

**Metodika pro statistické zpracování dat k identifikaci anomálií škodlivin
v životním prostředí**

**Doc. Dr. Vladimír Homola, Ph.D.
prof. Ing. Konstantin Raclavský, CSc.**

Ostrava, listopad 2011

Kontaktní údaje:

Institut geologického inženýrství, VŠB – Technická univerzita Ostrava
17. listopadu 15/2172, 708 33 Ostrava-Poruba
tel.: +420 59 699 1243, <http://geologie.vsb.cz>

1. Cíl metodiky

Metodika stanovuje postup při identifikaci anomálií ve výskytu škodlivin v pracovním a životním prostředí, a to z lineárně i prostorově rozložených dat. Používá přitom geostatistické nástroje implementované v řadě účelových aplikací, nebo dokonce jednoduše sestavitelných v běžně dostupných tabulkových procesorech. **Výhodou navrženého postupu** je jednak daleko *větší míra objektivizace* náhledu na zkoumané veličiny, jednak *řádové zvýšení rychlosti* získání požadovaných informací – obojí vzhledem k tradičně užívaným postupům využívajících standardní, většinou parametrické statistiky se značnou mírou ručních vstupů do procesu zpracování. Podstatnou výhodou je rovněž možnost *zmenšení počtu vstupních dat*, což s sebou nese i nezanedbatelný ekonomický efekt; vždyť právě data získávaná z terénu jsou jedna z nejdražších. **Vedlejším efektem** je pak možnost použití těchto postupů – a to v reálném čase – např. pro *modelování stavů veličin* získaných změnou vnějších parametrů, tj. řešení úloh typu „co by, kdyby“: odhady kritických stavů, předpovědi chování v budoucnu apod.

2. Princip metodiky

Metodika spočívá ve vyjmenování závazných kroků při zpracování a hodnocení definované množiny dat sledovaných látek pro získání definovaného výsledku. Jako zdrojová data metodika určuje logický formát tabulky relační databáze, v jednom nebo více sloupcích obsahující hodnoty sledované / sledovaných veličin, v jednom nebo ve dvou sloupcích jejich lineární nebo plošnou polohu. Tento zpracovávaný logický tvar může být buď jedna fyzická tabulka tabulkového procesoru nebo některého databázového formátu, nebo může jít o výsledek dotazu do několika různých datových zdrojů. Na takto definovaných datech stanovuje metodika posloupnost kroků data zpracovávajících. Jde o analýzu jejich variability, prostorového rozložení, lokálních odhadů a dalších. Kroky jsou stanoveny za účelem odhalení potenciálních anomálií sledovaných veličin ve sledovaném lineárním intervalu nebo plošné oblasti.

3. Oblast použití

Metodika je použitelná např. k výzkumným činnostem odhalujících bližší souvislosti mezi *kvalitou životních podmínek a zdravím obyvatelstva*, k *monitorování stavu a vývoje znečištění* daného regionu konkrétní škodlivinou, ke *kontrole kvality životního prostředí* v daném ukazateli atd. Aplikováno pak na konkrétní region, i k určení *expozice fyzických osob* danou škodlivinou.

Metodiku lze použít pro všechny kvantitativní veličiny mající prostorovou složku první nebo druhé dimenze. Limitující pro použití metodiky je počet zpracovávaných dat: metodika by neměla být používána pro méně než 10 hodnot v lineárním intervalu a méně než 30 hodnot v plošné oblasti.

Kontaktní údaje:

Institut geologického inženýrství, VŠB – Technická univerzita Ostrava
17. listopadu 15/2172, 708 33 Ostrava-Poruba
tel.: +420 59 699 1243, <http://geologie.vsb.cz>

4. Charakteristika objektů aplikace

4.1. Rozložení četnosti

Při analýze potenciálních anomálií v datovém souboru je zkoumání rozložení četnosti prvním krokem. I ten již může existenci anomálií vyloučit. Po event. odstranění odlehlých hodnot (např. pravidlem jednoho až tří sigma) a použitím některého testu (testy např. Kolmogorov – Smirnov, Pearson, Shapiro – Wilk a dalších) je na dané hladině významnosti přijata nebo nepřijata hypotéza o konkrétním rozdělení. Jestliže hypotéza není přijata, pokračuje se ve zkoumání dat některými z robustních metod. Zvláštní pozornost je třeba věnovat odlehlým hodnotám. Hrubé chyby se vylučují vždy. Ovšem reálné, zkoumané veličině vlastní extrémní hodnoty sice mohou v řadě případů působit značné potíže (např. při odhadu statistických charakteristik), pro zkoumání anomálií jsou však důležitým indikátorem.

4.2. Základní neparametrické statistiky

Pro hodnocení dat jednorozměrného datového souboru (bez ohledu na rozdělení) lze použít řadu neparametrických statistik. Především jde o dolní a horní omezení, oba kvartily, mezikvartilové rozpětí a medián – obecně pak $p\%$ kvantily. Pro účely prvotního pohledu na data jsou horní a dolní omezení rovna maximu a minimu. V kontrastu s neparametrickými statistikami se uvádí směrodatná odchylka a aritmetický průměr jako statistiky parametrické.

4.3. Krabicový graf

Patří mezi základní způsoby vizualizace základních charakteristik hodnocených dat. Ačkoliv jde o obecné znázornění vzájemných vztahů neparametrických statistik, lze při zkoumání anomálií dané veličiny vhodně využít. Konkrétní tvary tohoto grafu pak pomohou při stanovení dalšího postupu analýzy anomálií.

4.4. Semivariogram

Podstatnou částí analýzy anomálií je zkoumání variability dané veličiny v lineárním či prostorovém kontextu. Tato část se zabývá tím, jak se liší difference hodnot dané veličiny v závislosti na vzdálenosti. Náznověji: jak se liší difference dané veličiny ve dvou lokalitách „blízko“ sebe a „dál“ od sebe. Základem je zde průměrná kvadratická odchylka od průměrné hodnoty, ovšem jen těch dvojic dat, jejichž vzdálenost leží v intervalu $<h-dh; h+dh>$. Jde tedy o funkci vzájemné vzdálenosti. Právě z této funkce lze usoudit na některé anomálie.

4.5. Izolinie

Metody lokálních odhadů dovolují - např. aplikací na uzly obdélníkové sítě - modelovat pravděpodobné chování sledované veličiny na definovaném zájmovém území jako funkci polohy (obecně funkci dvourozměrných dat). Modelem je „plocha“, jejíž body v konstantní vzdálenosti od nulové roviny tvoří soustavu izolinií. Už zobrazení modelové plochy jako trojrozměrného objektu ukáže potenciální anomálie. Častější tzv. topografické zobrazení obsahující především izolinie má stejnou vypovídací schopnost, navíc dovoluje díky soustavě souřadné poměrně přesné odečty, lokalizaci dat a další.

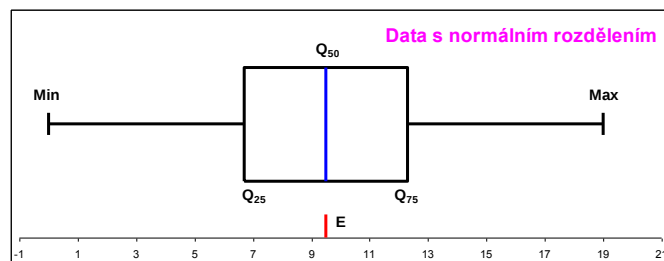
Kontaktní údaje:

Institut geologického inženýrství, VŠB – Technická univerzita Ostrava
17. listopadu 15/2172, 708 33 Ostrava-Poruba
tel.: +420 59 699 1243, <http://geologie.vsb.cz>

5. Postup

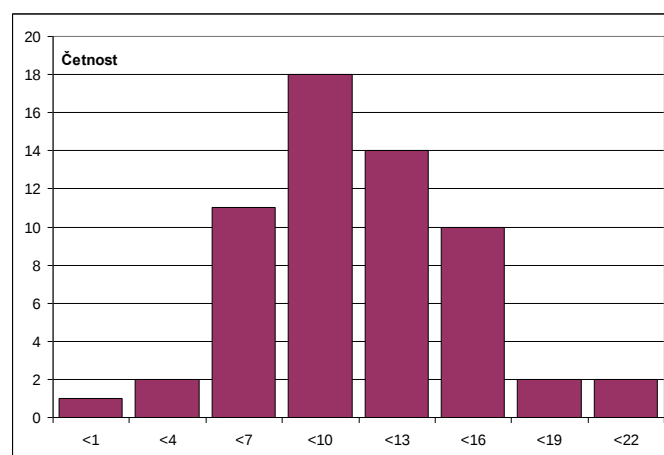
5.1. Rozložení četnosti

Prvním krokem při zkoumání rozložení četnosti dat je jejich grafické znázornění; už to napoví, zda má smysl pokračovat v testování dalšími metodami. Nejjednodušší a nejrychlejší je krabicový graf (box-and-whiskers plot). Vyznačuje polohu minima, dolního kvartilu (Q_{25}), mediánu (Q_{50}), horního kvartilu (Q_{75}) a maxima. Lze vyznačit i hodnotu průměru (E):



Pro normální rozložení teoreticky musí být graf symetrický s mezikvartilovým rozpětím rovným zhruba 1.35-násobku směrodatné odchylky. Je-li výrazně asymetrický nebo je-li hodnota mediánu výrazně odlišná od aritmetického průměru, nemají data normální rozdělení a pokračuje se ve zkoumání anomálií podle dalšího textu.

Podrobnější pohled na charakter dat podává histogram sestavený na základě třídního rozdělení četnosti:



Histogram četnosti vypovídá např. o symetrii resp. asymetrii rozdělení dat, o jejich rozmezí, v jakých úsecích se data kumulují nebo naopak o prázdných úsecích apod. Tato výpověď však závisí podstatně na volbě počtu tříd a jejich šířkách.

Podle [1] lze stanovit vyhovující hodnoty pro šířky h třídy a počet k tříd podle variačního rozpětí

$$h = 0.08 (X_{\max} - X_{\min}) \quad k = \text{int} |(X_{\max} - X_{\min}) / h| \quad (1)$$

nebo podle rozsahu dat

$$k = \text{int} |5 \ln(n)| \quad h = (X_{\max} - X_{\min}) / k \quad (2)$$

Při zpracování např. geochemických dat lze použít

$$k = 7 \quad h = \sigma \quad (3)$$

Kontaktní údaje:

Institut geologického inženýrství, VŠB – Technická univerzita Ostrava
17. listopadu 15/2172, 708 33 Ostrava-Poruba
tel.: +420 59 699 1243, <http://geologie.vsb.cz>

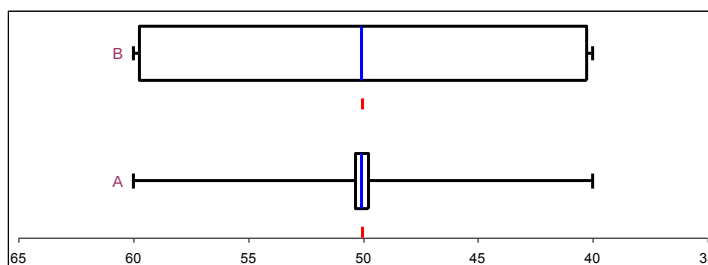
tak, že hranice tříd jsou tvořeny hodnotami $p + i \cdot \sigma$, kde p je střední hodnota, σ je směrodatná odchylka a i je z $\langle -3; 3 \rangle$. Dolní a horní mez intervalu hodnoty i se upraví podle konkrétní situace tak, že četnost ještě není nulová.

5.2. Neparametrické statistiky a box-plot

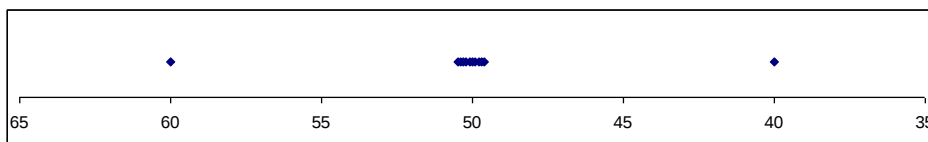
Stochastická data s potenciálními anomáliemi lze bez požadavku na jejich normalitu hodnotit neparametrickými statistikami, zvláště pak kvantily $Q_{0,25}$, $Q_{0,50}$ a $Q_{0,75}$ známé jako dolní kvartil, medián a horní kvartil. Výhodou je jejich poměrná robustnost proti odlehlým hodnotám.

Na základě uvedených kvantilů, minima a maxima lze zkonstruovat graf podávající celkový náhled na data zkoumané veličiny; graf má v originální literatuře název box-and-whiskers plot nebo jen box-plot, v české literatuře uváděný jako krabicový graf. Ukažme nyní, jakou vypovídající schopnost má v některých anomálních případech.

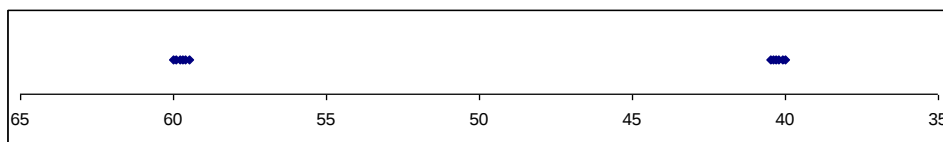
Mějme proto dva grafy A a B podle následujícího obrázku:



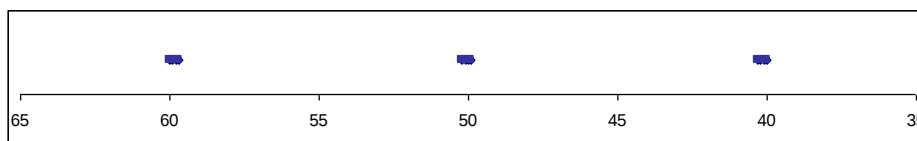
Graf A vypovídá o tom, že jistě celá polovina dat je soustředěna ve velmi úzkém hodnotovém rozpětí kolem mediánu. O rozložení čtvrtiny dat nad horním a čtvrtiny dat pod dolním kvartilem graf sice přímo nevypovídá, nejčastěji však graf A charakterizuje data, kde jen několik málo jednotlivých hodnot má extrémní hodnoty, a zbývající padnou do rozpětí jen o málo větší než je mezikvartilové rozpětí:



Graf B naopak vypovídá o tom, že jistě celá čtvrtina dat je soustředěna kolem maxima a celá čtvrtina dat je soustředěna kolem minima. O rozložení zbývajících dat mezi kvantily graf sice přímo nevypovídá, nejčastěji však graf B charakterizuje data, která vznikla „smícháním“ různých datových souborů; dvou:



nebo tří:



Kontaktní údaje:

Institut geologického inženýrství, VŠB – Technická univerzita Ostrava
17. listopadu 15/2172, 708 33 Ostrava-Poruba
tel.: +420 59 699 1243, <http://geologie.vsb.cz>

5.3. Semivariogram

Semivariogram je prostorová funkce udávající závislost kvadrátů odchylek zkoumané veličiny na vektoru polohy. Pro data získávaná na linii nebo ploše jde o závislost kvadrátů odchylek na vzájemné vzdálenosti. Logika je následující:

Zkoumejme, jak se hodnotově liší dvě hodnoty zkoumané veličiny y_i a y_j v závislosti na vzájemné vzdálenosti míst x_i a x_j , odkud hodnoty pocházejí. S ohledem na to, že jde o data stochastická, by mělo platit asi toto:

- Jestliže tato místa budou velmi blízko sobě, pak by hodnoty zkoumané veličiny měly být hodně podobné.
- Jestliže se místa budou od sebe vzdalovat, pak by se hodnoty zkoumané veličiny měly od sebe lišit tím více, čím je větší vzájemná vzdálenost ...
- ... ale zase ne až do nekonečna: místa, která jsou od sebe *hodně* vzdálená a která jsou od sebe *velmi hodně* vzdálená – tam už ten vzájemný rozdíl asi příliš nenaroste.

Mírou odlišnosti dvojice hodnot je jejich rozdíl. Průměrná hodnota kvadrátu vzájemných odchylek několika dvojic je *variance* (srov. *rozptyl*) a poloviční hodnota tohoto průměru je *semivariance*. Označíme-li semivarianci g , je

$$g(h) = \frac{1}{2 \cdot n} \sum_{|x_i - x_j| = h} (y_i - y_j)^2 \quad (4)$$

Graf funkce g se označuje jako *semivariogram*.

Pro diskrétní data nebývá obecně $\Delta_{ij} = |x_i - x_j| = k$, $k=1, 2, \dots$. V tom případě se postupuje následovně:

- $g(h)$ se určí pro $h=k \cdot h_0$, $k=1, 2, \dots$
- Dvojice y_i a y_j se v součtu (4) uplatní tehdy, je-li pro dané h (a tedy k takové, že $h=k \cdot h_0$) např.

$$\Delta_{ij} \in I_1 = ((k-1) \cdot h_0; k \cdot h_0) \text{ nebo } \Delta_{ij} \in I_2 = \left(\left(k - \frac{1}{2}\right) \cdot h_0; \left(k + \frac{1}{2}\right) \cdot h_0 \right) \quad (5)$$

- h_0 je stanoveno na základě konkrétních dat, konkrétní situace nebo požadavku analýzy většinou tak, aby do intervalu I_1 resp. I_2 padla alespoň jedna dvojice (x_i, x_j) .

Ukažme, jak se v semivariogramech projeví anomálie spojené s (např. geologickými) poruchami.

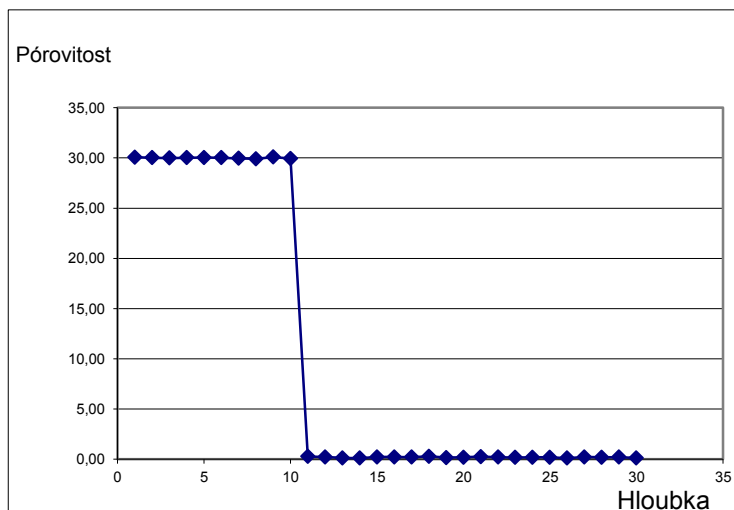
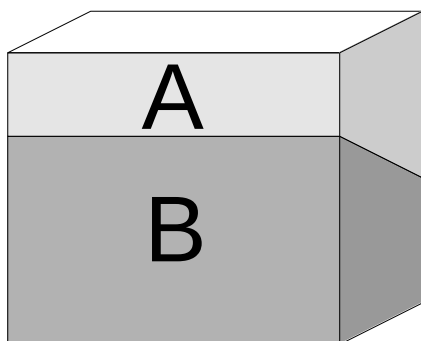
Mějme data popisující nějakou sledovanou veličinu vázanou na typ horniny. Mějme podloží tvořené dvěma horninovými bloky s oddělující plochou. Necht' data v hornině A mají výrazně odlišnou hodnotu než v hornině B. Může jít např. o sledování pórovitosti, přičemž A mohou být písky a B jíly.

Předpokládejme pro jednoduchost lineární strukturu tvořenou např. vrtem; hloubka, ze které byl odebrán vzorek, hloubka tedy je nezávisle proměnná x , pórovitost je závisle proměnná y .

Data proto mají jistou zhruba konstantní hodnotu pórovitosti až do hloubky, kde se písky a jíly stýkají. Od této hloubky pórovitost prakticky zcela vymizí. Situaci zachycuje následující obrázek vlevo, grafické znázornění dat podává obrázek vpravo.

Kontaktní údaje:

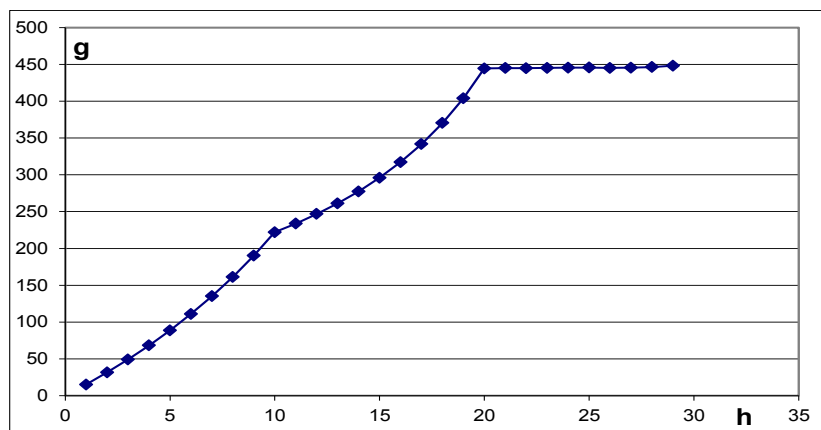
Institut geologického inženýrství, VŠB – Technická univerzita Ostrava
17. listopadu 15/2172, 708 33 Ostrava-Poruba
tel.: +420 59 699 1243, <http://geologie.vsb.cz>



Aplikujme nyní vztah (4). Necht' je h relativně malá vzdálenost. Pak v součtu (4) převažují členy s velmi blízkými hodnotami (a tedy $x_i - x_j$ je velmi malé) až na případ, že x_i bude v A a x_j v B nebo naopak: zde bude $x_i - x_j$ velké – avšak čím menší bude h , tím více bude malých Δ_{ij} a méně velkých a tedy průměr bude spíše menší (opačně se zvětšujícím se h).

Jestliže ovšem h přesáhne velikost většího úseku, pak je zaručeno, že vždy patří jedna hodnota x_i do A a druhá x_j do B a proto všechny sčítance v (4) budou mít maximální (a se zvětšujícím se h téměř konstantní) hodnotu; od této hodnoty se tedy dá očekávat konstantní hodnota semivariance.

Skutečný semivariogram dat z předchozího obrázku je tento:



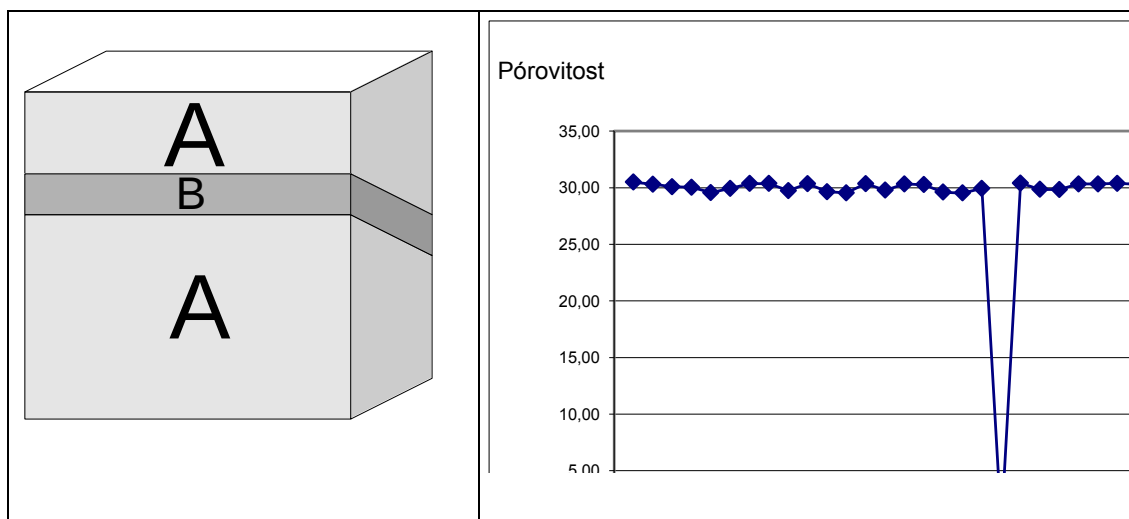
Na druhé straně mějme jiný stav. Necht' je jednolitý horninový masiv A porušen velmi úzkou vrstvou B. Necht' stejně jako v předchozím případě jsou A písky, B jíly, sledovaná veličina necht' je opět pórovitost. Situaci naznačuje obrázek vlevo. Data odpovídající naznačenému případu jsou na grafu vpravo:

Kontaktní údaje:

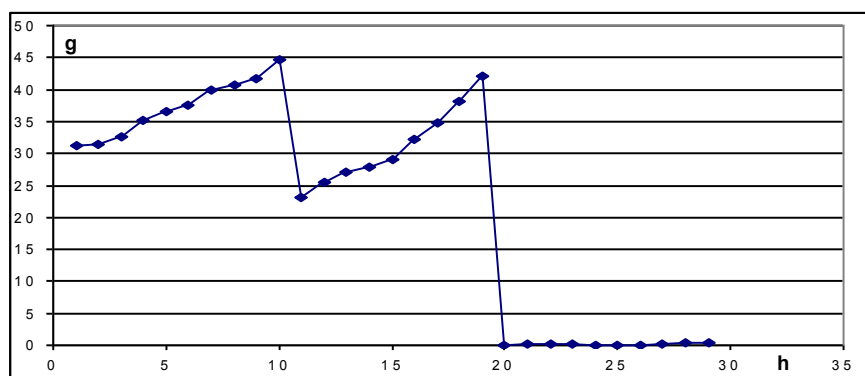
Institut geologického inženýrství, VŠB – Technická univerzita Ostrava

17. listopadu 15/2172, 708 33 Ostrava-Poruba

tel.: +420 59 699 1243, <http://geologie.vsb.cz>



Experimentální semivariogram odpovídající tomuto druhému druhu poruchy následuje:



Postup popsaný v předchozích odstavcích lze beze změny převést z lineárního na plošný problém. Odpovídá to vzorkům odebíraných na ploše (v terénu) v místech daných souřadnicemi $[x,y]$ a analyzovaných na sledovanou veličinu z . Vztah (4) platí analogicky po záměně y za z s tím, že vzdálenost dvou míst je spočtena podle Pythagorovy věty z rozdílu souřadnic x a y .

5.4. Mapa izolinií

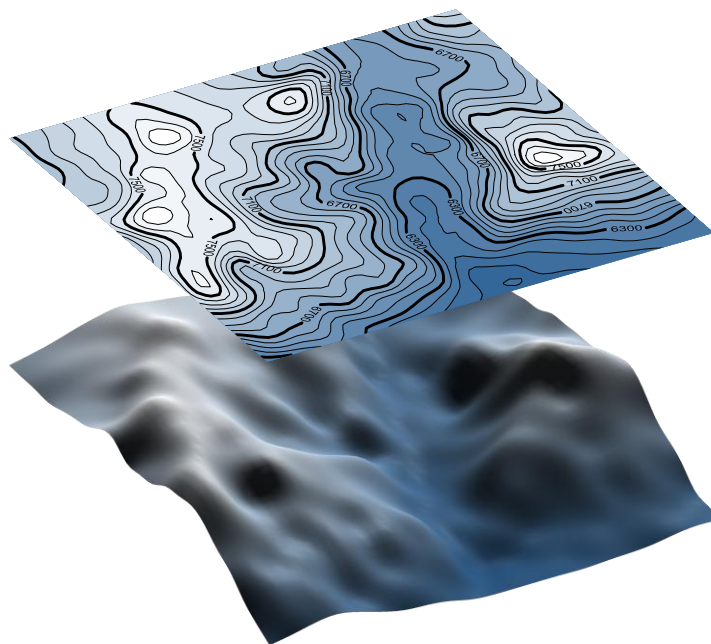
Zkoumaná veličina (např. konkrétní škodlivá látka) je na sledovaném území plošně distribuována. Nazíráno z trojrozměrného grafu, je tato distribuce znázorněna plochou, na níž lokální maxima představují vysoké hodnoty, lokální minima nízké hodnoty. Predikce anomálií zkoumané veličiny je tedy otázkou zkoumání tvaru této plochy.

Mapa izolinií, často také podle způsobu „ruční“ tvorby nazývaná konturová mapa, je dvourozměrná reprezentace trojrozměrných dat. Izolinie pak znázorňuje množinu bodů se stejnou hodnotou sledované veličiny v zájmovém území. Jinak také, jde o průsečnice modelované plochy s rovinou rovnoběžnou s rovinou XY v dané vzdálenosti od ní.

Názorně to ukazuje následující obrázek (převzato z [2]):

Kontaktní údaje:

Institut geologického inženýrství, VŠB – Technická univerzita Ostrava
17. listopadu 15/2172, 708 33 Ostrava-Poruba
tel.: +420 59 699 1243, <http://geologie.vsb.cz>



Základní úlohou je tedy vytvoření modelu této plochy. Je zřejmé, že není možno získat každý z jejího nekonečně mnoha jednotlivých bodů. Metodika navrhuje získat alespoň konečný počet jejích bodů tak, že jejich průmět do roviny XY tvoří pravidelnou obdélníkovou síť. I těchto bodů je však oproti známým reálným hodnotám (navíc odebíraným z nepravidelně rozmístěných lokalit) značný počet. Pro získání hodnot v uzlech sítě se proto použije některá z metod geostatistických odhadů. Následná konstrukce izoliní je pak jednoduchá geometrická úloha.

Jednou z metod nejčastěji používaných v geostatistických aplikacích pro bodové odhady hodnot je *krigování*. Pro její použití je zapotřebí jako její parametr zadat model semivariogramu (proto viz význam odstavce výše).

Krigování je jednou z metod, kterou by bylo možno označit jako "příspěvková". Hodnotu sledované veličiny v místě X si lze představit jako souhrn příspěvků z jednotlivých známých míst do X. Jednotlivé příspěvky jsou závislé především na jednotlivých známých hodnotách, a dále na vzdálenostech vzájemných i od bodu X, na struktuře sledované veličiny na celé ploše a dalších okolnostech.

Krigování proto stanoví odhad v místě X jako součet vážených známých hodnot

$$Z_X = \sum_{i=1}^n \mu_i \times Z_i \quad (6)$$

kde na váhy μ_i je kladena podmínka normalizace

$$\sum_{i=1}^n \mu_i = 1 \quad (7)$$

Odhad stanovený vztahem (6) se ovšem obecně liší od skutečné (nezměřené, nám neznámé) hodnoty sledované veličiny. Rozdíl mezi těmito dvěma hodnotami je nazýván *chybou odhadu*. Existuje jediná n-tice koeficientů μ_i taková, že chyba odhadu je právě pro tuto n-tici minimální. Určení těchto koeficientů je založeno na minimalizaci rozptylu odhadů σ^2 :

Kontaktní údaje:

Institut geologického inženýrství, VŠB – Technická univerzita Ostrava
17. listopadu 15/2172, 708 33 Ostrava-Poruba
tel.: +420 59 699 1243, <http://geologie.vsb.cz>

$$\sigma^2 = K_{X,X} - 2 \times \sum_i \mu_i \times K_{X,i} + \sum_i \sum_j \mu_i \times \mu_j \times K_{i,j} \quad (8)$$

v místě X , přičemž $K_{i,j}$ jsou kovariance mezi i -tým a j -tým známým bodem vázané se semivariancemi vztahem

$$K_{i,j} = g(\infty) - g(h_{i,j}) \quad (9)$$

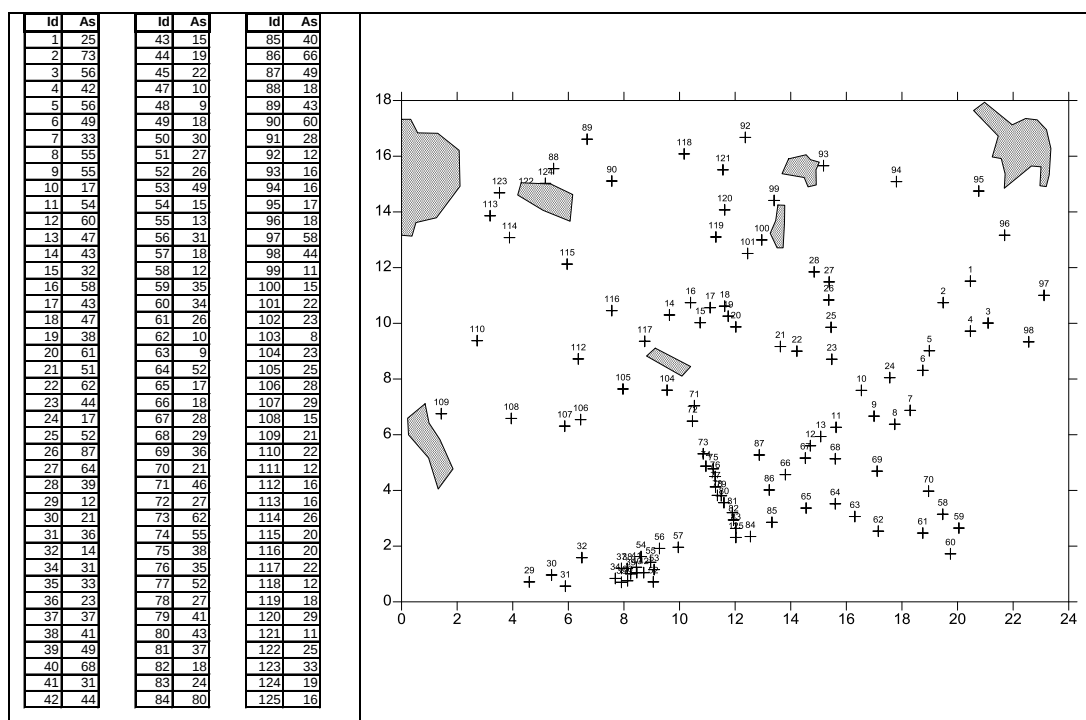
a hodnoty $K_{X,X}$ resp. $K_{X,i}$ jsou vypočítávány dle (9) z hodnot semivariací aproximovaných pro vzdálenosti $h_{X,X}$ resp. $h_{X,i}$.

6. Příklad použití

Jako příklad pro dokumentaci předchozích odstavců bylo zvoleno zkoumání anomálií v rozložení arzenu v půdách v terénu o rozloze cca. 24 x 18 kilometrů.

6.1. Zdrojová data

Bylo provedeno 123 odběrů vzorků půd, které byly následně analyzovány na arzen. Výsledky obsahující identifikaci odběrného místa a množství arzenu v [mg/kg] jsou uvedeny v následující tabulce vlevo, prostorová lokalizace míst pak na obrázku vpravo:

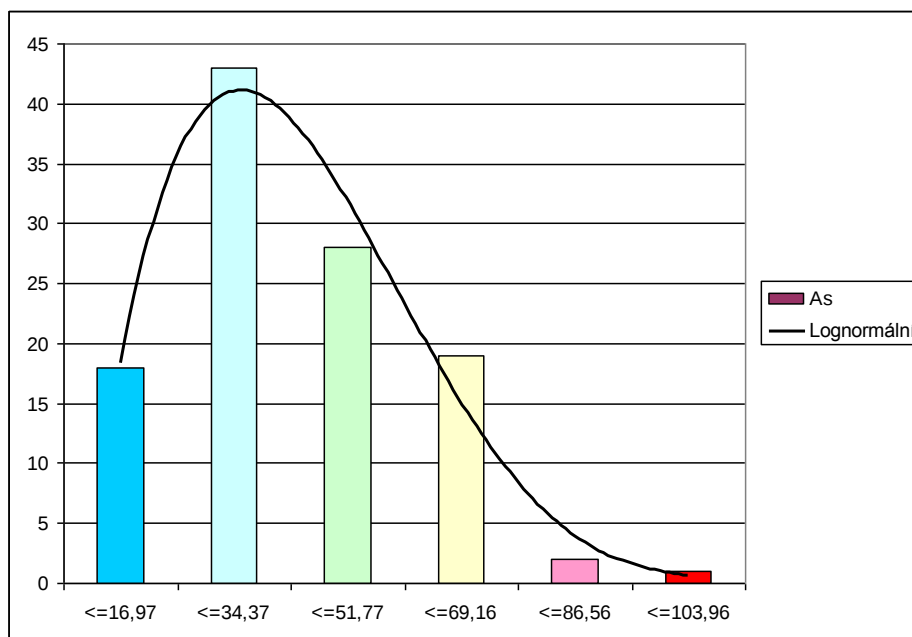


6.2. Rozložení četnosti

Střední hodnota dat As je $\bar{x} = 34,369$, směrodatná odchylka je $\sigma = 17,347$. Podle 5.1. byly jako hranice tříd zvoleny násobky směrodatné odchylky před a za průměrem, celkem 6 tříd. Histogram četnosti je uveden na následujícím obrázku:

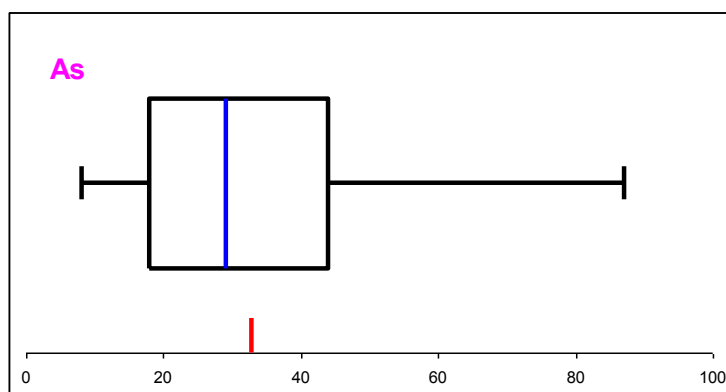
Kontaktní údaje:

Institut geologického inženýrství, VŠB – Technická univerzita Ostrava
17. listopadu 15/2172, 708 33 Ostrava-Poruba
tel.: +420 59 699 1243, <http://geologie.vsb.cz>



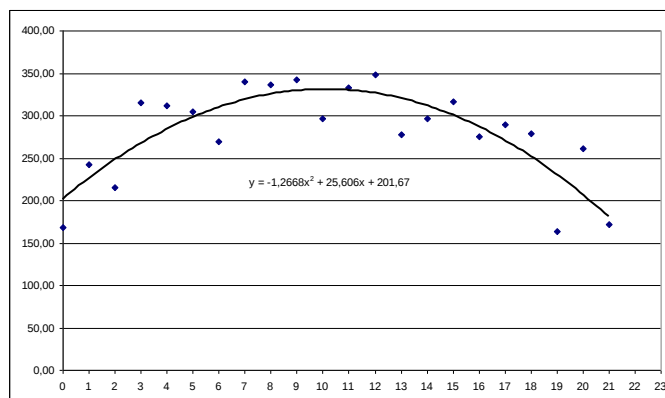
6.3. Krabicový graf

Krabicový graf odpovídá skutečně zhruba lognormálnímu rozložení a je uveden v následujícím obrázku:



6.4. Semivariogram

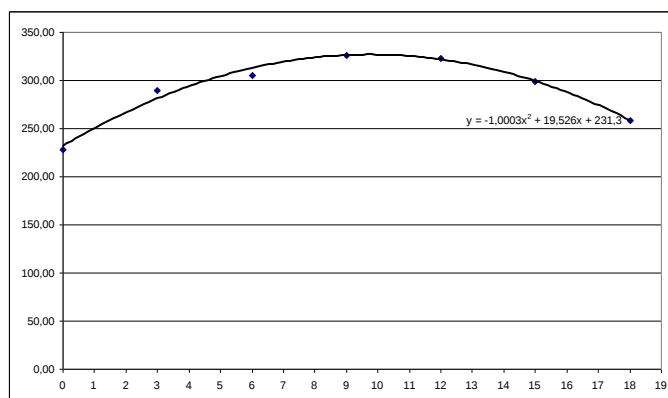
Podle odst. 5.3 je semivariogram předmětných dat pro $h_0 = 1$ následující:



Kontaktní údaje:

Institut geologického inženýrství, VŠB – Technická univerzita Ostrava
 17. listopadu 15/2172, 708 33 Ostrava-Poruba
 tel.: +420 59 699 1243, <http://geologie.vsb.cz>

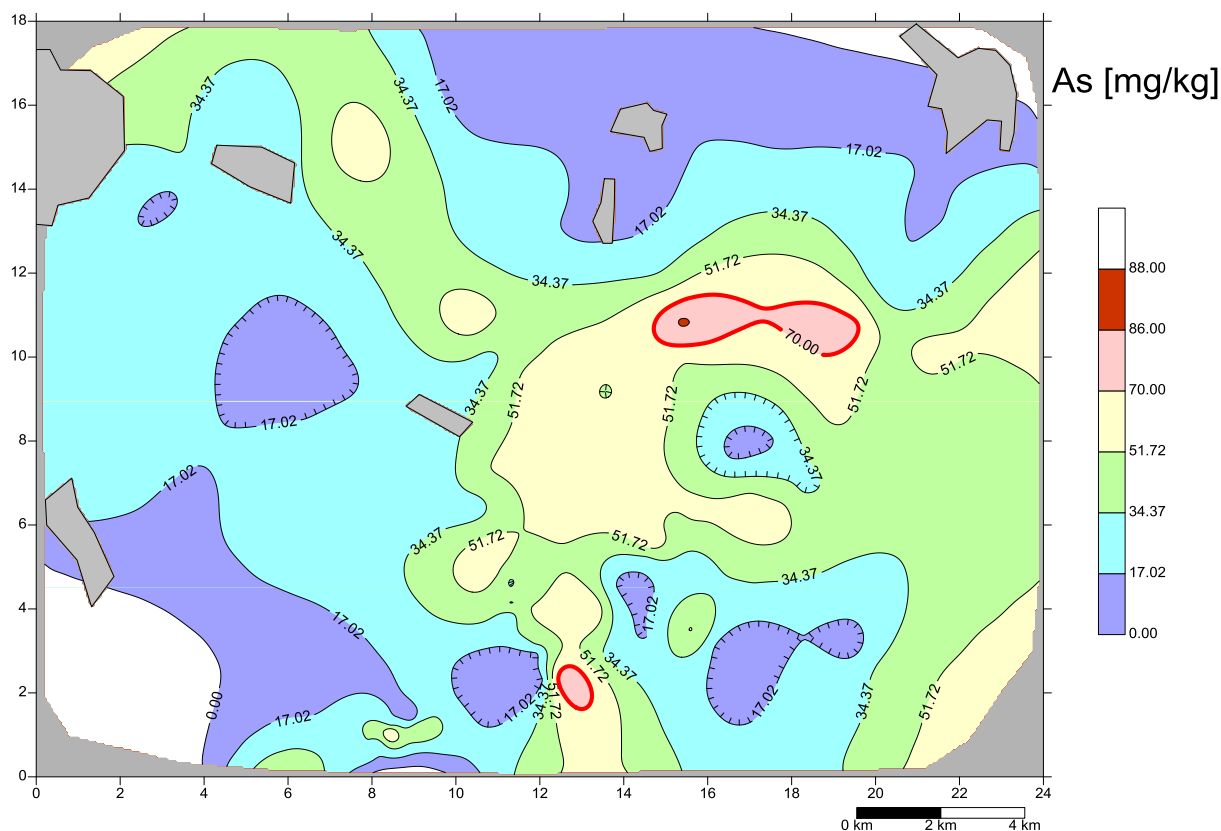
zatímco pro $h_0 = 3$ následující:



lze tedy zvolit jako teoretický semivariogram polynomický 2. stupně se zbytkovým rozptylem 223 (vyplývá z analýzy dalších hodnot $h_0=1,2,3,4$).

6.5. Mapa izoliní

Mapa izoliní (viz shora 5.4) je následující (barevné škálování je ve shodě s výše uvedeným histogramem četnosti):

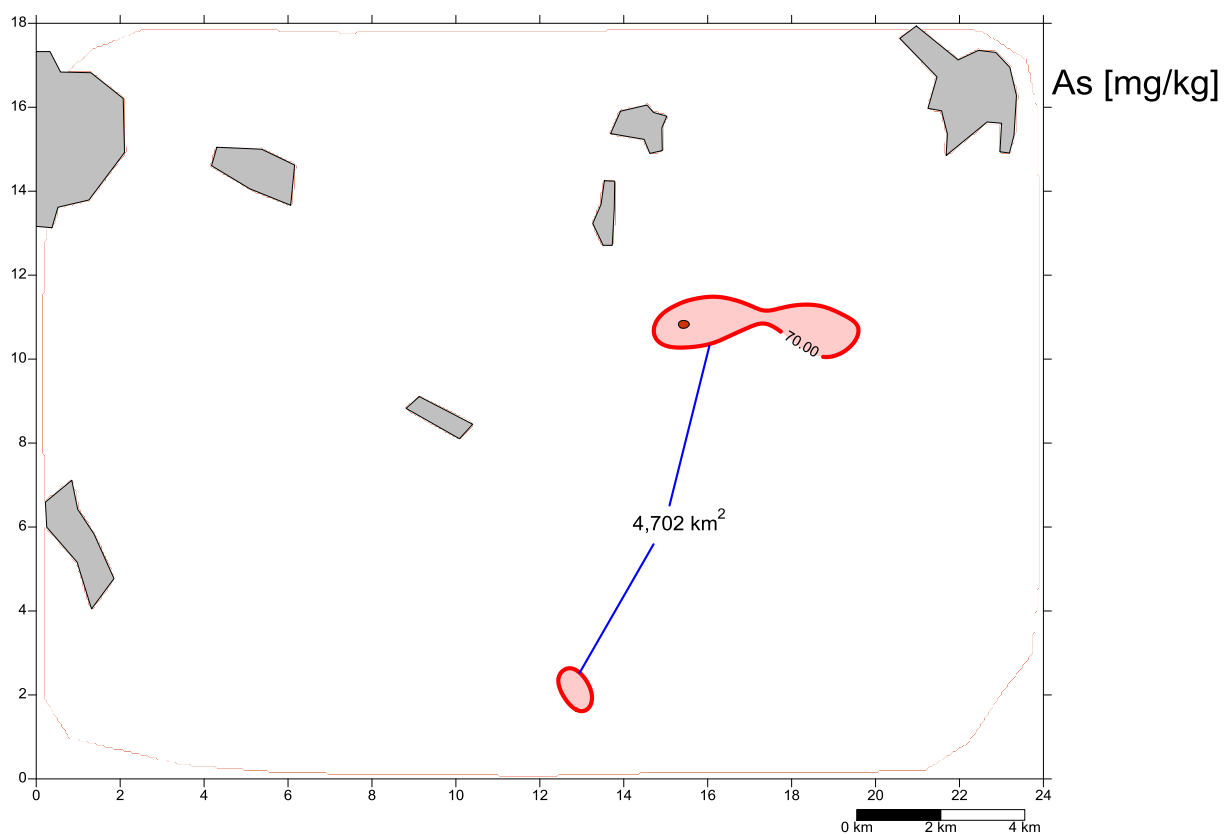


Mapa izoliní zřetelně vypovídá o anomálních hodnotách nejen přímo v konkrétním místě o souřadnicích [15.4 ; 10.8], ale i ve všech místech terénu, ve kterých je modelovaná hodnota větší než zadaný limit - zde hodnota 70 [mg/kg]. Zadaným limitem ovšem může být např. hranice dekontaminace pro staré ekologické zátěže nebo mnohé další indikační hodnoty.

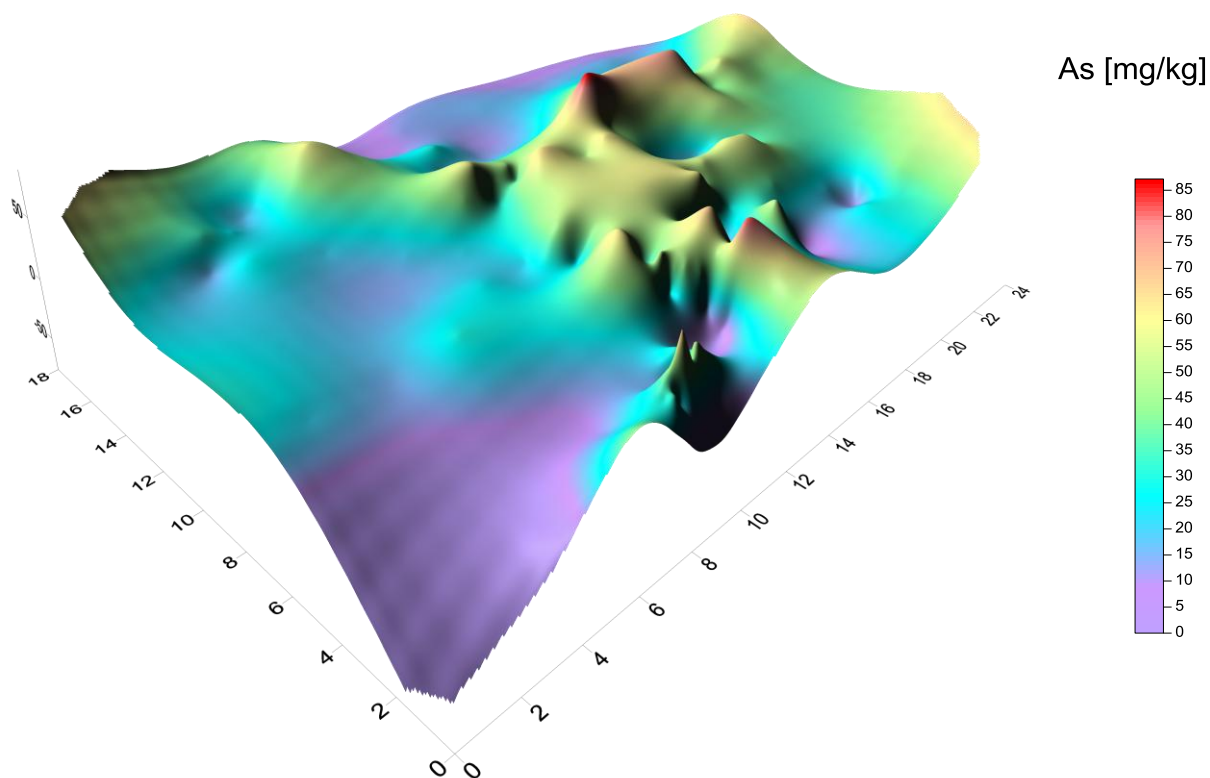
Jako jeden z vedlejších výstupů může být informace o celkové ploše terénu s nadlimitním obsahem:

Kontaktní údaje:

Institut geologického inženýrství, VŠB – Technická univerzita Ostrava
17. listopadu 15/2172, 708 33 Ostrava-Poruba
tel.: +420 59 699 1243, <http://geologie.vsb.cz>



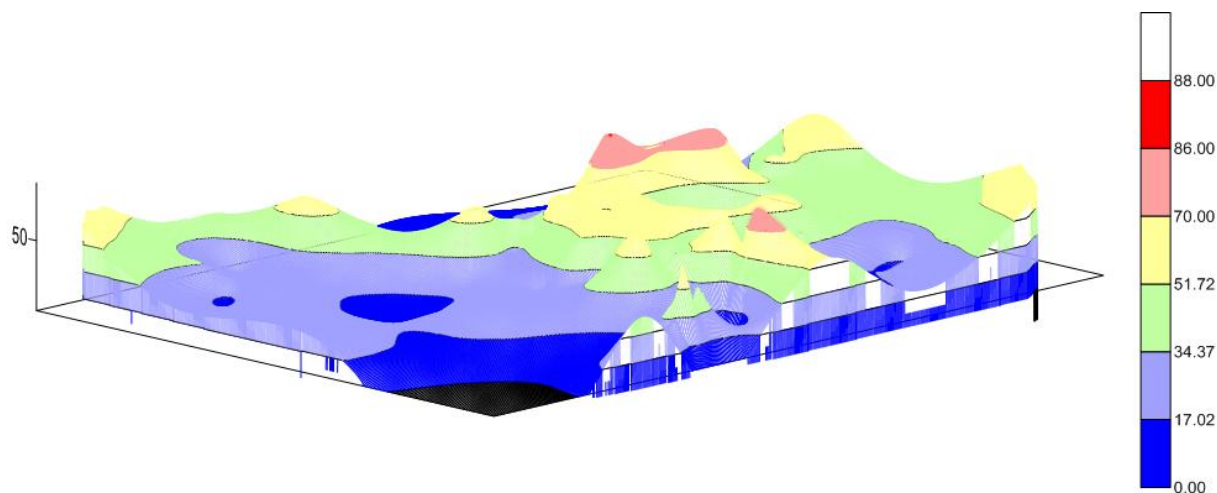
Na modelovanou plochu lze nazírat i prostorově jako na stínovaný reliéf:



Kontaktní údaje:

Institut geologického inženýrství, VŠB – Technická univerzita Ostrava
17. listopadu 15/2172, 708 33 Ostrava-Poruba
tel.: +420 59 699 1243, <http://geologie.vsb.cz>

nebo jako nespojitě vybarvený prostorový model:



Oba posledně uvedené obrázky slouží pouze pro rychlou orientaci, nehodí se např. pro přesné odečty.

7. Závěr

Uvedená metodika navrhuje standardní postup pro zpracování dat environmentální geochemie, kontaminace životního prostředí a environmentálních rizik. Je zejména vhodná pro zpracování údajů s plošnou distribucí, která je vyjádřena v topografickém zobrazení např. mapou izolinií. Metodika ukazuje význam posouzení statistického rozdělení dat v souboru pro volbu postupu dalšího zpracování.

Poděkování

Metodika byla podpořena z prostředků MŠMT ČR – OpVaVpi CZ. 1.05/2.1.00/03.0069.

Literatura

- [1] Schejbal, C. - Homola, V. – Staněk, F.: *Geoinformatika*. Pont, s.r.o. , Košice (2004), ISBN 80-967611-8-8.
- [2] *Surfer 8: User's Guide*. Golden Software, Inc. Colorado (2002).

Kontaktní údaje:

Institut geologického inženýrství, VŠB – Technická univerzita Ostrava
17. listopadu 15/2172, 708 33 Ostrava-Poruba
tel.: +420 59 699 1243, <http://geologie.vsb.cz>



ZDRAVOTNÍ ÚSTAV se sídlem v Ostravě

PARTYZÁNSKÉ NÁM. 7

702 00 OSTRAVA

Potvrzení

Zdravotní ústav se sídlem v Ostravě potvrzuje, že „Metodika pro statistické zpracování dat k identifikaci anomálií škodlivin v životním prostředí“ je nástrojem pro optimalizaci postupu k hodnocení anomálií. Metodika vychází z geochemické definice prahu anomálie a je využitelná i v případě, že nejsou definovány legislativní limity.

-- Metodika spočívá ve vyjmenování závazných kroků při zpracování a hodnocení množiny dat sledovaných látek pro získání definovaného výsledku. Jako zdrojová data metodika určuje logický formát tabulky relační databáze, v jednom nebo více sloupcích obsahující hodnoty sledované / sledovaných veličin, v jednom nebo ve dvou sloupcích jejich lineární nebo plošnou polohu. Na takto definovaných datech stanovuje metodika posloupnost kroků data zpracovávajících. Jde o analýzu jejich variability, prostorového rozložení, lokálních odhadů a dalších. Kroky jsou stanoveny za účelem odhalení potenciálních anomálií sledovaných veličin ve sledovaném lineárním intervalu nebo plošné oblasti.

-- Metodika statistického zpracování dat vychází z následujících testů a postupů: rozložení četnosti, základní neparametrické statistiky, krabicový graf, semivariogram, izolinie.

Vznik metodiky byl podpořen projektem MŠMT ČR – OpVaVpi CZ. 1.05/2.1.00/03.0069.

Metodika nebyla zpracována za finanční podpory Zdravotního ústavu se sídlem v Ostravě.

-- V Ostravě 22.11.2011

Zdravotní ústav se sídlem v Ostravě

Partyzánské nám. 7, 702 00 Ostrava

IČ: 71009396

23

RNDr. Petr HAPALA, ředitel
Zdravotní ústav se sídlem v Ostravě