příslušnou dvojici bodů.

Rozptyl chyb $\sigma_E^2(\mathbf{x})$ daný rovnicí (23) představuje kvadratickou formuli Q s celkovým počtem $n(\mathbf{x})$ vah $w_i^{SK}(\mathbf{x})$. Optimální váhy, tj. poskytující minimální rozptyl chyb odhadu, obdržíme položením všech $n(\mathbf{x})$ parciálních derivací rovno 0:

$$\frac{\partial Q(\mathbf{x})}{\partial w_i^{SK}(\mathbf{x})} = \sum_{j=1}^{n(\mathbf{x})} w_j^{SK}(\mathbf{x}) C_R(\mathbf{x}_i - \mathbf{x}_j) - C_R(\mathbf{x}_i - \mathbf{x}) = 0 \quad (25)$$

Rovnice (25) je známa jako *systém normálních rovnic* (Luenberger, 1969) nebo *simple kriging system* (Goovaerts, 1997).

Díky stacionaritě modelu platí $C_R(\mathbf{h}) = C(\mathbf{h})$ a rovnici (25) můžeme přepsat do konečného tvaru

$$\sum_{j=1}^{n(\mathbf{x})} w_j^{SK}(\mathbf{x}) C(\mathbf{x}_i - \mathbf{x}_j) = C(\mathbf{x}_i - \mathbf{x}) \quad (26)$$

pro $i = 1, \dots, n(\mathbf{x})$. Minimální rozptyl chyb, zvaný též *SK variance*, je získán dosazením rovnice (26) do rovnice (23) rozptylu chyb v kovarianční formě:

$$\sigma_{SK}^2(\mathbf{x}) = C(0) - \sum_{i=1}^{n(\mathbf{x})} w_i^{SK}(\mathbf{x}) C(\mathbf{x}_i - \mathbf{x}) \quad (27)$$

Maticová formulace. Simple kriging system (26) lze přepsat do maticové podoby, viz (Goovaerts, 1997):

$$\mathbf{K}_{SK} \mathbf{w}_{SK}(\mathbf{x}) = \mathbf{k}_{SK} \quad (28)$$

kde \mathbf{K}_{SK} je matice $n(\mathbf{x}) \times n(\mathbf{x})$ kovariancí $C(\mathbf{x}_i - \mathbf{x}_j)$:

$$\mathbf{K}_{SK} = \begin{bmatrix} C(\mathbf{x}_1 - \mathbf{x}_1) & \cdots & C(\mathbf{x}_1 - \mathbf{x}_{n(\mathbf{x})}) \\ \vdots & \ddots & \vdots \\ C(\mathbf{x}_{n(\mathbf{x})} - \mathbf{x}_1) & \cdots & C(\mathbf{x}_{n(\mathbf{x})} - \mathbf{x}_{n(\mathbf{x})}) \end{bmatrix} \quad (29)$$

$\mathbf{w}_{SK}(\mathbf{x})$ je sloupcový vektor vah,

$$\mathbf{w}_{SK}(\mathbf{x}) = \begin{bmatrix} w_1^{SK} \\ \vdots \\ w_{n(\mathbf{x})}^{SK}(\mathbf{x}) \end{bmatrix} \quad (30)$$

a \mathbf{k}_{SK} je sloupcový vektor kovariancí pro páry známých bodů \mathbf{x}_i s bodem predikovaným, \mathbf{x} ,

$$\mathbf{w}_{SK}(\mathbf{x}) = \begin{bmatrix} C(\mathbf{x}_1 - \mathbf{x}) \\ \vdots \\ C(\mathbf{x}_{n(\mathbf{x})} - \mathbf{x}) \end{bmatrix} \quad (31)$$

Váhy získáme výpočtem rovnice

$$\mathbf{w}_{SK}(\mathbf{x}) = \mathbf{K}_{SK}^{-1} \mathbf{k}_{SK} \quad (32)$$

2.4 Vážené průměrování metodou inverzních vzdáleností (IDW)

Jedna z nejpoužívanějších deterministických metod pro prostorové interpolace. Je relativně rychlá a snadná z hlediska výpočtu a stala se alternativou geostatistické metody kriging – zejména u modelů, kde z dostupného vzorku dat nelze sestavit vyhovující variogram (Lu and Wong, 2008).

Metoda je založena na předpokladu, že hodnota veličiny v predikovaném bodě je váženým průměrem hodnot okolních měřených bodů, přičemž váhy jsou určeny jako inverzní vzdálenost měřeného bodu od bodu interpolovaného – čím menší vzdálenost, tím větší váha. Jde o metodu

exaktní, lokální, deterministickou (Bartier and Keller, 1996). Užívaná zkratka metody IDW pochází z anglického výrazu *inverse distance weighted*.

Odhadovaná hodnota je vypočtena jako lineární kombinace vstupních hodnot:

$$\hat{z}(\mathbf{x}) = \frac{\sum_{i=1}^n w_i z(\mathbf{x}_i)}{\sum_{i=1}^n w_i} \quad (33)$$

kde w_i jsou příslušné váhy závislé na inverzi vzdálenosti.

Je-li mezi predikovaným a měřeným bodem vzdálenost d_i , potom platí (Bartier and Keller, 1996):

$$\hat{z}(\mathbf{x}) = \frac{\sum_{i=1}^n d_i^{-r} z(\mathbf{x}_i)}{\sum_{i=1}^n d_i^{-r}} \quad (34)$$

kde r je parametr metody IDW definovaný uživatelem; obvyklá hodnota se volí z intervalu $\langle 1; 5 \rangle$ a ovlivňuje, v jakém poměru klesá hodnota váhy (tj. vliv) měřeného bodu s rostoucí vzdáleností od bodu interpolovaného.

Metoda IDW často produkuje povrch, který je charakteristický koncentrickými strukturami kolem interpolovaných bodů (tzv. bulls eyes) (Ježek and kolektiv, 2008; Bartier and Keller, 1996). Protože IDW je založena na lokálním průměrování, nikdy neposkytuje odhady mimo rozsah hodnot měřených bodů.

3 Instalace programu

Vlastní program `Gorint` je skriptem psaným v programovacím jazyce `R`. `R` je platformou pro statistické výpočty všeho druhu a poskytuje i grafické výstupy. Prostřednictvím velké škály implementovaných funkcí je v `R` umožněno provádět řadu statistických a grafických technik (lineární a nelineární modelování, statistické testování, analýzy časových řad aj.), stejně tak je možné si pro jednotlivé funkce naprogramovat algoritmy přesně na míru.

Pro spuštění algoritmů v `R`, a tedy i programu `Gorint`, je potřeba mít nainstalovanu konzoli `R` (The R Project for Statistical Computing), kterou lze volně stáhnout z oficiálních webových stránek tohoto projektu: <http://mirrors.softliste.de/cran/>. Zde zvolíte operační systém Vašeho PC a následně se necháte vést nabídkou.

Poté zde stačí pouze otevřít a spustit soubor `Gorint.R`. Následné operace a výpočty v tomto programu jsou popsány níže.

4 Použití programu

4.1 Vstupní data

Vstupní data se do programu **Gorint** (a obecně do **R**) načítají z textového souboru (soubory s příponou `*.txt`, `*.dat`). Seznam vstupních bodů, z nichž chceme interpolovat povrch, musí být v souboru uspořádán do jednoduché matice, v níž jsou souřadnice bodů X , Y , Z členěny do samostatných sloupců oddělených mezerou, tabulátorem nebo čárkou. První řádek matice je hlavičkou obsahující označení sloupců. Jako desetinný operátor používejte výhradně tečku:

1	y	x	z	c	s
2	831274.79	1154028.35	1183.41	5.271	0.422
3	831274.80	1154028.42	1183.40	4.783	-0.112
4	831435.61	1154351.35	1186.94	3.325	-0.227
5	830172.38	1156260.08	1367.70	8.664	1.109
6	830172.29	1156259.98	1367.67	8.541	0.997

Pořadí sloupců není důležité, stejně tak může matice obsahovat i další sloupce, které k výpočtu nepotřebujeme. Poté, co uživatel provede načtení dat ze vstupního souboru do matice, se kterou dále pracuje (viz sekci 4.3), před výpočtem si sám nadefinuje čísla sloupců, které chce k interpolaci použít (viz sekci 4.2 – řádek 2-4 kódu v závěru této sekce).

4.2 Přehled parametrů

V tabulce 1 jsou uvedeny veškeré parametry (argumenty) funkcí programu **Gorint**, které je možné zadat. Parametry je nutné zadat dříve, než vyvoláme funkci, mezi jejíž argumenty tyto parametry patří.

Parametr	Definiční obor, popis parametru
<code>file</code>	string Cesta a název souboru se vstupními daty (string nutno psát vždy do uvozovek).
<code>x_coord_col</code>	integer Pořadí sloupce ve vstupním souboru (bráno zleva), ve kterém jsou x -ové souřadnice bodů
<code>y_coord_col</code>	integer Pořadí sloupce ve vstupním souboru (bráno zleva), ve kterém jsou y -ové souřadnice bodů
<code>z_observed_col</code>	integer Pořadí sloupce ve vstupním souboru (bráno zleva), ve kterém jsou z -ové souřadnice bodů (interpol. veličina)
<code>crossval</code>	{0;1} Parametr, který rozhoduje, zda bude interpolováno v požadovaném gridu (volba crossval = 0), nebo zda interpolace proběhne pouze ve vstupních bodech pro účel crossvalidace – program pak porovná měřené hodnoty s hodnotami interpolovanými (volba crossval = 1)
<code>grid_resolution</code>	real+ Rozlišení výstupního rastru, tj. délka hrany čtvercové buňky v gridu. Zadává se v jednotkách indentických s jednotkami souřadnic vstupních bodů
<code>lag</code>	real+ Parametr semivariogramu, udává vzdálenost h mezi páry bodů (viz sekci 2.3.1)
<code>cut</code>	real+ Doplňkový parametr semivariogramu, který udává nejzazší vzdálenost pro dosah (range) a semivariogramu – zde bude variogram na ose x ukončen
<code>nugget</code>	real+ Parametr semivariogramu, udává zbytkový rozptyl c_0 , tj. hodnotu semivariance pro $h = 0$ (viz sekci 2.3.1)
<code>type_of_experimental_variogram</code>	{1;2;3} Udává typ použitého experimentálního semivariogramu 1 = konvenční typ dle vztahu (1), 2 = robustní typ dle vztahu (3), 3 = robustní typ založený na mediánu dle vztahu (4),

Tabulka 1: Přehled parametrů a jejich popis

Parametr	Definiční obor, popis parametru
<code>type_of_theoretical_variogram</code>	<code>{1;2;3;4}</code> Udává typ použitého teoretického semivariogramu <code>1</code> = exponenciální typ dle vztahu (5), <code>2</code> = sférický typ dle vztahu (6), <code>3</code> = gaussovský typ dle vztahu (7), <code>4</code> = mocninný typ dle vztahu (8)
<code>n</code>	<code>integer</code> Počet okolních bodů vstupujících do interpolace každého predikovaného bodu
<code>max_radius</code>	<code>real+</code> Maximální povolená vzdálenost od predikovaného bodu, ze které jsou brány okolní body pro interpolaci. Zadává se v jednotkách indentických s jednotkami souřadnic vstupních bodů
<code>r</code>	<code><1;5></code> Exponent interpolační techniky IDW (viz sekci 2.4, vztah 33)
<code>pocet_kal</code>	<code>integer</code> Počet vstupních bodů, které se nebudou účastnit procesu interpolace, ale budou vybrány pro validaci
<code>title</code>	<code>string</code> Název (label) výstupního semivariogramu (nutno zadat do uvozovek)
<code>out_dir</code>	<code>string</code> Cesta k adresáři, kam se budou ukládat výstupy (zadejte včetně závěrečného lomítka)

Tabulka 1: Přehled parametrů a jejich popis (pokračování)

Ukázka nastavení parametrů uvedených v tabulce 1:

```

1 file = 'C:/Muj_pracovni_adresar/soubor_s_daty.txt'
2 x_coord_col = 2
3 y_coord_col = 1
4 z_observed_col = 3
5 crossval = 0
6 grid_resolution = 5
7 lag = 5
8 cut = 80
9 nugget = 0
10 type_of_experimental_variogram = 1
11 type_of_theoretical_variogram = 4
12 n = 8

```

```

13 max_radius = 100
14 r = 3
15 pocet_kal = 400
16 title = 'název mého rastru '
17 out_dir = 'C:/Muj_pracovni_adresar/vystupy/'

```

4.3 Přehled hlavních funkcí, jejich argumentů a způsobu zadávání

Veškeré funkce programu **Gorint** je třeba nejprve načíst a zkompilovat v konzoli R, což se provede následujícím jednoduchým příkazem:

```
1 source('C:/Muj_pracovni_adresar/Gorint.R')
```

V kulatých závorkách a uvozovkách je uvedena cesta ke skriptu tohoto programu.

V následujícím přehledu jsou uvedeny funkce programu **Gorint** s příslušnými argumenty. Argumenty je třeba zadat do kulatých závorek ihned za název funkce.

- Funkce **read.table**: standardní funkce R, která slouží k načtení dat ze zadaného vstupního souboru:

```
1 in_data = read.table(file, header = TRUE)
```

Má-li vstupní soubor v prvním řádku hlavičku, zadejte **header = TRUE**, v opačném případě bude **header = FALSE**.

- Funkce **shuffle**: promíchá seznam vstupních bodů, což je nezbytné, chceme-li následně provést náhodný výběr vzorku dat, na kterém lze otestovat úspěšnost interpolačního procesu (verifikace). Takto vybrané body se proto neúčastní interpolace výsledného rastru.

```
1 shuffled_data = shuffle(in_data)
```

- Funkce **make_grid**: vytvoří prázdný grid, který bude následně naplněn interpolovanými hodnotami:

```
1 Out_raster = make_grid(in_data, grid_resolution, crossval)
```

- Funkce **variogram**: vytvoří empirický semivariogram zadaného typu a proloží ho teoretickým semivariogramem, opět dle volby jeho typu:

```
1 semivariogram = variogram(in_data, lag, cut, nugget,
    type_of_experimental_variogram, type_of_theoretical_variogram,
    title)
```

Tato funkce je implementována do obou funkcí krigování (**ok** a **uk**), proto není nutné ji spouštět takto samostatně. K proložení empirického semivariogramu teoretickým je zde využita funkce nelineárních nejmenších čtverců **nls** (Nonlinear Least Squares, více viz <http://stat.ethz.ch/R-manual/R-patched/library/stats/html/nls.html>).

- Funkce **ok**: metodou *ordinary kriging* interpoluje hodnoty do připraveného rastru:

```
1 Z_ok = ok(in_data , out_raster , lag , cut , nugget ,  
           type_of_experimental_variogram , type_of_theoretical_variogram , n ,  
           r , crossval , title , max_radius )  
2 map_rgl(Z_ok , title )  
3 map_sp(Z_ok , title )  
4 map_fields(Z_ok , title )
```

Výstupem této funkce je textový soubor s interpolovanými hodnotami a vedle toho pdf soubor s obrázkem výsledného rastru, k čemuž slouží funkce **map_sp**, **map_rgl** nebo **map_field**. Pro tento účel je v R potřeba mít nainstalovány balíčky funkcí **rgl**, **sp** (nebo **gstat**) a balík **fields**. Tyto komponenty lze zdarma nainstalovat přímo v prostředí konzole R pomocí nabídky v horním řádku.

- Funkce **uk**: metodou *universal kriging* interpoluje hodnoty do připraveného rastru:

```
1 Z_uk = uk(in_data , out_raster , lag , cut , nugget ,  
           type_of_experimental_variogram , type_of_theoretical_variogram , n ,  
           r , crossval , title , max_radius )  
2 map_field(Z_uk , title )
```

K vizualizaci je možné opět využít funkce uvedené výše, např. **map_field**.

- Funkce **idw**: metodou *IDW* interpoluje hodnoty do připraveného rastru:

```
1 Z_idw = idw(in_data , out_raster , n , r , crossval , max_radius )  
2 map_field(Z_idw , title )
```

Formát výstupního rastru je identický s předešlými metodami interpolace.

5 Případová studie – ukázka aplikace

Ukázka interpolace rastru bodově měřené koncentrace zinku ve svrchní vrstvě půdy užitím metody ordinary kriging v programu **Gorint**. Tato cvičná data byla změřena na pokusné lokalitě na Příbramsku a jsou přiložena k programu.

5.1 Vstupní kód programu

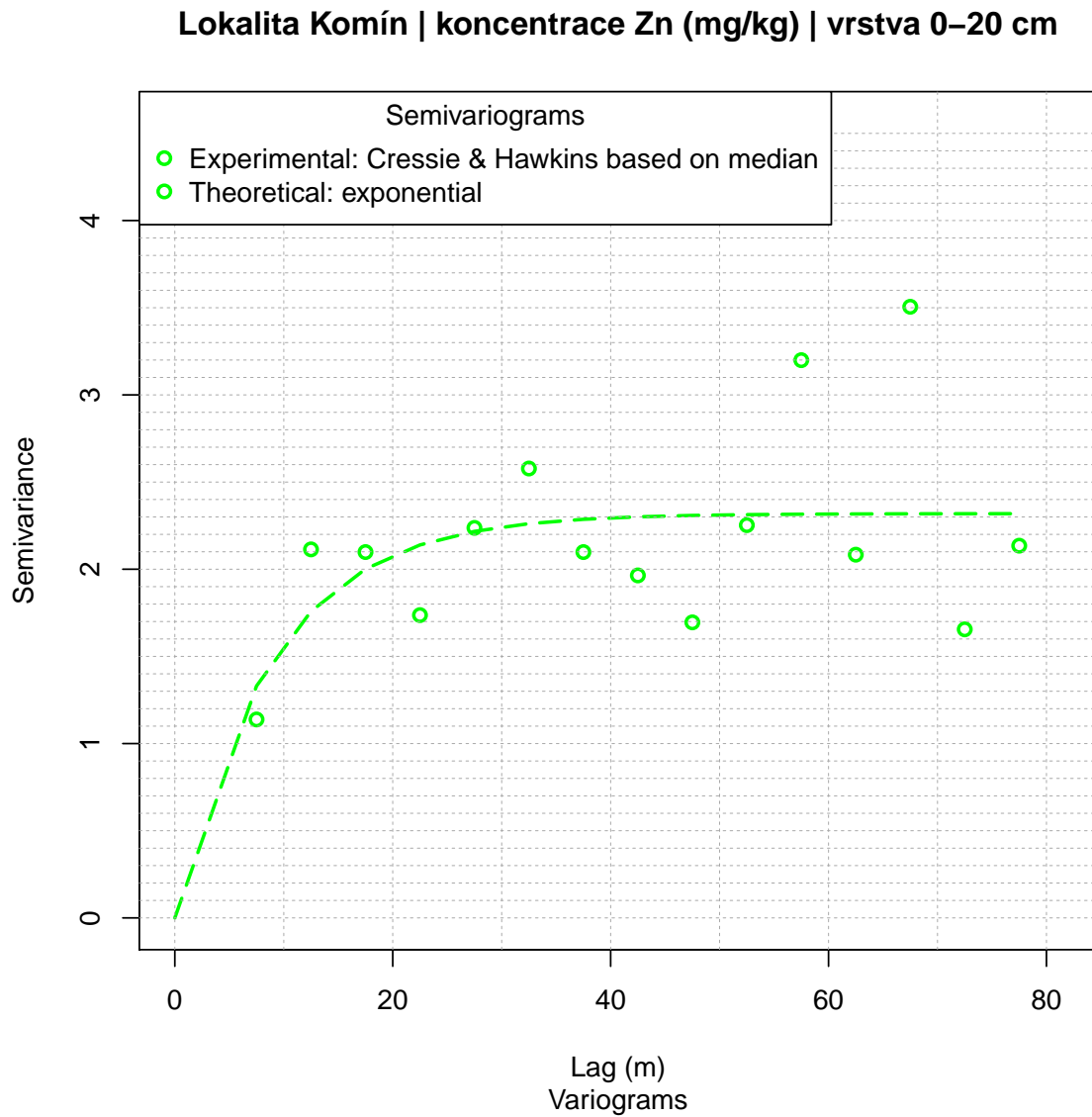
Řádek 1 ukazuje cestu ke vstupnímu souboru s daty. V řádcích 2-4 se zadává pořadí sloupců jednotlivých souřadnic, jak jsou uspořádané ve vstupním datovém souboru. Vlastní přiřazení souřadnic do matice `in_data`, se kterou program dále pracuje, je řízeno řádky 20-23 (viz níže). Řádky 5-17 ukazují načtení parametrů (argumentů funkcí). Řádkem 18 načítáme vlastní program **Gorint** se všemi jeho funkcemi, řádkem 19 načítáme data ze vstupního souboru do matice `in_data`. Následující čtyři řádky 20-23 je nutné zadat, aby došlo ke správnému přiřazení zadaných souřadnic bodů do příslušných sloupců matice `in_data`. Další řádky řídí aplikaci jednotlivých funkcí popsaných v sekci 4.3.

```

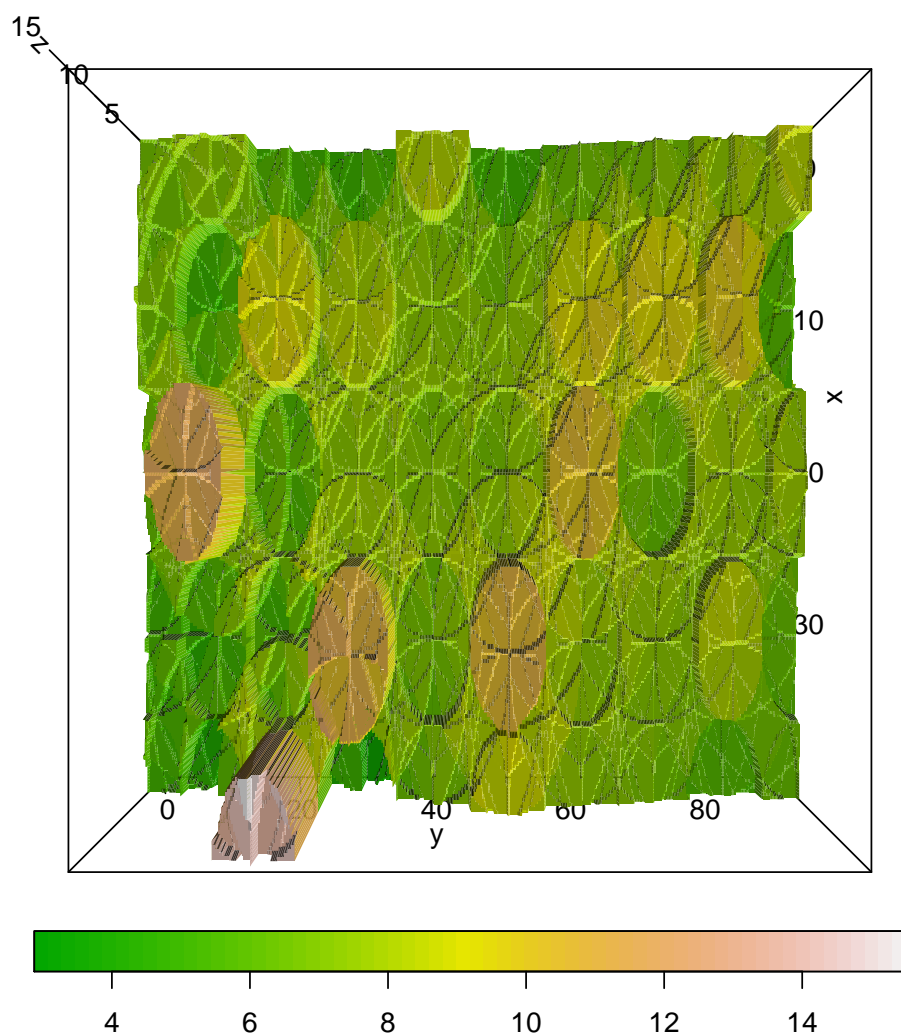
1 file = 'C:/Muj_pracovni_adresar/soubor_s_daty.txt'
2 x_coord_col = 1
3 y_coord_col = 2
4 z_observed_col = 3
5 crossval = 0
6 grid_resolution = 0.2
7 lag = 5
8 cut = 80
9 nugget = 0
10 type_of_experimental_variogram = 3
11 type_of_theoretical_variogram = 1
12 n = 8
13 max_radius = 40
14 r = 2.5
15 pocet_kal = 0
16 title = 'Lokalita Komín | Koncentrace Zn (mg/kg) | vrstva 0–20 cm'
17 out_dir = 'C:/Muj_pracovni_adresar/vystupy/'
18 source('C:/Muj_pracovni_adresar/Gorint.R')
19 in_data = read.table(file, header = TRUE)
20 x_coord = in_data[, x_coord_col]
21 y_coord = in_data[, y_coord_col]
22 z_observed = in_data[, z_observed_col]
23 in_data = data.frame(x_coord, y_coord, z_observed)
24 out_grid = make_grid(in_data, grid_resolution, crossval)
25 Z_ok = ok(in_data, out_raster, lag, cut, nugget,
            type_of_experimental_variogram, type_of_theoretical_variogram, n,
            r, crossval, title, max_radius)
26 map_fields(Z_ok, title)

```

5.2 Výstupy programu



Obrázek 1: Empirický semivariogram (body) dle modelu popsaném ve vztahu (4) (Cressie and Hawkins, 1980) proložený teoretickým exponenciálním semivariogramem (5)



Obrázek 2: Výsledný rastr získaný interpolační technikou ordinary kriging dle zadaných parametrů uvedených v sekci 5.1

Literatura

- Bartier, P. M., Keller, C. P., 1996. Multivariate interpolation to incorporate thematic surface data using inverse distance weighting (idw). *Computers and Geosciences* 22 (7), 795–799, copyright © 1996 Elsevier Science Ltd.
- Burrough, P., McDonnell, R., 1998. *Principles of Geographical Information Systems*. Oxford University Press, New York, USA, 336 p., ISBN: 0-19-823365-5.
- Chaplot, V., Darboux, F., Bourennane, H., Leguédais, S., Silvera, N., Phachomphon, K., 2006. Accuracy of interpolation techniques for the derivation of digital elevation models in relation to landform types and data density. *Geomorphology* 77, 126–141, © 2006 Elsevier B.V.
- Collins, S. H., Moon, G. C., 1981. Algorithms for dense digital terrain models. *Photogram. Engineering and Remote Sensing* 47, 71–76.
- Cressie, N., Hawkins, D. M., 1980. Robust estimation of the variogram: I. *Mathematical Geology* 12 (2), 115–125.
- El-Sheimy, N., Valeo, C., Habib, A., 2005. *Digital terrain modeling: acquisition, manipulation, and application*. Artech House remote sensing library. Artech House, Inc., Canton Street, Norwood, ISBN: 1-58053-921-1.
- Fairfield, J., Leymarie, P., 1991. Drainage network from grid digital elevation models. *Water Resources Research* 27 (5), 709–717.
- Goovaerts, P., 1997. *Geostatistical for Natural Resources Evaluation*. Applied Geostatistic. Oxford University Press, Inc., New York, USA, ISBN: 978-0-19-511538-3.
- Hengl, T., Gruber, S., Shrestha, D. P., 2003. *Digital terrain analysis in ILWIS: lecture notes and user guide*. Department of Earth Systems Analysis, International Institute for Geo-information Science and Earth Observation (ITC), Enschede, Netherlands.
- Heritage, G. L., j. Milan, D., Large, A. R. G., Fuller, I. C., 2009. Influence of survey strategy and interpolation model on dem quality. *Geomorphology* 112, 334–344, copyright © 2009 Elsevier Ltd.
- Ježek, J., 2002. *Výuka statistiky pro nestatistiky: Využití programu MATLAB při výuce geostatistiky* (s. 48–56). ÚAMVT, PřF UK Praha, Albertov 6, 128 43 Praha 2, © Matfyzpress 2002.
- Ježek, J., kolektiv, 2008. *Geostatistika: metody prostorových interpolací, Kriging*. ÚAMVT, PřF UK Praha, Albertov 6, 128 43 Praha 2, učební texty; nepublikováno.
- Li, Z., Zhu, Q., Gold, C., 2005. *Digital Terrain Modeling: Principles and Methodology*. Library of Congress. CRC Press, Boca Raton, Florida, USA, ISBN: 0-415-32462-9.
- Lu, G. Y., Wong, D. W., 2008. An adaptive inverse-distance weighting spatial interpolation technique. *Computers and Geosciences* 34, 1044–1055, 2008 © Elsevier Ltd.
- Luenberger, D. G., 1969. *Optimization by Vector Space Methods*. John Wiley and Sons, New York, USA.

- Miller, C., LaFlamme, R. A., 1958. The digital terrain modeling – theory and application. *Photogrammetric Engineering* 24 (3), 433–442.
- Mitášová, H., Mitáš, L., Harmon, R. S., 2005. Simultaneous spline approximation and topographic analysis for lidar elevation data in open source gis. *Geoscience and Remote Sensing Letters* 0 (0).
- Moore, I. D., Grayson, R. B., Ladson, A. R., 1991. Digital terrain modeling: A review of hydrological, geomorphological, and ecological applications. *Hydrological Processes* 5, 3–30, copyright © 1991 by John Wiley and Sons Ltd.
- O’Callaghan, J. F., Mark, D. M., 1984. The extraction of drainage networks from digital elevation data. *Computer Vision, Graphics and Image Processing* 28, 323–344.
- Quinn, P., Beven, K., Chevallier, P., Planchon, O., 1991. The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. *Hydrological Processes* 5, 59–79.
- Sarangi, A., Madramootoo, C. A., Enright, P., 2006. Comparison of spatial variability techniques for runoff estimation from a canadian watershed. *Biosystems Engineering* 95 (2), 295–308.
- Wilson, J. P., Gallant, J. C., 2000. *Terrain Analysis: Principles and Applications*. John Wiley & Sons, Inc., New York, USA, 479 p., ISBN: 0-471-32188-5.