



Ministerstvo životního prostředí

Metodická příručka MŽP

Hodnocení průzkumu a sanací

Ministerstvo životního prostředí

2011

OBSAH:

1.	Úvod	5
2.	Obecné základy statistického hodnocení dat	7
2.1.	Základní statistické pojmy	7
	Statistická data, výběrový soubor	7
	Statistické hodnocení dat	7
	Reprezentativnost statistických dat	8
	Třídění dat statistických souborů, histogramy	8
	Frekvenční a distribuční funkce	10
	Základní charakteristiky statistických souborů	10
	Rozdělení (distribuce) statistického souboru	11
	Základní statistické parametry souborů	12
	Směrodatná odchylka	13
	Vztah základní - výběrový soubor	13
	Krabicové grafy	14
3.	Aplikace statistických metod v kontaminační geologii a hydrogeologii	15
3.1.	Obecný postup při statistickém zpracování dat	15
3.2.	Základní rozdělení statistických metod	15
3.3.	Testy rozdělení statistických souborů	15
3.4.	Statistická nezávislost dat	17
3.5.	Úprava datových souborů s velkým počtem nízkých hodnot	18
3.6.	Odlehlé hodnoty	19
3.7.	Testování hypotéz	19
3.8.	Testy významnosti	22
3.9.	P-hodnoty	22
3.10.	T-testy	23
3.11.	Multi-srovnávací strategie	23
3.12.	Statistické intervaly	24
3.13.	Časové řady	28
3.14.	Závislost dvou náhodných kvantitativních veličin - regrese a korelace	30
3.15.	Geostatistika	31
3.16.	S-křivky	32

Metodická příručka
Hodnocení průzkumu a sanací

3.17. Shluková analýza	33
3.18. Rekapitulace použití statistických metod v kontaminační geologii a hydrogeologii ...	35
4. Strategie a metodika vzorkování	39
4.1. Konceptuální model lokality	39
4.2. Strategie vzorkování	39
4.3. Vzorkovací schémata	41
Jednoduché náhodné vzorkování	41
Systematické náhodné vzorkování (v síti)	42
Stratifikované náhodné vzorkování	44
4.4. Kompozitní (směsné) vzorkování	45
4.5. Reprezentativní koncentrace	46
4.6. Průměrovací - expoziční plocha a její velikost	48
4.7. Vzorkování zemin ve vertikálním profilu	50
4.8. Ohnisko kontaminace	50
4.9. Pozad'ové hodnoty	51
4.10. Zajištění jakosti vzorkovacích prací	52
4.11. Vzorkovací plán	53
5. Hustota vzorkování, počty vzorků	53
5.1. Minimální počet vzorků pro statistické hodnocení dat	53
Výkopy	54
Vzorkování hromad zemin při ex-situ dekontaminaci	55
Vzorkování zemin při sanaci in situ	55
Počet vzorků pro porovnání výsledků sanace s pozadím	55
5.2. Minimální počet vzorků při úsudkovém vzorkování	56
5.3. Předpisové – smluvní - požadavky na minimální počty vzorků	57
Minimální plošná hustota vzorkovací sítě při předběžném průzkumu lokality	57
Hodnocení úspěšnosti sanace nebo přípustnosti určitého funkčního využívání lokality porovnáváním s limity, odvozenými ve vztahu k ochraně zdraví či ekosystémů	57
Porovnávání kontaminace zemin nesaturované zóny s limitem (kritériem) pro zeminy z hlediska přestupu do podzemních vod	58
Porovnávání kontaminace zemin s limitem (kritériem) pro zeminy z hlediska přímého kontaktu či ingesce	58
Porovnávání kontaminace půdního vzduchu s kritérii kontaminace pro půdní vzduch z hlediska inhalace	59

Metodická příručka
Hodnocení průzkumu a sanací

Minimální počty vzorků pro hodnocení koncentrace ve dně a stěnách sanačních výkopů...	59
Minimální počty vzorků pro charakteristiku zemin na hromadách	60
Lokalizace a počty vzorků při sanaci zemin in situ	61
5.4. Monitoring podzemních vod	61
Detekční monitoring	61
Kontrolní monitoring	61
Sanační monitoring	62
Postsanační monitoring	62
Pozadíový monitoring	62
5.5. Statistické metody hodnocení monitoringu podzemních vod	62
Obecné požadavky na počty a frekvenci vzorků při monitoringu podzemních vod	62
Minimální požadavky na hustotu vzorkování při statistickém hodnocení trendů kontaminace v podzemní vodě	63
Minimální požadavky na počet vzorků podzemních vod pro aplikaci různých statistických metod	63
Porovnávání kontaminace podzemní vody s limity podle analýzy rizika	63
Porovnávání kontaminace podzemní vody vůči pozadí	64
Porovnávání kontaminace podzemních vod s kritérii pro ochranu podzemní vody	64
Porovnávání kontaminace podzemních vod s kritérii pro ochranu povrchových vod	65
Porovnávání obsahu těkavých složek v podzemní vodě s kritérii kontaminace podzemních vod z hlediska inhalace	65
Základní zásady monitoringu přírodní atenuace	65
6. Výběr literatury	67

Metodická příručka Hodnocení průzkumu a sanací

Metodická příručka MŽP pro hodnocení průzkumu a sanací byla zpracována sekcí technické ochrany životního prostředí pod vedením ing. Karla Bláhy, CSc., a PhDr. Ivo Hlaváče na odboru ekologických škod pod vedením RNDr. Pavly Kačabové, ředitelky odboru. Odbornými garanty zpracování byli ing. David Topinka a Mgr. Ivana Vávrová.

Metodická příručka byla zpracována ing. Jiřím Tylčerem.

Cennými radami i přímo doplněním některých kapitol přispěl ke zpracování této příručky jako oponent RNDr. František Pastuszek.

1. ÚVOD

Aplikace statistických metod pro hodnocení kontaminačních průzkumů a sanací horninového prostředí předpokládá dodržení řady poměrně přísných požadavků na metodiku vzorkování, které musí být zohledněny již v projektu průzkumných prací.

Intenzivní prosazování statistických přístupů v kontaminační geologii je spojeno s přelomem 80tých a 90tých let minulého století, především v návaznosti na rozvoj tohoto oboru v USA po přijetí zákona CERCLA (Superfund). Publikace americké federální agentury US EPA z těchto let (EPA 1989a, 1989b, 1992a, 1992b, 1992c, 1992d) jsou dodnes základními a vyčerpávajícími materiály k tomuto tématu.

Doposud však rozvoj používání statistických metod nenaplnil původní očekávání. Příčiny jsou mimo jiné v tom, že pokud neslevíme z požadavků na spolehlivost, znamená jejich aplikace zpravidla významně vyšší náklady na vzorkování a analýzy oproti zavedeným zvyklostem. Mnohdy nelze garantovat splnění teoretických předpokladů použití statistiky v důsledku složitého charakteru přírodního prostředí a složité distribuce kontaminace na znečištěných lokalitách.

V praxi průzkumů a sanací horninového prostředí se tak běžně vyskytují situace, kdy by použití přísně statistických metod pro zpracování a hodnocení získaných dat nebylo korektní a mohlo by vést k zásadně chybným závěrům.

V neposlední míře je zde bariérou vzdělanostní profil typického řešitele průzkumných a sanačních akcí. Ve většině by důsledná aplikace statistických přístupů vyžadovala spoluúčast specializovaného profesionála a to nejen při hodnocení výsledků, ale již ve fázi projektové přípravy. Nekorektní využívání statistiky by jinak mohlo přinést více škod než užitku.

Na druhé straně je právě z těchto důvodů žádoucí, aby alespoň základní představu o principech, možnostech a omezeních statistického hodnocení dat měl každý, kdo se ve své práci posuzováním kontaminace horninového prostředí zabývá, ať již z jakéhokoli úhlu pohledu.

Požadavkům praxe na jednodušší přístupy ke vzorkování a hodnocení dat při průzkumech a sanacích vychází vstřícně různé směrnice a jiné metodické materiály, zaváděné v některých environmentálně vyspělých zemích od druhé poloviny devadesátých let. Vedle USA lze příkladem uvést Holandsko, Velkou Británii, Kanadu, Austrálii.

Ve většině jsou si tyto směrnice pojetím a často i v detailech velmi blízké. Stanovují zásady vzorkování a sjednocují zásady zpracování a interpretace dat jak pro přístupy s využitím

Metodická příručka Hodnocení průzkumu a sanací

statistických metod, tak pro vzorkování na základě odborného úsudku. Zpravidla směrnice určují i minimální počty vzorků v závislosti na vzorkovaném médiu, vzorkovací strategii, rozsahu prací a charakteru řešeného problému. Mnohdy jsou explicitně vyjmenovány i požadované či doporučené statistické metody a podmínky jejich aplikace - zpravidla s tím, že odůvodněně a se souhlasem příslušného úřadu lze použít i metody jiné. V některých státech jsou k dispozici i podrobné manuály pro aplikaci takových směrnic v praxi.

Hodnota těchto směrnic a manuálů je v tom, že se již osvědčily jako pragmatický, avšak kvalifikovaný kompromis, vyhovující podmínkám a možnostem běžné praxe průzkumů a sanací. Nejsou v žádném případě v rozporu s teoretickými předpoklady aplikace statistických postupů, které však nejsou považovány za jediné použitelné.

Tyto koncepční přístupy byly inspirací i pro předkládanou metodickou příručku, která má kompilační charakter a je založena na podrobné rešerši podkladů výše diskutovaného druhu. V závěru příručky je uveden výčet použité a doporučené literatury. Zahraniční legislativní normy samotné však v tomto přehledu uváděny nejsou. Pokud je na ně v textu odkazováno, pak jen stručně, bez plných citací.

Příručka si klade za cíl, být referenční pomůckou při výběru metodiky a rozsahu vzorkování a při volbě metodiky hodnocení získaných dat, při hodnocení dosažení cílů sanace i při projektování a hodnocení výsledků průzkumů znečištění zemin a podzemních vod.

Příručka nemůže suplovat chybějící závaznou směrnici vzorkování pro naši českou praxi. Na to je příliš rozsáhlá - preferencí byla snaha poskytnout základní poučení o tématu v celé jeho komplexnosti a na obecnější úrovni. Je zařazena i kapitola s pokusem o populární vysvětlení základních principů statistického hodnocení dat tak, aby příručka mohla být pomocí i těm, jejichž vzdělanostní profil je naprosto odlišný.

Na druhé straně bylo usilováno o maximální možnou stručnost a a přehlednost. Příručka se proto nezabývá technickými aspekty správné vzorkovací praxe, nýbrž jen problematikou strategie vzorkování. Ze stejného důvodu neaspiruje tato příručka na to, aby byla podrobným manuálem pro statistické výpočty. K dispozici je v tomto směru buď řada učebnic obecného charakteru nebo ještě lépe některé z publikací z připojeného seznamu literatury, které jsou zaměřeny přímo na hodnocení průzkumů a sanací horninového prostředí. Navíc, k běžné dispozici je dnes řada počítačových programů pro statistické hodnocení dat. K seznamu literatury je připojen i výběr některých ověřených statistických programů, které jsou bezúplatně ke stažení z Internetu. Širokou škálou běžnějších statistických operací disponuje i MS Excel.

2. OBECNÉ ZÁKLADY STATISTICKÉHO HODNOCENÍ DAT

Účelem textu této kapitoly je:

- poskytnout základní informaci o pojmech a principech statistického zpracování a hodnocení dat těm, kteří se při své práci setkávají s výsledky těchto postupů, avšak jejich vzdělanostní základ je zcela odlišný,
- prezentace referenčního přehledu základních výpočtových vzorců pro pracovníky, kteří statistické zpracování dat při vyhodnocování průzkumů horninového prostředí provádějí nebo výsledky takového hodnocení musí z různých úhlů posuzovat a využívat, avšak tato činnost pro ně není vzhledem k jejich vzdělanostnímu profilu úplně rutinní záležitostí.

Je pravděpodobné, že pro většinu čtenářů je tato kapitola zbytečná a mohou ji přeskočit.

2.1. Základní statistické pojmy

Statistická data, výběrový soubor

Statistická data jsou data, která máme o zkoumaném objektu či jevu k dispozici a která splňují určitá kritéria, podmiňující možnost jejich zpracování metodami matematické statistiky. Ta využívá aparátu teorie pravděpodobnosti.

Statistická data mohou být úplná a zahrnovat celou základní **populaci** (čili **základní soubor**). Při průzkumech horninového prostředí však bývá pravidlem, že k dispozici máme jen omezený soubor dat, jen určitý počet výsledků laboratorních analýz. Hovoříme pak o **výběrovém souboru** nebo též o **vzorku základní populace**. Počet prvků výběru, počet prvků výběrového souboru (v našem případě počet analýz) „n“ se nazývá **rozsah výběru**. Z jednoho základního souboru můžeme provést více takových výběrů. Výběry nemusí být shodného rozsahu.

Zkoumání populací pomocí výběrů je typické pro většinu oborů – například volební průzkumy se také neprovádějí šetřením u všech občanů republiky, ale jen u početně omezeného vzorku.

Zkoumáním a hodnocením výběrových souborů se snažíme o co nejreprezentativnější postižení vlastností základních souborů.

Statistické hodnocení dat

Statistické hodnocení dat využívá aparátu teorie pravděpodobnosti a slouží v praxi k těmto základním účelům:

- **odhady** – z dat výběrového souboru odhadujeme na vlastnosti celé populace,
- **testování hypotéz** – ze souboru dat který máme k dispozici usuzujeme, zda zkoumaný jev či objekt vykazuje určitou vlastnost nebo zda má některé shodné prvky s jinými soubory,
- **modelování** – soubor či soubory získaných dat slouží ke zkoumání vztahů mezi různými vlastnostmi či parametry zkoumaného jevu či objektu.

Typickým příkladem odhadu je stanovení průměrné hodnoty koncentrace kontaminantu v zeminách zkoumané lokality. Výsledky statistického zpracování analýz odebraných vzorků

vztahujeme na celou lokalitu (tj. na celou populaci) - stanovenou průměrnou koncentraci považujeme za reprezentativní pro celou lokalitu.

Příkladem testování hypotéz je, když na základě výsledků analýz odebraných vzorků podzemních vod posuzujeme zda bylo dosaženo cílů sanace, zda je koncentrace daného kontaminantu pod sanačním limitem.

Nejjednodušším druhem statistického modelování je regresní analýza, kdy se snažíme o stanovení závislosti mezi dvěma vlastnostmi zkoumaného objektu, snažíme se formulovat hypotézu o vztazích mezi nimi. Příkladem může být odvození grafu, vyjadřujícího na základě výsledků laboratorních analýz vztah mezi obsahem určitého kontaminantu v zemině a v půdním vzduchu.

Složitějším typem statistického modelování je například vykreslování map kontaminace s využitím tzv. krigování, jak to provádí známý program Surfer.

Reprezentativnost statistických dat

Reprezentativnost statistických dat posuzujeme podle toho, s jakou výstižností či spolehlivostí výsledky hodnocení výběrových souborů charakterizují vlastnosti celé populace (celého základního souboru). Na reprezentativnost statistických dat má vliv více různorodých faktorů:

Statistická náhodnost vzorků výběrového souboru je základním předpokladem reprezentativnosti dat, ze kterých budeme usuzovat na vlastnosti celé populace. Při vzorkování zemin musí mít každé místo lokality stejnou pravděpodobnost, že z něj bude odebrán vzorek pro laboratorní analýzu. Pokud použita **strategie vzorkování** zemin nebo podzemních vod vyhovuje tomuto základnímu kritériu, hovoříme o statisticky náhodném vzorkování.

Statistická nezávislost dat: výběrový soubor který má sloužit k odhadu velikosti určitého parametru populace nesmí obsahovat data, která by byla mezi sebou nějak závislá. Například stanovení průměrné koncentrace škodliviny v podzemní vodě v areálu lokality by mohlo být zavádějící, pokud koncentrace podléhá sezónním výkyvům a hodnocené vzorky z vrtů by nepocházely ze shodného data (resp. z užšího časového intervalu).

Metody vzorkování a analýz rovněž ovlivňují reprezentativnost statistického hodnocení zásadním způsobem. V některých případech by bylo velmi problematické porovnávat výsledky analýz vzorků, odebraných či analyzovaných různými metodami. Reprezentativnost ovlivňují i **chyby** při vzorkování, při manipulaci se vzorky, při jejich transportu, při jejich uchovávání v laboratoři, při samotných analýzách. Chyby tohoto druhu lze dělit na nahodilé a systematické. Eliminaci a minimalizaci chyb tohoto druhu má za úkol **kontrola kvality** vzorkování a analýz. Pro posouzení kvality vzorkování a analýz slouží kontrolní odběry a analýzy vzorků a porovnávací statistické testy výsledků.

Při zajištění všech ostatních předpokladů je pro reprezentativnost statistického hodnocení dat rozhodující **rozsah výběrových souborů**. S rostoucím počtem odebraných a analyzovaných vzorků zemin nebo podzemních vod vzrůstá přesnost statistického stanovení vypočtených reprezentativních hodnot (např. průměrů). Lze vypočíst, **s jakou pravděpodobností (spolehlivostí)** se bude například vypočtený průměr koncentrace nacházet při daném počtu vzorků uvnitř intervalu určité velikosti.

Třídění dat statistických souborů, histogramy

Základní statistické hodnocení dat výběrových souborů je založeno na jejich uspořádání a roztřídění podle velikosti a podle četnosti výskytu každé z naměřených hodnot. Grafickým

znázorněním tohoto třídění jsou základní histogramy (frekvenční) a histogramy kumulativních četností (distribuční). Princip histogramů vysvětluje následující obrázek č.1.

Často se naměřená data pro účely statistického zpracování třídí do velikostních tříd. Obrázek číslo 2 vysvětluje princip takového třídění. Pro příklad byla použita primární naměřená data z tabulky obrázku č. 1, velikost třídního intervalu byla zvolena 5 mg/kg.

**Obr. 1. ROZTŘÍDĚNÍ DAT VÝBĚROVÉHO SOUBORU
A GRAFICKÁ PREZENTACE VÝSLEDKŮ VZORKOVÁNÍ**

souhrnná datová tabulka

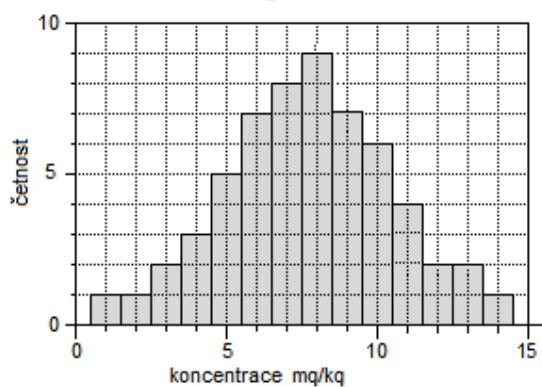
naměřená data			
naměřené koncentrace mg/kg	četnost hodnot	kumulativní četnost	
		absolutní	relativní %
0	0	0	0
1	1	1	2
2	1	2	4
3	2	4	7
4	3	7	12
5	5	12	21
6	7	19	33
7	8	27	47
8	9	36	62
9	7	43	74
10	6	49	84
11	4	53	91
12	2	55	95
13	2	57	98
14	1	58	100
15	0		

celkem 58

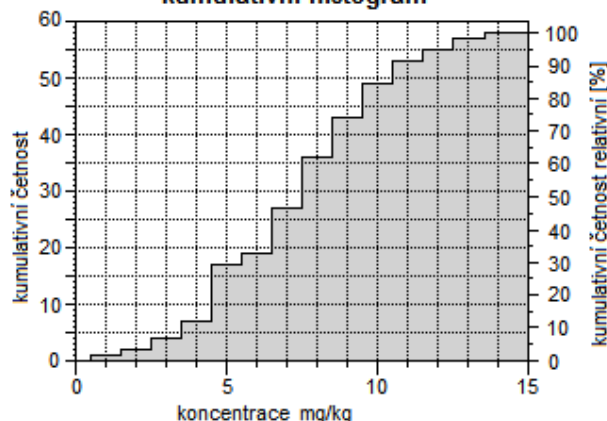
Aritm. průměr = 7,7 mg/kg

Medián = 8 mg/kg

základní histogram četnosti



kumulativní histogram



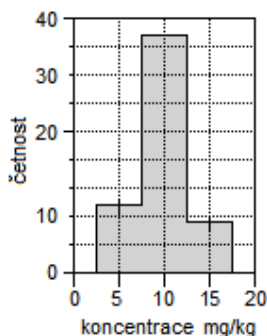
Obr. 2. USPOŘÁDÁNÍ DAT DO TŘÍD

souhrnná datová tabulka

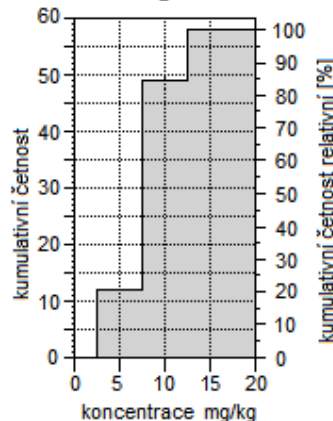
změřená data			
třída (interval koncen- trace) mg/kg	četnost hodnot	kumulativní četnost	
		absolutní	relativní %
0 - 5	12	12	21
5 - 10	37	49	84
10 - 15	9	58	100

celkem 58

základní histogram
četnosti



kumulativní
histogram



Jde o tentýž datový soubor jako v obrázku č. 1.

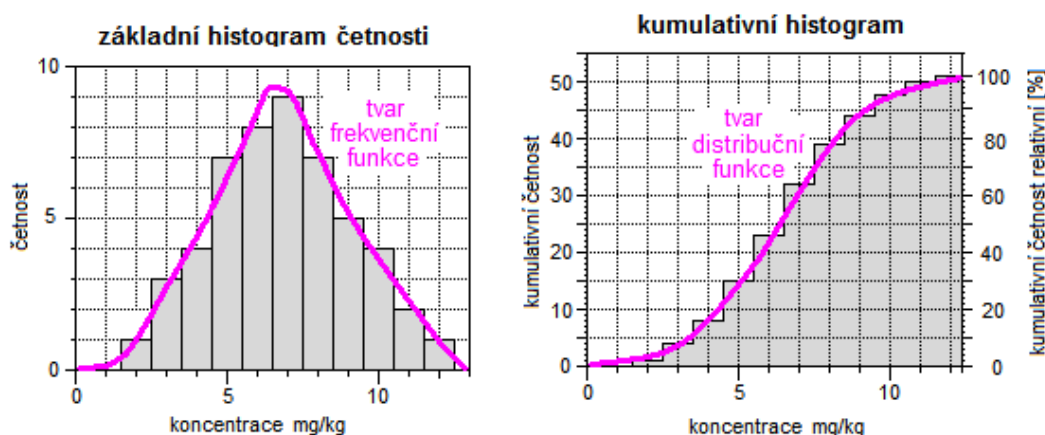
Frekvenční a distribuční funkce

Histogramy statistických souborů mívají typický tvar, se kterým souvisí:

- tak zvaná frekvenční funkce spojité náhodné veličiny v případě základního (frekvenčního) histogramu
- tak zvaná distribuční funkce spojité náhodné veličiny v případě kumulativního (distribučního) histogramu.

Následující obrázek č. 3 naznačuje souvislost mezi tvary histogramů a tvarem frekvenční a distribuční funkce.

**Obr. 3. VZTAH MEZI HISTOGRAMEM A TVREM
FREKVENČNÍ A DISTRIBUČNÍ FUNKCE**



Základní charakteristiky statistických souborů

Charakteristiky statistického souboru jsou číselné hodnoty které výstižně charakterizují (reprezentují) celý soubor a jeho základní vlastnosti. Rozlišujeme charakteristiky rozsahu souboru, charakteristiky polohy a charakteristiky rozptylu:

- charakteristika rozsahu = počet prvků výběrového souboru,
- charakteristiky polohy – jsou to střední hodnoty jako průměr, medián, modus,
- charakteristiky rozptylu (čili variability) - indikují spolehlivost charakteristik polohy.

Je evidentní, že jinak budeme nahlížet na informaci, že průměrná koncentrace škodliviny v zemině je 20 mg/kg a výsledky jednotlivých analýz se pohybují v rozpětí od 18 do 22 mg/kg, nebo že průměr je tentýž, ale výsledky jednotlivých analýz se pohybují v rozpětí od 5 do 50 mg/kg.

Mezi charakteristiky variability patří především: rozptyl (=variance) a směrodatná odchylka (=deviace), dále variační rozpětí, variační koeficient.

Dalšími, často používanými charakteristikami souborů jsou například percentily (též zvané kvantily), konfidenční intervaly, atd.

V dalším textu uvedená tabulka č.1. uvádí vzorce pro výpočty základních charakteristik statistických výběrových souborů, popřípadě i některá vysvětlení k nim.

Tabulkové procesory – např. Excel – umí provádět výpočty těchto i dalších charakteristik automaticky.

Rozdělení (distribuce) statistického souboru

Rozdělení statistického souboru charakterizuje, jakým způsobem jsou jednotlivé hodnoty rozloženy kolem hodnoty střední, jaký tvar má histogram četností výskytu jednotlivých hodnot, jaký tvar má frekvenční funkce (viz obrázek 4).

Podle charakteru svého rozdělení lze naprostou většinu souborů dat náhodných veličin reálného světa roztrždit do několika základních skupin.

Základní význam mají tak zvané **rozdělení normální čili Gaussovo** a **rozdělení lognormální (LN rozdění)**.

- **Normální rozdění:** graf frekvenční funkce tohoto rozdělení má typický zvonovitý tvar, souměrný kolem střední hodnoty (přičemž vodorovné osy se oba konce křivky dotýkají až v $\pm$ nekonečnu). V praxi nejběžnější typ rozdění.

Normální rozdění má mezi všemi rozděními výsadní postavení. Normálnímu rozdění se totiž za určitých podmínek blíží i jiná rozdění náhodných veličin a to tím více, čím je rozsah výběrových souborů větší (tzv. zákon velkých čísel, centrální limitní teorém¹). Z toho rovněž vyplývá, že pro soubory s velkým počtem dat jsou menší odchylky od normální distribuce akceptovatelné.

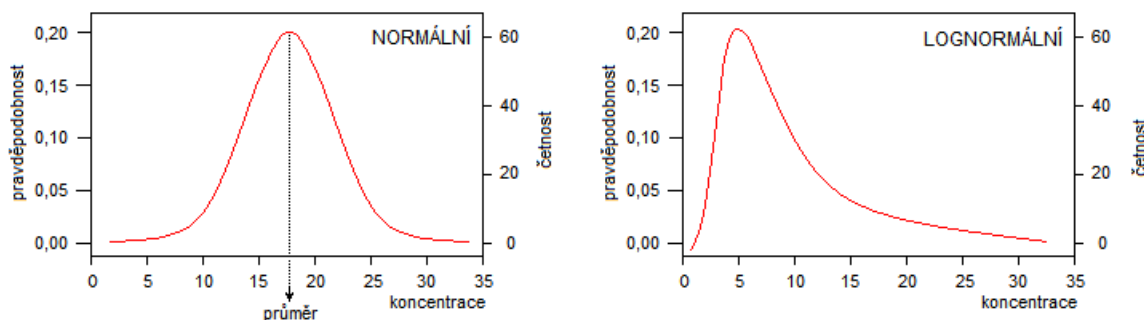
- **Lognormální rozdění:** výskyt určitého podílu výrazně vyšších hodnot způsobuje, že graf frekvenční funkce je nesymetrický, protažený silně doprava.

Důležitou vlastností lognormálního rozdění je, že logaritmy originálních dat mají rozdění normální. Výpočty statistických charakteristik lognormálních souborů jsou založeny na této vlastnosti. Do vzorců pro výpočty charakteristik souborů s normálním rozdělením se dosazují logaritmy naměřených dat.

V environmentálních disciplínách se vyskytuje lognormální rozdění poměrně často, existuje pro to teoretické odůvodnění. Lognormální rozdění se často vyskytuje právě u kontaminace horninového prostředí.

Pro výpočty charakteristik souborů s lognormálním rozdělením nelze použít data s nulovými hodnotami (logaritmus nuly neexistuje).

Obr. 4. NORMÁLNÍ A LOGNORMÁLNÍ ROZDĚLENÍ



Neparametrické soubory jsou takové, u kterých charakter rozdění dat nedokážeme definovat. Zpravidla tomu tak bývá kvůli malému počtu dat - při nedostatečném rozsahu souboru. Často je příčinou odchylek souborů od normálního rozdění vyšší výskyt

¹ Centrální limitní teorém (zjednodušeně) říká, že pokud ze základního souboru provedeme větší počet výběrů většího rozsahu ($n \geq 30$), pak rozdění průměrů těchto výběrů bude mít normální rozdění bez ohledu na charakter rozdění základního souboru,

Metodická příručka Hodnocení průzkumu a sanací

extrémních (odlehklých hodnot) a vyšší výskyt hodnot velmi nízkých (pod mezí detekce). Tyto jevy mohou být způsobeny tím, že datový soubor obsahuje víc než jednu populaci (část dat pochází z ohnisek znečištění, druhá část z nekontaminovaného okolí).

Pro testování dat neparametrických souborů se nemohou používat standardní testy pro parametrické soubory. Obecně vyžaduje využívání neparametrických metod pro dosažení stejné úrovně spolehlivosti větší soubory dat než v případě souborů parametrických.

Základní statistické parametry souborů

Tab.1.	Výpočet základních charakteristik výběrových souborů, jejich interpretace					
charakteristika	normální (N) rozdělení			lognormální (LN) rozdělení		
průměr	Aritmetický průměr: $A = \bar{x} = \frac{1}{n} \sum x_i$			Geometrický průměr: $G = \bar{x}_g = e^{\varphi} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$ $\varphi = \ln G = \ln \bar{x}_g = \frac{1}{n} \sum \ln x_i$		
medián	$\tilde{x} (= x_{50})$ = prostřední z hodnot výběru, uspořádaného podle velikosti. Polovina (50%) dat výběru je menších nebo rovno $\tilde{x}$, 50% dat je větších. Medián není tolik ovlivňován krajními hodnotami jako průměr. Při sudém počtu dat výběru je medián aritmetickým průměrem dvou středních hodnot. Výrazný rozdíl mezi průměrem a mediánem indikuje nesymetrické rozdělení souboru dat nebo vliv tzv. odlehklých hodnot (hodnot abnormálně se odchylujících od hodnot ostatních). U LN rozdělení při větším rozsahu souboru medián = geometrický průměr.					
modus	Nejčastěji se vyskytující hodnota souboru (hodnota s nejvyšší četností).					
rozptyl (variance) (= průměr čtverců odchylek od průměru hodnot)	$s^2 = \frac{1}{n} \cdot \sum (x_i - \bar{x})^2 = \frac{\sum x_i^2}{n} - \bar{x}^2$ Nezaokrouhlovat zde $\bar{x}$, má to velký vliv na přesnost určení s^2			$\eta = s_g^2 = (\ln SG)^2 = \frac{1}{n} \cdot \sum (\ln x_i - \ln \bar{x}_g)^2 =$ $= \frac{1}{n} \cdot \sum \left(\ln \frac{x_i}{\bar{x}_g} \right)^2 = \frac{\sum (\ln x_i)^2}{n} - \left(\frac{\sum \ln x_i}{n} \right)^2$		
směrodatná (standardní) odchylka (deviace)	$SD = s = \sqrt{s^2}$			$s_g = \ln SG = \sqrt{\eta} = \sqrt{s_g^2} = \sqrt{(\ln SG)^2}$ $SG = e^{\sqrt{\eta}} = e^{\sqrt{s_g^2}} = e^{\sqrt{(\ln SG)^2}}$		
variační koeficient	$V = \frac{s}{\bar{x}}$			$V = \frac{\sqrt{\eta}}{\varphi}$		
rozpětí	Rozdíl mezi nejvyšší a nejnižší hodnotou datového souboru: $R = x_{\max} - x_{\min}$					
percentil (kvantil)	$x_p, x(p)$ Uvedené procento „p“ dat souboru je menších nebo rovno $x(p)$, (100-p)% dat je větších.					
interval kolem střední hodnoty (průměru), ve kterém leží určité procento p všech dat výběrového souboru	meze intervalu: $\left[\bar{x} - (u_p \cdot s) \right] \text{ až } \left[\bar{x} + (u_p \cdot s) \right]$			logaritmy hodnot mezí intervalu: $\left[\varphi - (u_p \cdot \sqrt{\eta}) \right] \text{ až } \left[\varphi + (u_p \cdot \sqrt{\eta}) \right]$		
	u _p ve vzorcích nabývá těchto hodnot:			hodnoty mezí intervalu:		
	p %	99	95,4	95	90	80
	u _p	2,576	2,00	1,96	1,645	1,282
	p %	75	70	68,2	60	50
u _p	1,15	1,046	1,00	0,84	0,676	
			čili: $\left(\frac{G}{SG^{u_p}} \right) \text{ až } \left(G \cdot SG^{u_p} \right)$			

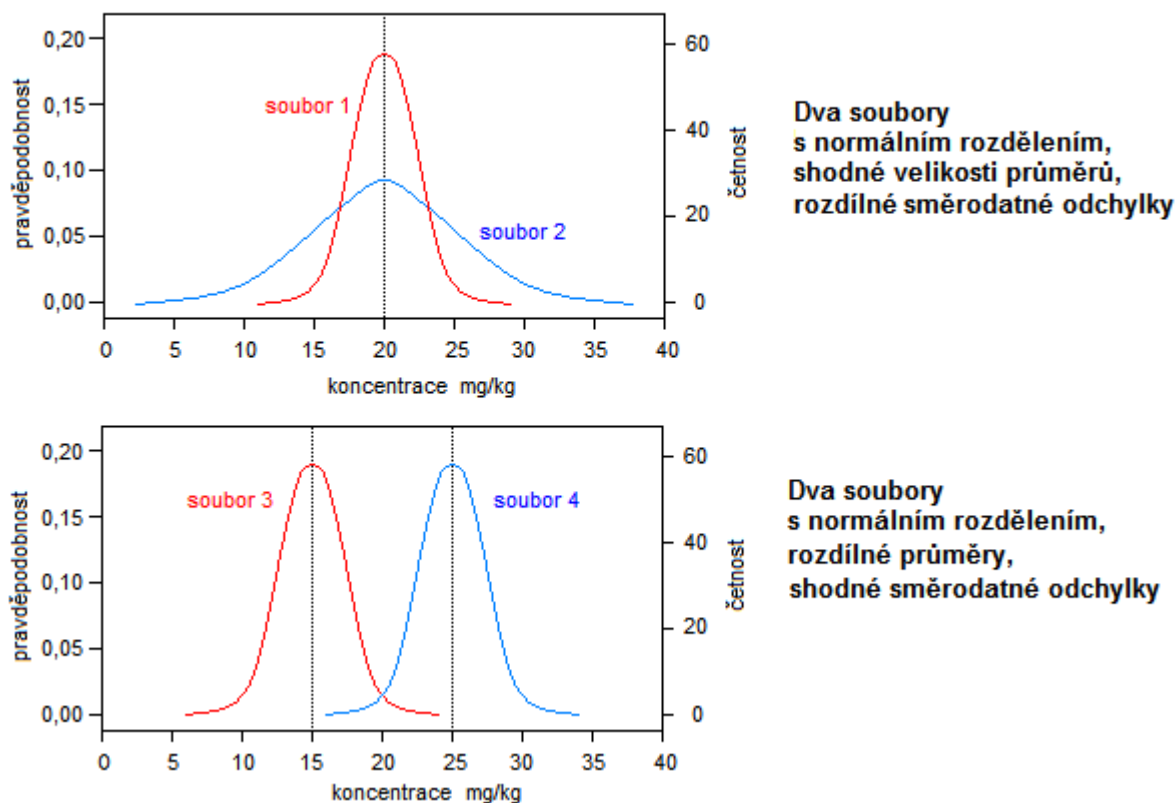
Průměr a směrodatná odchylka vstupují prakticky do všech výpočtů při testování souborů parametrickými metodami.

Zmínku si zde zaslouží upozornění na to, že při výpočtech reprezentativních koncentrací pro hodnocení on-site expozice při analýzách humánního rizika používáme aritmetický průměr bez ohledu na skutečný charakter rozdělení souboru dat ze vzorkování. Aritmetický průměr je reprezentativní hodnotou ve vztahu k náhodnému pohybu potenciálního příjemce kontaminace v areálu lokality.

Směrodatná odchylka

Směrodatná (standardní) odchylka je měřítkem disperze dat – toho, jak dalece se jednotlivé hodnoty odchylují od průměru. Ilustrativně vysvětluje tento význam směrodatné odchylky následující obrázek č. 5.

obr. 5. VÝZNAM SMĚRODATNÉ ODCHYLKY



U normálního rozdělení platí, že 68,2% všech dat souboru se odchyluje od průměru nanejvýš o hodnotu směrodatné odchylky (tzv. centrální rozpětí), 68% všech dat je v intervalu $(\bar{x} \pm s)$.

V intervalu do dvou směrodatných odchylek na každou stranu od průměru se potom nachází kolem 95% všech dat souboru. Jen 2,5% dat je menších než $(\bar{x} - 2s)$ a jen 2,5% dat je větších než $(\bar{x} + 2s)$.

Vztah základní - výběrový soubor

Ze základního souboru lze vytvořit více souborů výběrových. Parametry jednotlivých výběrových souborů se mohou navzájem lišit.

Metodická příručka Hodnocení průzkumu a sanací

Při průzkumu kontaminace horninového prostředí by takových výběrů mohlo být v podstatě nekonečně mnoho. Výsledky dvou etap vzorkování na téže lokalitě se nebudou v praxi nikdy shodovat.

Základní soubor je tudíž reprezentován výběrem jen s omezenou přesností.

Vlastnosti (parametry) základního souboru musíme odhadovat z parametrů výběru, který máme k dispozici.

Následující tabulka uvádí přehled charakteristik základního a výběrového souboru:

Tab. 2. Charakteristiky základního a výběrového souboru				
Parametry →	četnost	průměr	rozptyl	směrodatná odchylka
neznámé parametry základního souboru	N	μ	δ^2	δ
parametry studovaného výběru	n	$\bar{x}$	s^2	s
		$\bar{x} = \frac{1}{n} \cdot \sum x_i$	$s^2 = \frac{1}{n} \cdot \sum (x_i - \bar{x})^2$	$s = \sqrt{s^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$
odhad parametrů základního souboru z dat výběru		$\hat{\mu}$	$\hat{\delta}^2$	$\hat{\delta}$
		$\hat{\mu} = \bar{x} = \frac{1}{n} \cdot \sum x_i$	$\hat{\delta}^2 = s^2 \cdot \frac{n}{n-1} = \frac{1}{n-1} \cdot \sum (x_i - \bar{x})^2$	$\hat{\delta} = \sqrt{\hat{\delta}^2} = s \cdot \sqrt{\frac{n}{n-1}} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$

Krabicové grafy

Krabicové grafy se často používají pro grafické zobrazení základních charakteristik statistických souborů. Poskytují názornou představu o jejich centrální tendenci a variabilitě.

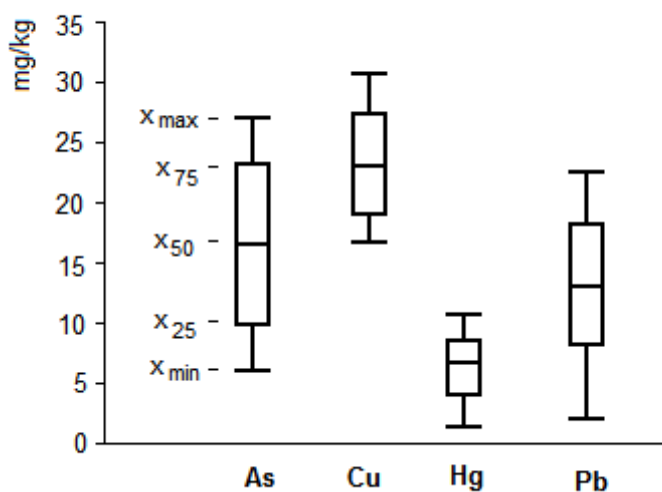
V krabicovém grafu jsou zobrazeny:

- percentil 25%
- percentil 50% (=medián)
- percentil 75%
- minimální hodnota souboru
- maximální hodnota souboru.

Připojený obrázek č. 6 vysvětluje konstrukci krabicového grafu a současně uvádí příklad jeho použití pro názorné zobrazení a porovnání výsledků vzorkování zemin na vybrané těžké kovy.

U datových souborů s normálním rozdělením rozděluje medián „krabici“ na dvě stejné poloviny.

Obr. 6. Krabicové grafy



3. APLIKACE STATISTICKÝCH METOD V KONTAMINAČNÍ GEOLOGII A HYDROGEOLOGII

3.1. Obecný postup při statistickém zpracování dat

Reprezentativnost statistických výpočtů je podmíněna dodržáním předpokladů pro použití jednotlivých metod. Než začneme s výpočty tak musíme prověřit, zda nashromážděná data odpovídají předpokladům. Nejdůležitějšími faktory které musí být otestovány jsou:

- charakter rozdělení dat,
- statistická nezávislost dat.

Následně se musíme vyrovnat s těmito problémy:

- prověření extrémních - odlehlých hodnot,
- výskyt nízkých hodnot.

Dalším krokem je výpočet základních charakteristik polohy a rozptylu pro prošetřovaný soubor (průměr, medián, směrodatná odchylka, popřípadě percentily, popř. konfidenční intervaly).

Následuje provedení různých statistických testů podle povahy řešeného problému.

Konečným krokem je interpretace výsledků a formulování závěrů.

3.2. Základní rozdělení statistických metod

Obecně lze statistické metody rozdělit do dvou základních kategorií:

- parametrické,
- neparametrické.

Volba použití parametrické či neparametrické metody závisí na charakteru rozdělení souboru, na počtu dat a na podílu hodnot pod mezí detekce.

Parametrické metody předpokládají normální resp. lognormální distribuci dat a nezávislost dat. Při porovnávání souborů musí být přibližně shodný jejich rozptyl.

Neparametrické metody pro porovnávání souborů předpokládají přibližně shodný tvar distribuce a přibližně shodné rozptyly.

Neparametrické metody nejsou závislé na normální distribuci dat (charakter distribuce není třeba testovat a data není nutno transformovat), jsou odolnější k výskytu odlehlých (extrémních) hodnot a lze je použít i tehdy, když je větší podíl dat pod mezí detekce.

Některá literatura nedoporučuje používat při hodnocení kontaminace ani logaritmickou transformaci dat. Jako nejrobustnější metodu pro testování jakýchkoliv souborů s nesymetrickým rozdělením doporučuje aplikaci Chebychevova teorému (viz Cl:aire, 2008).

3.3. Testy rozdělení statistických souborů

Prvním krokem jakéhokoli statistického zpracování dat musí být výběr správného distribučního modelu - charakter rozdělení studovaného souboru musí být otestován. Při testu

provádíme vlastně vyrovnání statistických dat - histogramem četností se snažíme proložit vhodnou teoretickou křivku.

Pro toto testování rozdělení se obecně doporučuje (např. DEQ, 2002), aby prošetřovaný soubor měl **nejméně 9 hodnot**. Testy by sice šlo provádět i se 3 daty, jejich průkaznost by však byla velmi nízká.

Obecně je postup tento:

- Některým testem prověříme, zda data vyhovují kritériu normálního rozdělení.
- Když ne, tak provedeme znovu test na normalitu se zlogaritmovanými daty. Když zlogaritmovaná data vyhovují, má soubor lognormální distribuci. V tom případě data zlogaritmuje a s takto upravenými daty dále pracujeme jako s normálním souborem².
- Nemají-li data ani normální, ani lognormální distribuci, pak musíme buď hledat další vyhovující teoretické rozdělení nebo musíme použít neparametrické metody hodnocení, založené vesměs na třídění dat podle velikosti.

V odborné literatuře lze nalézt různé metody pro testování charakteru rozdělení souborů (testování normality). nejběžnějšími jsou tyto:

Účinnou prověrkou charakteru rozdělení souboru je **vykreslení tzv. pravděpodobnostního grafu** pro normální nebo zlogaritmovaná data prošetřovaného souboru.

Výpočet koeficientu šikmosti je málo pracným a jednoduchým postupem (automaticky jej umí i MS Excel). Koeficient šikmosti γ_1 nám indikuje asymetrii dat vzhledem ke středu distribuce:

$$\gamma_1 = \frac{\frac{1}{n} \sum (x_i - \bar{x})^3}{\left(\frac{n-1}{n} \right)^{\frac{3}{2}} \cdot s^3}$$

Data z normální distribuce mají $\gamma_1 = 0$, asymetrické distribuce mají $\gamma_1 \neq 0$, podle toho zda je pravý ($\gamma_1 > 0$) nebo levý ($\gamma_1 < 0$) „ocas“ distribuční křivky delší a tenčí než ten druhý. Lognormální distribuce má γ_1 pozitivní ($\gamma_1 > 0$).

Je-li absolutní hodnota $|\gamma_1| < 1$ a soubor je malý (typicky do 25 vzorků), je to indikace dostatečné normality souboru (ev. jeho log-normality, testujeme-li zlogaritmovaná data).

Nejdříve testujeme originální data a když $|\gamma_1| > 1$, tak data zlogaritmuje a testujeme log-normalitu.

Při použití tohoto testu bychom měli mít na zřeteli, že ne každé symetrické rozdělení musí být normální (např. Studentovo rozdělení).

Často bývá v literatuře doporučován test **chí-kvadrát**, avšak US EPA (1992) jej výslovně nedoporučuje pro problémy kontaminace horninového prostředí.

Shapiro-Wilkův test (pro menší soubory do 50 vzorků), resp. jeho modifikace **Shapiro - Francia test** pro soubory větších rozsahů jsou podle US EPA (1992) nejkvalitnější.

² Některá literatura (např. Cl:aire, 2008) nedoporučuje pro výpočet konfidenčních limitů logaritmicou transformaci dat a preferuje použití neparametrických testů (např. Chebychevův teorém) pokud nezlogaritmovaná data nevyhovují normalitě.

3.4. Statistická nezávislost dat

Ověření nezávislosti dat je důležité pro většinu statistických procedur. Soubory závislých dat vykazují nižší variabilitu než odpovídá skutečnosti. Závislost dat může vést k podcenění skutečné variability mezi daty a může závažně ovlivnit reprezentativnost statistických hodnocení. Příklad: pět odečtů téhož teploměru pět minut po sobě nemůže podat dostatečnou informaci o průměrné denní teplotě ovzduší.

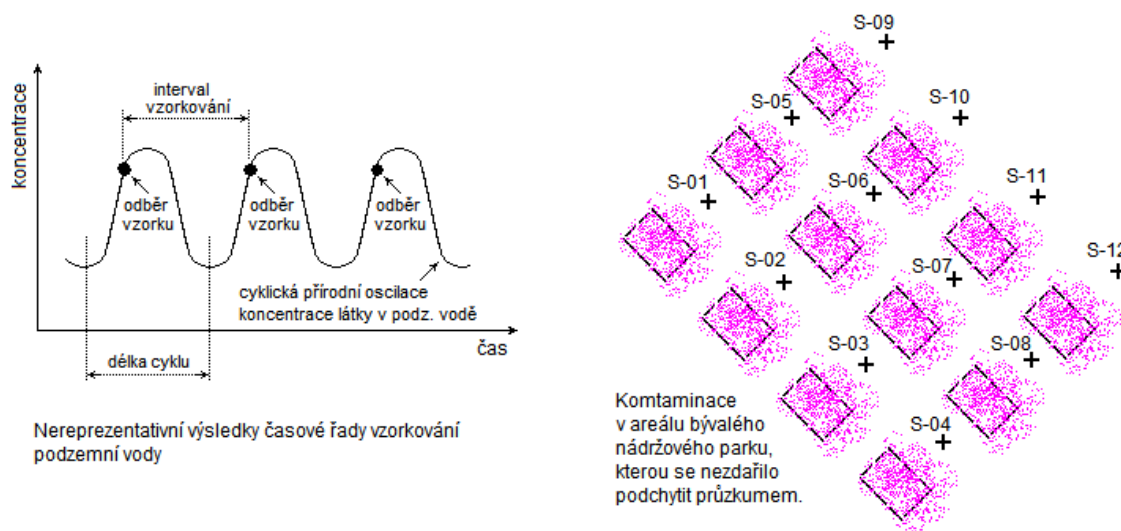
Pro testování statistické nezávislosti dat v souborech uvádí literatura (např. EPA 1992) různé testy - například testy autokorelace resp. von Neumannův poměrový test pro neparametrické soubory.

Při průzkumech kontaminace testy na statistickou nezávislost dat často opomíjíme. Do značné míry spoléháme na to, že nezávislost dat zajišťujeme již preventivně volbou vhodné strategie vzorkování a na to, fyzická nezávislost vzorků je zpravidla i garancí nezávislosti statistické.

Je však dobré mít na zřeteli, že statistická nezávislost dat nemusí být vždy samozřejmostí. Nejobvyklejším a typickým případem tohoto druhu jsou sezónní oscilace koncentrací některých složek chemismu podzemních vod. Nevhodně zvolený vzorkovací interval při neznalosti cykličnosti chemismu může vést k nerepresentativní statistické interpretaci výsledků vzorkování. Ilustruje to připojený obrázek 7 vlevo. Cykličnost sezónních vlivů je specifickým případem závislosti dat.

Cykličnost nemusí být jen časová. Na pravé straně téhož obrázku 7 je znázorněna situace, kdy se průzkumem nepodařilo zjistit rozsáhlou kontaminaci v areálu bývalého nádržového parku, jelikož bylo uplatněno vzorkovací schéma pravidelné sítě bez vědomí dřívějšího rozmístění nádrží.

Obr. 7. Oscilace dat



Jinou možností výskytu cyklické oscilace by mohla být plošná distribuce zbytkové kontaminace zemin na lokalitě po sanaci ventingem za použití pravidelné sítě ventovacích vrtů.

Příkladem jiného druhu fyzické závislosti dat by mohl být odběr vzorku vody z monitorovacího vrtu a pozdější odběr vzorku z dalšího vrtu po směru proudění na téže proudnici, přičemž by časová prodleva mezi odběry odpovídala době, za kterou podzemní voda urazí tuto vzdálenost.

3.5. Úprava datových souborů s velkým počtem nízkých hodnot

Nízké hodnoty (souhrnně ND) mohou být těchto druhů (podle označení v laboratorní praxi):

- U - nedetekovaná hodnota - signál je nerozlišitelný od pozadíového šumu, je pod detekčním limitem analytické metody (MDL),
- J,E - J (pro organiku) nebo E (pro anorganiku) - je detekována určitá nízká koncentrace, jde o pozitivní vzorky, ale koncentrace je tak nízká, že je nejistota v kvantifikaci - hodnota je pod praktickým kvantifikačním limitem analytické metody (PQL).

Výskyt velmi nízkých hodnot komplikuje vyhodnocování statistických souborů. Jejich velký počet závažným způsobem deformuje rozdělení dat. Při výpočtech statistických parametrů u souborů s lognormálním rozdělením nelze za hodnoty pod mezí detekce dosazovat nulu (logaritmus nuly není definován).

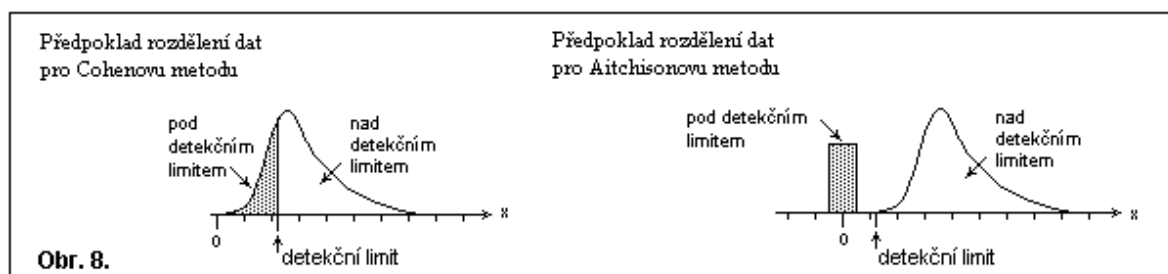
U hodnot pod MDL je skutečná koncentrace někde mezi nulou a MDL. Doporučuje se, dosazovat do souborů hodnoty $1/2$ MDL.

U hodnot pod PQL je skutečná koncentrace někde mezi MDL a PQL. Doporučuje se, do souborů dosazovat buď laboratorním protokolem uváděnou koncentraci nebo PQL.

Základní doporučení (dle US EPA 1992):

- jestliže ND tvoří do 15% souboru, nahraď je buď $1/2$ MDL nebo PQL a pokračuj v běžné parametrické analýze dat,
- jestliže počet ND tvoří nad 15% všech dat souboru, pak při t-testech nebo ANOVA analýzách použij neparametrické testy (Wilcoxon-Rank_Sum test pro porovnávání dvou skupin dat nebo Kruskal-Wallisův test pro více než dvě skupiny dat – viz např. US EPA 1992),
- jestliže počet ND tvoří 15% až 50% všech dat souboru, pak při výpočtech statistických intervalů použij Cohenovu nebo Aitchisonovu úpravu dat (viz např. US EPA 1992),
- jestliže počet ND tvoří nad 50% všech dat souboru, pak pro vyhodnocení používej statistické metody pro neparametrické soubory.

Zda pro úpravu datových souborů použít Cohenovu nebo Aitchisonovu metodu se rozhoduje na základě konstrukce modifikovaných pravděpodobnostních grafů. Cohenův postup předpokládá, že hodnoty pod mezí detekce zapadají do daného rozdělení získaných dat, zatímco Aitchison předpokládá, že odpovídají nulové kontaminaci a jsou mimo rozdělení souboru dat, tvořeného pouze hodnotami nad detekčním limitem. Viz obrázek:



Vyskytují-li se v souboru analýz z kontaminované lokality jak vysoké hodnoty koncentrací, tak vysoký počet nízkých hodnot, indikuje to, že lokalita není homogenním celkem z hlediska předpokladů pro aplikaci statistických metod hodnocení dat. Může jít o situaci, že značná část vzorků s nízkými hodnotami reprezentuje přírodní pozadí a jen část vzorků kontaminovanou

plochu. Je nutno se zamyslet nad plošnou distribucí dat a s přihlédnutím k povaze řešeného problému rozhodnout, zda nerozčlenit lokalitu do několika dílčích celků, hodnocených statisticky samostatně. Inspirací zde může být např. holandská metodika, která doporučuje tento postup pro stanovení průměrných koncentrací v kontaminačním mraku pro účely hodnocení rizik z kontaminace:

- Zeminy: vymezit v ploše lokality areál, kde jsou koncentrace nad součtem $1/2$ limitu C plus pozadí (není-li pozadí stanoveno, bere se nulové). Data z této plochy pak tvoří soubor, pro který se stanoví geometrický průměr a další charakteristiky. Limitem C se zde myslí obecně limit přípustné kontaminace (nesměšovat s Metod. pokynem MŽP).
- Podzemní voda: vymezit v ploše lokality areál, kde jsou koncentrace nad limitem C, potom z dat v tomto areálu určit geometrický průměr a další charakteristiky.

Smyslem tohoto postupu je, aby při analýzách rizika nedocházelo k "ředění" kontaminace a na druhou stranu, abychom riziko nenadhodnocovali.

Detekční limit analytické metody MDL by měl být obecně na úrovni cca $1/10$ kritické koncentrace (limitu).

3.6. Odlehlé hodnoty

Odlehlými hodnotami nazýváme ta data, která evidentně vybočují z distribuce ostatních dat prošetřovaného souboru ve směru k hodnotám vysokým (extrémním).

Je nutno v takovém případě prověřit, zda odlehlá hodnota nemůže být důsledkem hrubé chyby vzorkování nebo analýzy. Při vylučování odlehlých hodnot ze souboru je nutno postupovat s velkou opatrností, zvláště u souborů s lognormálním rozdělením. Vyloučení odlehlé hodnoty ze souboru musí být vždy jmenovitě a konkrétně odůvodněno. Je třeba rovněž zvážit, zda výskyt odlehlých hodnot neindikuje, že data nepocházejí z jedné populace, že lokalita není statisticky homogenním celkem. Nejčastěji extrémní hodnoty pocházejí z ohnisek kontaminace o jejichž existenci jsme nevěděli a proto je nezohledňoval vzorkovací plán.

Formální test na odlehlé hodnoty provádíme jen při podezření, že velmi vysoké koncentrace některých vzorků se sice vymykají zbytku dat, avšak nejde o chybu. Test předpokládá, že zbytek dat má normální rozdělení. Jelikož při lognormálním rozdělení bývají velmi vysoké hodnoty menší části vzorků pravidlem, provádíme raději test přímo s logaritmy dat. Tím se vyhneme tomu, abychom neklasifikovali data odpovídající regulárnímu lognormálnímu rozdělení dat chybně jako odlehlé hodnoty.

Nejběžnější testy na odlehlé hodnoty:

- Grubbsův parametrický test extrémních odchylek (Reisenauer 1973, též EPA 1992),
- Dixonův neparametrický test extrémních odchylek (Reisenauer 1973).

Neformálně, avšak názorně a jednoduše odlehlé hodnoty vyniknou při vykreslení tzv. pravděpodobnostního grafu pro normální nebo zlogaritmovaná data prošetřovaného souboru.

3.7. Testování hypotéz

Vyslovování závěrů ze statistického hodnocení dat je vždy rozhodováním mezi dvěma scénáři, mezi dvěma hypotézami.

Shoda dat s hypotézou ještě neznamená, že hypotéza je pravdivá. Naopak, data odporující hypotéze ukazují na to, že hypotéza pravdivá není. Hypotézu nelze na základě dat dokázat, lze

ji však na základě dat vyvrátit. Toto je okolnost zásadní důležitosti. Podle ní v praxi volíme nulovou hypotézu v závislosti na tom, co by bylo menším průšvihem v případě, kdybychom z dat vyvodili chybný závěr. Nulová hypotéza H_0 je formulována tak, aby ji mohla data vyvrátit v případě, že není pravdivá. H_0 bývá tedy opak toho, co chci dokázat.

Rozdíly ve výpočtových postupech při rozdílně formulovaných nulových hypotézách se mohou jevit subtilní, dopady na závěry jsou však zásadní.

Existují dvě možnosti jaká je skutečnost: H_0 platí nebo neplatí. Naše rozhodnutí může být též dvojitý: H_0 zamítám nebo nezamítám. Rozhodovací proces lze znázornit touto tabulkou:

Tab. 3. Testování hypotéz		Typ rozhodnutí (přijetí hypotézy):	
		akceptuji H_0	zamítám H_0 , přijímám H_A
skutečnost	H_0 je pravdivá	správný závěr (jeho pravděpodobnost = $1-\alpha$)	chybný závěr - (chyba I. druhu - její pravděpodobnost = α)
	H_0 je nepravdivá	chybný závěr (chyba II. druhu - její pravděpodobnost = β)	správný závěr (jeho pravděpodobnost = $1-\beta$)

Hodnoty pravděpodobnosti α , β jsou kritické úrovně jistoty, se kterou se vyslovujeme k platnosti či neplatnosti hypotéz. Při statistickém hodnocení nám vždy zůstane určitá nejistota zda jsme učinili správný závěr, jelikož na vlastnosti celého základního souboru (celé populace) usuzujeme z výběru omezeného rozsahu. Ten reprezentuje základní soubor jen s určitou mírou spolehlivosti. Chyby I. i II. druhu jsou vlastní rozhodování na základě statistických dat a nelze je nijak úplně eliminovat. Čím menší pravděpodobnost chyby I. druhu jsme ochotni připustit, tím větší vytváříme předpoklad pro chybu II. druhu. Čím menší α , tím větší β . Za lepší ochranu před chybou I. druhu se platí větší pravděpodobností chyby II. druhu. Cílem statistických metod je minimalizace falešných závěrů I. druhu.

Hladina významnosti α pro zamítnutí nulové hypotézy se při testování v kontaminační hydrogeologii obvykle volí na úrovni 0,05. Lze ji pro daný datový soubor také vypočítat a podle výsledku pak konkrétněji posuzovat přesvědčivost dat z hlediska zamítnutí či přijetí nulové hypotézy (viz kap. 9 – tzv. p-hodnoty).

V předpisech některých amerických států se připouští při hodnocení rizika pro ne-rezidenční scénáře hladina významnosti $\alpha = 0,20$ pro zamítnutí nulové hypotézy.

Obecně je postup při testování hypotéz tento:

- Formuluji nulovou hypotézu H_0 ,
- definuji požadovanou hladinu významnosti,
- poté prověřuji - testuji, zda data jsou či nejsou v rozporu s H_0 .

Pokud jsou data v rozporu s H_0 , přijímám alternativní hypotézu H_A , která je negací nulové hypotézy.

Nulovou hypotézu zamítám, pokud je podle testu uspořádání dat takové, že je za předpokladu platnosti nulové hypotézy velmi nepravděpodobné.

Metodická příručka
Hodnocení průzkumu a sanací

Formulace nulové hypotézy v závislosti na charakteru řešeného problému:

V principu si klademe v kontaminační geologii/hydrogeologii vždy jednu ze dvou základních otázek, kterým odpovídají rozdílně formulované nulové hypotézy:

1)	Řešený problém – co nás zajímá:	Je lokalita čistá? Jsou k dispozici dostatečně přesvědčivé důkazy o tom, že koncentrace škodliviny na lokalitě je nižší než sanační limit či jiný limit přípustné kontaminace stanovený na základě ochrany zdraví či ekosystémů (tj. limit odvozený zpětným propočtem z úrovně akceptovatelného rizika)?
	Nulová hypotéza H_0	Lokalita je kontaminovaná, pokud nebude prokázán opak. Takto formulovanou nulovou hypotézu pro porovnávání koncentrací na lokalitě s limity přípustné kontaminace volíme s ohledem na to, že chceme mít vyšší jistotu pro závěr, že lokalita je čistá. Prohlášení čisté lokality za „špinavou“ považujeme za chybu méně závažnou – nedojde v tomto případě k podcenění rizika ohrožení zdraví jejich uživatelů.
	Alternativní H_A	Lokalita kontaminovaná není.
	Použití v situacích:	Hodnocení výsledků sanace zemin nebo podzemních vod na základě vzorkování po ukončení sanace Postsanační monitoring podzemních vod na lokalitě či výstupním profilu proudění podzemních vod z lokality Kontrolní monitoring podzemních vod na lokalitě či na výstupním profilu od ní – kontrola dodržování limitu přípustné kontaminace (např. kontrolní monitoring skládky) Posuzování možností využívání lokality s určitou úrovní kontaminace zemin, půdního vzduchu či podzemních vod pro určitý způsob funkčního využívání, vymezený limity přípustné kontaminace Postsanační monitoring půdního vzduchu či ovzduší v objektech v areálu sanované lokality
2)	Řešený problém - co nás zajímá:	Je lokalita kontaminovaná? Jsou k dispozici dostatečně přesvědčivé důkazy o tom, že koncentrace škodliviny na lokalitě je vyšší než požadovaná koncentrace resp. úřední limit přípustné kontaminace odvozený od pozadí?
	Nulová hypotéza H_0	Lokalita kontaminovaná není, pokud nebude prokázán opak. Takto formulovanou nulovou hypotézu volíme s ohledem na to, že nechceme zbytečně vyvolávat planý poplach, chceme mít vyšší jistotu pro závěr, že na lokalitě se opravdu vyskytuje kontaminace.
	Alternativní H_A	Lokalita je kontaminovaná.
	Použití v situacích:	Porovnávání dat o kontaminaci zemin či půdního vzduchu kontaminované lokality s hodnotami pozadí Detekční monitoring podzemních vod na sledované lokalitě nebo na výstupním profilu proudění podzemních vod - detekce vzniku kontaminace nad úrovní pozadí (např. v areálu skládky)

3.8. Testy významnosti

Jsou to kvalitativní testy k prověřování hypotéz o souborech dat. Potřebujeme se např. přesvědčit:

- zda zkoumaný výběr pochází ze základního souboru s určitým rozdělením,
- zda dva výběry pocházejí z jednoho a téhož základního souboru,
- zda lze pohlížet na odlehlou hodnotu jako na důsledek hrubé chyby měření,
- zda je možno považovat studovaný soubor za náhodně uspořádaný,
- jak významně se liší výsledky analýz od kontrolních duplicitních analýz,
- zda se soubor dat z lokality významně liší od souboru dat pozadových, atd.

Takových otázek může být celá řada. Odpovídáme na ně pomocí testů významnosti.

Obecný postup při testech významnosti:

- formulujeme nulovou hypotézu,
- volíme hladinu významnosti (zpravidla 0,01 nebo 0,05, nejvýše 0,1),
- vypočteme hodnotu testovacího kritéria a porovnáme s tabelovanou kritickou hodnotou,
- formulujeme závěr - zamítáme či akceptujeme nulovou hypotézu,
- popřípadě vypočteme tzv. p-hodnoty,
- interpretujeme výsledek.

Ke všem testům potřebujeme mít k dispozici odpovídající statistické tabulky. Každému testovacímu kritériu přísluší určité rozdělení. Kritické hodnoty pro jednotlivá rozdělení, t.j. hodnoty, které příslušná náhodná veličina překročí s danou pravděpodobností, bývají tabelovány. Na základě porovnání vypočítané hodnoty testovacího kritéria s příslušnou kritickou hodnotou odpovídající náhodné veličiny zamítáme či přijímáme nulovou hypotézu.

Pokud srovnáváme navzájem dva statistické soubory dat (dva výběry), pak hovoříme o párovém testu. Pokud test použijeme pro porovnání jednoho datového souboru s fixní číselnou veličinou (například se sanačním či jiným limitem), hovoříme o jedno-výběrovém testu.

3.9. P-hodnoty

Srovnání vypočtené hodnoty testovacího kritéria s tabulkovou kritickou hodnotou je jen informací typu ano/ne, informací zda pravděpodobnost platnosti nulové hypotézy je větší nebo menší než hladina významnosti, na které testování provádíme.

Místo takového prostého srovnávání vypočtených hodnot testovacího kritéria s kritickými hodnotami se někdy ještě navíc používá postup, při kterém se vypočítá přímo pravděpodobnost p , se kterou studovaná data zapadají do rozdělení odpovídajícího nulové hypotéze. Tyto tzv. **p-hodnoty** nám podávají podrobnější informaci o pravděpodobnosti platnosti nulové hypotézy. Použití p-hodnot umožňuje také relativní porovnávání přesvědčivosti hodnot z různých souborů.

3.10. T-testy

Jsou to poměrně jednoduché a v praxi kontaminační hydrogeologie nepoužívanější testy pro srovnávání významnosti rozdílů průměrů dvou souborů dat.

Předpoklady použití: soubory nezávislých dat, normální rozdělení, zhruba stejné rozptyly.

T-testy jsou vcelku tolerantní k mírným odchylkám od normální distribuce dat, jsou ale citlivé na výskyt odlehklých hodnot.

Pro neparametrické soubory je variantou t-testu tzv. **Wilcoxon-Rank-Sum test** (viz např. US EPA, 1992) nebo **Čebyčevův (Chebychevův) teorém** (Claire, 2008³). Pro zvýšení spolehlivosti závěrů při použití Wilcoxon-Rank-Sum testu doporučuje US EPA (2000) navíc použít ještě tzv. **Kvantilový test**, zvláště pro soubory s velkým počtem hodnot pod mezí detekce. Minimální doporučený počet vzorků pro aplikaci Wilcoxon-Rank-Sum testu je nejméně 4 vzorky pro každou skupinu dat.

3.11. Multi-srovnávací strategie

Porovnávání více než dvou souborů dat. Předpoklady aplikace: soubory nezávislých dat, normální rozdělení, zhruba stejné rozptyly.

Na lokalitách máme v praxi často větší počet monitorovacích vrtů a v každém vrtu sledujeme více parametrů, které musíme porovnávat s limitem či pozadím. Kdybychom testovali každý vrt a každý parametr v něm individuálně porovnávacím párovým t-testem, budou se nám kumulovat nejistoty a podstatně se zvyšuje riziko konečného chybného výroku (lokalita je kontaminovaná, i když ve skutečnosti není)⁴. Navíc, pro větší počet souborů by bylo takové individuální porovnávání komplikované a pracné.

Místo série párových t-testů volíme tedy vhodněji multiporovnávací strategie typu **Omnibus**, kam patří i tzv. analýza variancí **ANOVA**. Data z několika vrtů jsou sdružena dohromady a testována jako celek. Tento postup nám nespecifikuje, který ze souborů vykazuje statisticky významnou odchylku, avšak pouze to, že přinejmenším v jedné skupině dat se taková odchylka vyskytuje. To musíme prověřit následným ad-hoc testováním jednotlivých skupin dat (párovými resp. jednovýběrovými testy).

Typicky se ANOVA používá při detekčním monitoringu podzemních vod, kdy data z pozadových vrtů tvoří jednu skupinu a data ze sledované lokality (z výstupu vod ze sledované lokality) skupinu druhou. Metoda je citlivá na prostorovou variabilitu dat z lokality a má malou spolehlivost při detekci úzkých kontaminačních mraků. Náklady vzorkování narůstají, protože jsou zapotřebí nejméně 4 vzorky z každého vrtu (Weber, 1995).

Variantou ANOVY pro neparametrické soubory je **Kruskal-Wallisův test**. Předpokládá se zhruba stejné rozdělení souborů, to ale nemusí být nutně symetrické.

Podrobněji k těmto typům testů viz například US EPA, 1992.

Alternativní strategií je tzv. **retesting**, zvláště pro malé lokality, kde by bylo riskantní vyslovovat definitivní závěry z malého počtu dat. Vhodnějším postupem zde je, odvodit z existujících dat nejprve predikční a toleranční limity. Pro vyslovení definitivních závěrů je

³ Tato literatura doporučuje Chebychevův test jako vhodnější (robustnější) alternativu i pro ty soubory s nesymetrickým rozdělením, kde by šlo použít t-test se zlogaritmovanými původními daty.

⁴ Tak pro 20 nezávislých porovnávání, z nichž každé bude mít samo o sobě pravděpodobnost chybného závěru 1%, bude celková nejistota z každé vzorkovací kampaně $[1-(1-0,01)^{20}] = 18\%$.

rozhodující teprve jejich konfrontace s výsledky dalšího vzorkování. Formalizované postupy viz US EPA 1992.

3.12. Statistické intervaly

Princip použití statistických intervalů pro hodnocení kontaminace: horní nebo dolní mez statistického intervalu (v závislosti na povaze řešeného problému) porovnáváme s limity či s pozadím, popřípadě s nověji změřenými hodnotami. Rozeznáváme limity konfidenční, toleranční a predikční. Nejběžněji se používají intervaly kolem průměrů. Mohli bychom je však určovat i pro percentily či rozptyly.

Intervaly můžeme vypočítat z parametrů výběru, který máme k dispozici. Obecný vzorec pro výpočet mezí statistického intervalu datového souboru kolem průměru:

$$\text{mez intervalu} = \bar{x} \pm K_{(1-\alpha, n)} \cdot S$$

kde „ $\bar{x}$ “ je průměr prošetřovaného souboru, znaménko „+“ či „-“ platí alternativně pro horní resp. dolní mez intervalu, „K“ je faktor přebíraný ze statistických tabulek pro zvolenou úroveň nejistoty „ α “ a pro daný počet vzorků „n“, „S“ je směrodatná odchylka prošetřovaného souboru.

Šíře statistického intervalu závisí na velikosti směrodatné odchylky výběrového souboru, na rozsahu výběru (tj. počtu dat, která jsou k dispozici) a na zvolené úrovni spolehlivosti.

Typicky platí, že pokud budou vypočteny intervaly ze stejných dat a se stejnou úrovní jistoty, pak nejširší budou intervaly toleranční, potom predikční, nejužší budou konfidenční.

Pokud mají data lognormální rozdělení, pak se intervaly konstruují pro zlogaritmovaná data a meze intervalů se porovnávají s logaritmy limitů resp. s logaritmy naměřených dat či s logaritmy požadovaných hodnot.

3.12.1. Konfidenční intervaly

Konfidenční interval (též interval spolehlivosti) vymezuje rozpětí hodnot, ve kterém se s určitou mírou jistoty $(1-\alpha)$ bude nacházet určitý parametr základního souboru. Nejistota, tj. pravděpodobnost, že se parametr základního souboru nenachází uvnitř konfidenčního intervalu $= \alpha$. Dolní mez konfidenčního intervalu se běžně označuje LCL, horní mez UCL (z angličtiny).

Konfidenční interval pro průměr

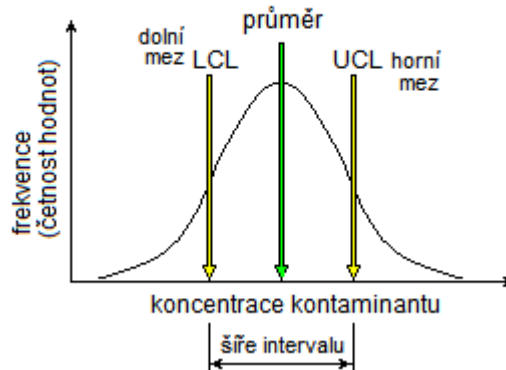
Nejčastěji stanovujeme konfidenční intervaly pro průměry výběrového souboru. Pro praxi mohou mít význam například ještě i konfidenční intervaly pro určité kvantily.

Pojem konfidenčního intervalu pro průměr vysvětluje vedlejší obrázek č.9.

Průměr se může ve skutečnosti nacházet kdekoli uvnitř konfidenčního intervalu.

Se vzrůstem počtu vzorků se šířka intervalu pro zvolenou úroveň jistoty zužuje.

Obr. 9. KONFIDENČNÍ INTERVAL PRO PRŮMĚR



Čím větší je výběrový soubor, tím se odhad parametru základního souboru upřesňuje a šíře intervalu spolehlivosti se zužuje. Minimální počet dat pro výpočet konfidenčních intervalů jsou 4, ale pokud není k dispozici alespoň 8 - 10 dat, jsou extrémně široké. Je však dobré si uvědomit, že spolehlivost neroste s počtem vzorků lineárně. Od určité velikosti souboru je již příspěvek dalších vzorků k poznání zkoumaného objektu nevýznamný.

Akceptovatelná nejistota při hodnocení dat z průzkumu kontaminace horninového prostředí je zpravidla 5%, nejvýše 10% (= spolehlivost 95 resp. 90%). Spokojíme-li se s nižší spolehlivostí, šíře konfidenčního intervalu se zúží.

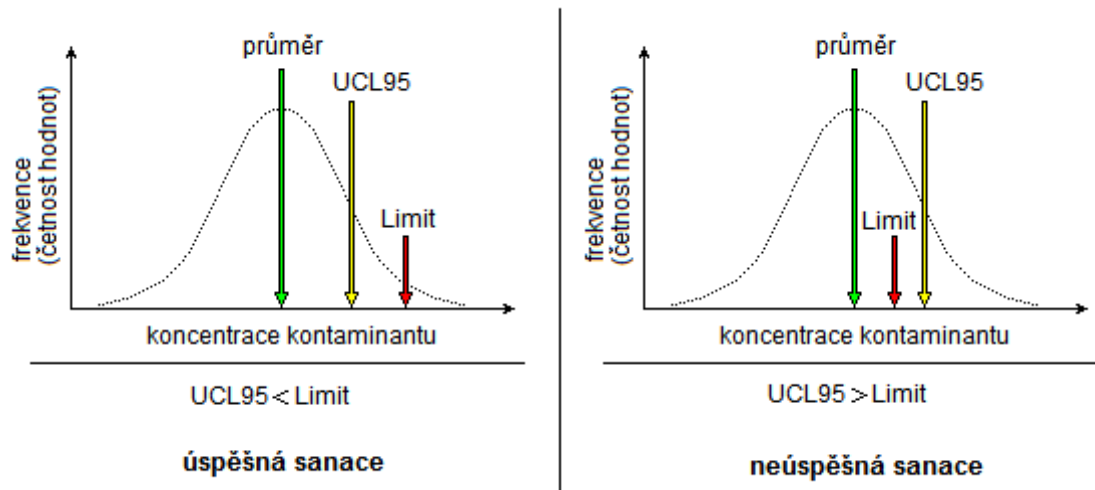
Příklad použití konfidenčního intervalu pro průměr při hodnocení výsledků sanace:

Pokud má být vzorkováním potvrzeno, že zbytková kontaminace zemin či podzemních vod sanované lokality je se zvolenou úrovní jistoty nad či pod stanoveným sanačním limitem, porovnáváme v zájmu ochrany lidského zdraví s tímto limitem nikoliv průměr ze vzorkování, nýbrž horní mez konfidenčního intervalu pro průměr (platí pro situaci, kdy byl sanační limit odvozen ve vztahu k ochraně zdraví, zpětným propočtem z úrovně akceptovatelného rizika).

Interpretace situace na levé straně obrázku 10: Je-li UCL95 těsně pod hranici sanačního limitu, pak je 5% pravděpodobnost, že je chybný závěr znící: lokalita je kontaminovaná⁵. S 95% jistotou lze konstatovat, že sanace byla úspěšná.

Interpretace situace na pravé straně obrázku 10: Je-li UCL95 těsně na hranici sanačního limitu, pak je 5% pravděpodobnost, že je chybný závěr znící: lokalita je čistá. S 95% jistotou lze konstatovat, že sanace byla neúspěšná.

Obr. 10. Konfidenční interval pro průměr při posuzování dosažení cílů sanace



Alternativou k použití UCL95 pro porovnávání zbytkových koncentrací se sanačním limitem je použití jedno-výběrového t-testu pro parametrické soubory nebo např. Čebyševova jednostranného testu pro neparametrické soubory (viz US EPA 1992 či Claire, 2008).

Použití konfidenčního intervalu pro průměr při analýze rizika:

Při výpočtu rizika z kontaminace horninového prostředí dosazujeme za expoziční koncentraci nikoliv průměr ze vzorkování, nýbrž horní mez konfidenčního intervalu UCL pro tento průměr. Je to opět v zájmu ochrany lidského zdraví.

⁵ Zdánlivě krkolomný interpretační výrok souvisí s formulací tzv. nulové hypotézy, kterou prověříme statistickým testem.

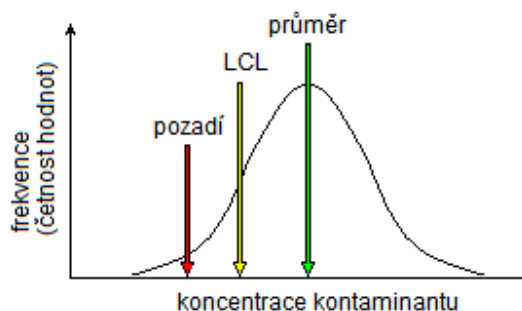
Pokud hodnotíme úroveň kontaminace či úspěšnost sanace podzemních vod, pak musíme mít na zřeteli tyto další okonosti: pokud jsou limity přípustné kontaminace stanoveny pro využívání vody jako pitné, pak nemůžeme v zájmovém území použít pro hodnocení přípustnosti zbytkové kontaminace ani průměr, ani UCL průměru ze všech vrtů v zájmové oblasti. Bezpečnost využívání vody jako pitné musí být zajištěna pro každé místo zvodně bez ohledu na konkrétní lokalizaci jímacího objektu. Každý vzorkovací bod musíme porovnávat s limitem individuálně.

Porovnávání s pozadím: abychom předcházeli ukvapeným závěrům o existenci kontaminace, tak při porovnávání s pozadím použijeme spodní mez konfidenčního intervalu LCL95 pro průměr koncentrací zemin či podzemních vod na lokalitě či na výstupu podzemních vod z areálu sledované lokality⁶.

Je-li již LCL95 nad pozadím, pak je to s 95% jistotou významná indikace toho, že v areálu lokality dochází ke kontaminaci nad úroveň pozadí.

Situaci znázorňuje vedlejší obrázek č. 11 →

Obr. 11. Použití LCL pro průměr



Je-li již LCL průměru nad pozadím, pak je průměr s vysokou mírou jistoty skutečně nad pozadím.

Pokud nemáme pozadí určeno jedinou hodnotou ale statistickým souborem, pak jej neporovnáváme s LCL, ale použijeme párový t-test pro porovnání se souborem dat z lokality.

Při monitoringu podzemních vod nenacházejí konfidenční intervaly - kromě zde jmenovaných případů - příliš využití. Nepodávají totiž dostatečnou informaci o nadprůměrných koncentracích. Ty nás totiž často zajímají především a to jak při posuzování negativních důsledků kontaminace, tak při sledování trendů. Významnější využitelnost mají při monitoringu podzemních vod intervaly toleranční a predikční.

Výpočet konfidenčního intervalu pro aritmetický průměr:

Minimální počet dat pro výpočet konfidenčního intervalu je 9 (DEQ, 2002), data musí být statisticky nezávislá a pocházet z jedné populace (potvrdit testem rozdělení souboru).

Průměr základního souboru μ bude s pravděpodobností $p = (1-\alpha) = CL$ ležet v konfidenčním intervalu pro parametr μ .

Meze konfidenčního intervalu pro aritmetický průměr:	$\left(\bar{x} - u_p \cdot \frac{s}{\sqrt{n-1}} \right) \leq \mu \leq \left(\bar{x} + u_p \cdot \frac{s}{\sqrt{n-1}} \right)$	
	spodní mez LCL	horní mez UCL

kde u_p je t.zv. normovaná náhodná veličina z tabulek pro danou úroveň spolehlivosti:

spolehlivost p [%]	68,2	90,0	95,0	99,0
u_p	1,00	1,645	1,96	2,576

Pro výběry malého rozsahu ($n \leq 30$) nahrazujeme u_p hodnotou $t_{\alpha, n}$ Studentova rozdělení, kterou nalezneme v běžných statistických tabulkách (např. tab. VII v Reisenauer, 1973) pro

⁶ Souvislost s formulací tzv. nulové hypotézy.

danou úroveň spolehlivosti a pro počet stupňů volnosti $v = (n-1)$, kde n je počet prvků výběru.

Konfidenční interval pro geometrický průměr:

Existuje více metod (viz například US EPA, 1992), nelze použít prosté dosazení zlogaritmovaných dat do výše uvedeného vzorce pro výpočet UCL aritmetického průměru.

Vypočti aritmetický průměr zlogaritmovaných dat ϕ a rozptyl zlogaritmovaných dat $SD^2 = \eta$. Potom:	
$LCL_{\alpha} = \exp\left(\phi + 0,5\eta + \frac{\sqrt{\eta} \cdot H_{\alpha}}{\sqrt{n-1}}\right)$	$UCL_{(1-\alpha)} = \exp\left(\phi + 0,5\eta + \frac{\sqrt{\eta} \cdot H_{(1-\alpha)}}{\sqrt{n-1}}\right)$
Kde H_{α} pro CL 95% ($H_{0,05}$) a pro CL90% ($H_{0,10}$) viz tabulky EPA 1992, str. 59, 60	$H_{(1-\alpha)}$ viz tabulky US EPA, 1989.

Pro výpočet konfidenčních limitů pro neparametrické soubory lze nalézt v odborné literatuře celou řadu postupů. Některá literatura (Cl:aire, 2008) zpochybňuje jejich obecnou využitelnost a dokonce nedoporučuje případně používat ani logaritmickou transformaci pro lognormální soubory. Jako nejrobustnější metodu pro testování neparametrických souborů doporučuje aplikaci Chebychevova teorému (Cl:aire, 2008).

3.12.2. Toleranční intervaly při monitoringu podzemních vod

Toleranční interval zahrnuje zvolenou část populace (např. 95% možných výsledků měření - hovoříme zde o tak zvaném pokrytí) s určitou úrovní jistoty, odpovídající tzv. tolerančnímu koeficientu.

Například toleranční interval s pokrytím 95% a s tolerančním koeficientem 95% zahrne v průměru 95 procent všech možných výsledků měření (analýz) s jistotou 95%. Při menším počtu odebraných vzorků proto bývá vybočení některé analýzy nad horní mez tolerančního intervalu (UTL) ojedinělým jevem a indikuje vznik kontaminace.

Při větším počtu vzorků je ovšem vybočení nad UTL možné, aniž to nutně znamená kontaminaci (při úrovni jistoty 95% to může být statisticky 1 z 20 vzorků). Volíme pak obvykle **převzorkování** vrtu nebo zvýšíme frekvenci vzorkování, než vyslovíme definitivní závěr.

Použití při monitoringu podzemních vod (EPA 1992):

- Detekční monitoring – porovnávání vrtů z lokality s pozadím. Když koncentrace v některém vzorku z lokality nebo po směru proudu od ní překročí horní mez tolerančního limitu UTL vypočteného pro datový soubor z pozadového vrtu (z vrtu proti směru proudění podzemních vod ke sledované lokalitě), je to významná indikace toho, že na lokalitě dochází ke kontaminaci.
- Detekční monitoring – porovnávání výsledků vzorkování v jednom a tomtéž vrtě bez kontaminace. Z dosavadních dat se vypočte UTL. Jeho případné překročení při dalším vzorkování je významnou indikací toho, že na lokalitě dochází ke kontaminaci.
- Kontrolní monitoring - Když vypočtený UTL pro data z vrtu v areálu lokality nebo pod ní překročí stanovený legislativní limit, pak je to významná indikace toho, že na lokalitě dochází ke kontaminaci nad přípustnou mez. Toto použití UTL je zvláště vhodné pro případy, kdy máme k dispozici více než 5 hodnot.

Toleranční limity lze konstruovat pro parametrické i neparametrické soubory. Parametrické toleranční intervaly lze počítat i pro lognormální distribuci. Konstrukce neparametrických intervalů je založena na třídění dat. Jejich použití je zvláště vhodné pro soubory s velkým počtem hodnot pod mezí detekce. Již z principu vyplývá, že neparametrické toleranční intervaly lze konstruovat pouze pro data z vrtů bez kontaminace.

Podrobnou metodiku i příklady pro výpočet tolerančního intervalu uvádí např. literatura US EPA, 1992.

3.12.3. Predikční intervaly při monitoringu podzemních vod

Predikční interval zahrnuje s určenou pravděpodobností výsledky určitého počtu budoucího vzorkování (může to být i jedno budoucí vzorkování). Budeme-li po určitou dobu vzorkovat pozadový vrt a získaná data použijeme k sestrojení predikčního intervalu, potom výsledky budoucího vzorkování podzemních vod na odtoku z lokality by měly být se zvolenou mírou jistoty v tomto intervalu (nezmění-li se ovšem charakter distribuce). Predikční interval zahrnuje $100(1-\alpha)$ procent populace.

Použití (dle EPA 1992):

- Detekční monitoring: srovnávání dat z vrtu na lokalitě či na odtoku podzemních vod ze sledované lokality s horním predikčním limitem UPL, který byl vypočten z vrtů monitorujících pozadí (proti směru proudění podzemních vod). Bude-li výsledek novější analýzy nad horní mezí predikčního intervalu UPL, je to významná indikace vzniku kontaminace podzemních vod.
- Detekční monitoring – Porovnávání výsledků vzorkování podzemní vody z jednoho a téhož vrtu, který je až dosud bez kontaminace. Z dosavadních výsledků se vypočte UPL. Jeho případné překročení při dalším vzorkování by bylo významnou indikací toho, že na lokalitě dochází ke kontaminaci.

Podrobnou metodiku i příklady pro výpočet predikčních intervalů uvádí např. literatura US EPA, 1992. Lze konstruovat predikční intervaly i pro neparametrické soubory.

Účinnost použití predikčních intervalů klesá, pokud obsahují více než 50% dat pod mezí detekce. Jejich použití je tudíž zvláště vhodné pro případy, kdy se monitorují konstituenty, které jsou i nekontaminovaných podzemních vodách běžně přítomny (např. sírany aj.).

3.13. Časové řady

Časová řada je soubor naměřených dat uspořádaný podle časové osy. V kontaminační hydrogeologii se grafy časových řad využívají k hodnocení a prognózování časového vývoje koncentrací znečištění podzemní vody v jednotlivých monitorovacích vrtech.

Grafy často umožňují již z prostého náhledu usuzovat na dosavadní trendy a budoucí vývoj koncentrací. Zpravidla však interpretaci časových řad komplikuje to, že jejich průběh bývá výsledkem interference více faktorů. Jejich identifikace a analýza nemusí být jednoduchá.

Obecně mívá časová řada tyto složky:

- složka periodická - sezónního (S) a/nebo delšího - cyklického (C) charakteru,
- složka trendová (T),
- složka nepravidelná – náhodná (ϵ).

Nejčastěji řešenou úlohou v kontaminační hydrogeologii bývá analýza trendové složky, což vyžaduje dekompozici časové řady, tj. její „očistění“ od složek ostatních, její „vyhlazení“.

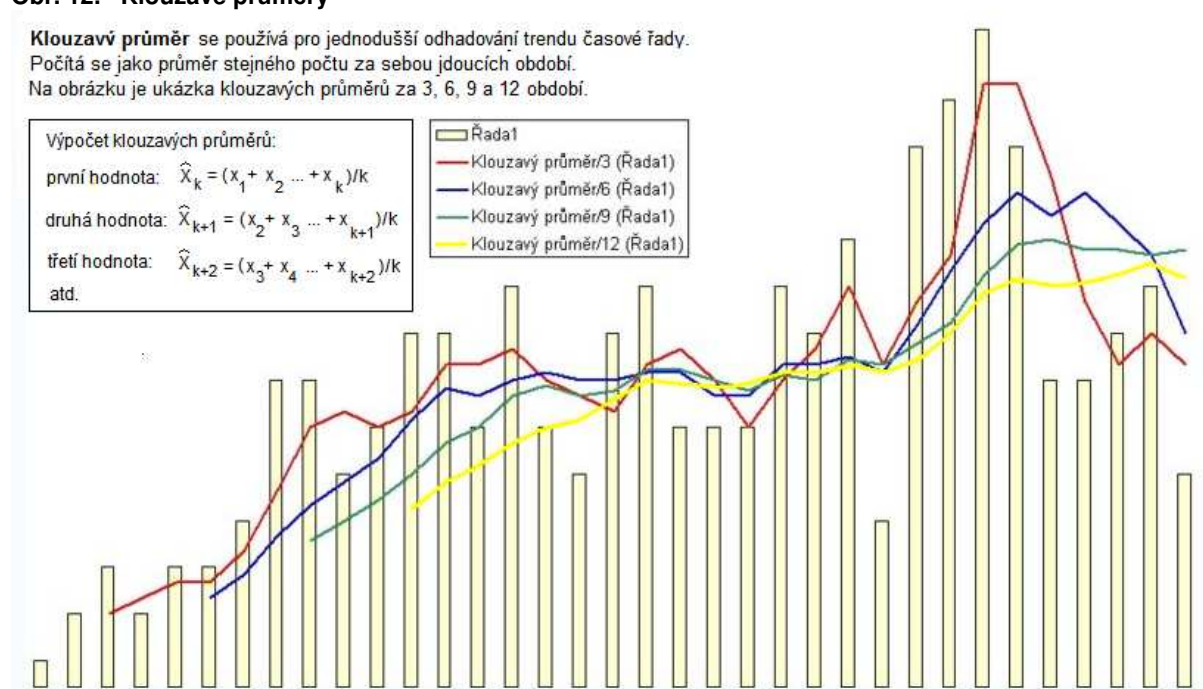
K dispozici je pro tyto účely řada statistických metod, z nichž ty základní patří do těchto dvou skupin:

- adaptivní postupy,
- regresní analýza.

Existují ještě další metody analýzy časových řad, které jsou však již výrazně náročnější na matematický aparát (zejména Box-Jenkinsova metodologie, spektrální analýza aj.).

Názornější vizuální identifikaci trendů umožní v řadě případů již jednoduchý výpočet a zobrazení tzv. **klouzavých průměrů**. Tato adaptivní metoda „vyhlazuje“ původní časovou řadu, přičemž toto vyhlazení je tím větší, čím je větší délka klouzavých průměrů (tj. kolik předchozích hodnot zahrnuje výpočet klouzavého průměru pro daný bod na časové ose). Princip a použití metody znázorňuje připojený obrázek č. 12. Obrázek je rovněž ilustrativní pro pochopení principu „vyhlazování“ časových řad vůbec.

Obr. 12. Klouzavé průměry



Další poměrně jednoduchou a v praxi často využívanou adaptivní metodou pro identifikaci vzestupného trendu je kombinovaná **Shewhart - CUSUM metoda** (EPA, 1992a), u které se sekvence dosud získaných dat použije k výpočtu limitů, jejichž překročení je považováno za významné. Shewhart - CUSUM procedura je použitelná, pokud nejméně 50% hodnot časové řady je nad detekčním limitem.

Podobně nám mohou pro indikaci vybočení z dosavadní řady vzorkování posloužit meze tolerančního či predikčního intervalu.

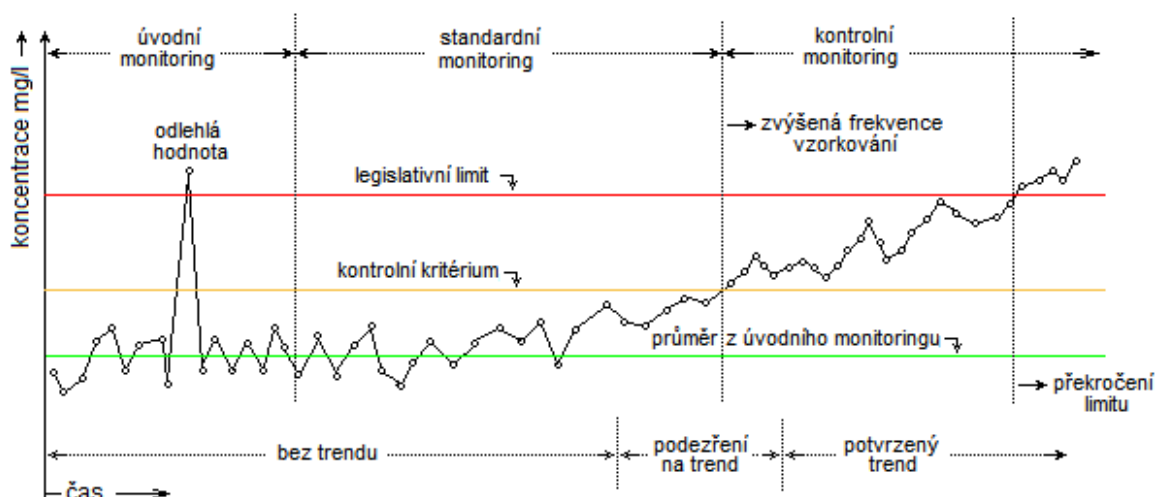
Poměrně ještě nenáročné na matematický aparát jsou též metody **exponenciálního vyrovnávání** časových řad, nabízející se rovněž k použití pro jejich vyhlazování a k identifikaci trendů.

Regresní analýza se používá ke stanovení trendů a modelování budoucího vývoje koncentrací na základě aproximace trendu matematickými funkcemi, vkládanými do originální časové řady se snahou o co nejtěsnější korelaci.

Pro neparametrické soubory je variantou regresní analýzy **Mann-Kendallův test** (pro monotónní trendy, tj. trendy jen sestupné nebo jen vzestupné). Viz Gilbert, 1987, nebo US EPA, 1996. Existuje varianta tohoto testu s možností eliminace sezónních vlivů.

V každém případě potřebujeme mít k odhadu trendů a prognózování budoucího vývoje časových řad k dispozici data z tak zvané úvodní etapy monitoringu, která nám slouží jako východisko a vstupní soubor pro aplikaci statistických metod. Je žádoucí, aby etapa úvodního monitoringu podchycovala přinejmenším jeden celý rok, často doporučovaný minimální počet vzorků je 8. Jako názorná ukázka významu úvodní etapy slouží následující obrázek č. 13.

Obr. 13. Monitoring kvality podzemní vody – graf časové řady



Pro další monitoring po úvodní etapě je vhodné stanovení tzv. kontrolního koncentračního kritéria, jehož případné překročení budeme interpretovat jako první signál možné změny dosavadního trendu vývoje koncentrací v čase. Zpravidla to bude interní podnět k tomu, abychom věnovali zvýšenou pozornost dalším výsledkům, abychom provedli kontrolní vzorkování a popřípadě zvýšili dosavadní frekvenci odběru vzorků. Vhodné nastavení kontrolního kritéria je věcí odborné úvahy. Mohou to být hodnoty stanovené například statistickými metodami (horní meze tolerančních či predikčních intervalů aj.), může to být také například překročení průměru z úvodní etapy o hodnotu směrodatné odchylky či jejího násobku nebo též překročení dosavadního maxima a podobně. Též se nabízí například poloviční hodnota oficiálního limitu, jehož překročení již znamená opravdu problém, který je nutno řešit s úřadem.

Analýza časových řad se všeobecně doporučuje jen pro vrty bez kontaminace, nikoliv pro hodnocení vývoje kontaminačních mraků.

3.14. Závislost dvou náhodných kvantitativních veličin - regrese a korelace

Když se na jednom prvku současně měří více veličin které mohou být na sobě závislé, získáme vícerozměrný soubor.

U dvourozměrných souborů jsou měřeny hodnoty dvou proměnných, takže se získávají dvojice výsledků měření: (x_1, y_1) ; (x_2, y_2) ;... až (x_n, y_n) . Jednou z proměnných může být například i čas.

Závislosti, kde určité hodnotě nezávisle proměnné veličiny odpovídá jediná hodnota veličiny závisle proměnné, se nazývají závislosti funkční. Neexistuje-li mezi závisle proměnnými a nezávisle proměnnými veličinami jednoznačná funkční závislost, jde o závislosti statistické. Funkční závislost je extrémním případem závislosti statistické.

Při korelační analýze se prověřuje, zda mezi veličinami existuje nějaký vztah, zda data jsou na sobě závislá. Platí ovšem, že korelační závislost nemusí být vždy a nevyhnutelně funkční.

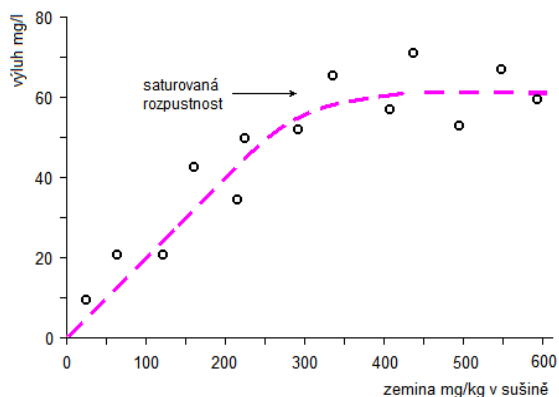
Účelem regresní analýzy je hledání charakteru závislosti mezi korelovatelnými daty. Předpokládáme, že mezi proměnnými je funkční závislost a hledáme funkci, která tento vztah závislosti aproximuje.

Korelace resp. regrese mohou být lineární nebo nelineární. Čáry proložené shluky naměřených bodů zveme regresními či korelačními přímkami resp. křivkami. V běžné literatuře lze nalézt používané výpočetní metody pro jejich stanovení.

Nejčastěji používáme lineární korelace resp. regrese. Důvodem je jejich jednoduchost.

Při lineární regresi přitom nepředpokládáme, že závislost mezi proměnnými musí být nutně lineární v celém svém rozsahu. Často ale máme důvody předpokládat, že je rozumně aproximovatelná lineární funkcí v intervalu, který je předmětem našeho zájmu.

Velmi opatrně je nutno postupovat při extrapolacích mimo rozsah naměřených dat (zvláště nebezpečné jsou extrapolace k nule).



Obr. 14. Korelace mezi koncentrací škodliviny v sušině a ve výluhu

Uvedený příkladový obrázek č. 14 zobrazuje vztah mezi koncentrací škodliviny v sušině a ve výluhu, odvozený ze souboru laboratorních stanovení. Až do úrovně saturované rozpustnosti lze závislost aproximovat lineární regresí.

Těsnost závislosti mezi dvěma proměnnými při lineární regresi vyjadřuje tzv. koeficient korelace. Podmínkou pro jeho výpočet je, aby základní soubor měl t.zv. dvojrozměrné normální rozdělení. Koeficient korelace může nabývat hodnot od -1 do +1. Hodnota $r = 0$ značí, že mezi proměnnými není žádný vztah, při $r = \pm 1$ závislost statistická přechází v závislost funkční.

3.15. Geostatistika

Klasická statistika modeluje variabilitu vzorků jako náhodný fenomén, předpokládá absolutní vzájemnou nezávislost dat. Geostatistika předpokládá určitou prostorovou kontinuitu mezi vzorky, která je je funkcí vzdálenosti mezi vzorky.

Většina parametrů horninového prostředí i kontaminace se však mění spojitě. Parametry dvou blízkých bodů jsou si podobnější než parametry vzdálenějších bodů, nejsou navzájem nezávislé. Tuto vlastnost zveme prostorovou závislostí dat. Geostatistické metody umožňují analýzu, hodnocení a charakterizování prostorově korelovatelných dat.

Geostatistika je dvoustupňovým procesem:

- v prvním kroku se definuje model, popisující prostorový funkční vztah mezi daty, tzv. semovariogram (funkce rozdílů hodnot koncentrací vždy mezi dvěma vzorkovacími body a vzájemné vzdálenosti bodů)
- v druhém kroku se provádí **krigování** - každému bodu plochy se přisoudí určitá hodnota parametru na základě vlivu okolních bodů, přičemž míra vlivu okolí je definována funkcí semi-variogramu.

Geostatistika umožňuje přisuzovat určitou velikost sledovaných parametrů i nevzorkovaným bodům a navíc umožňuje hodnotit přesnost interpretace.

Geostatistika je nástrojem, který umožňuje vyhodnotit i vzorkování, kde rozložení vzorkovacích bodů není založeno na statistických zásadách, nýbrž na odborném úsudku.

Metodu krigování využívá například hojně rozšířený PC program Surfer.

Reprezentativnost map automaticky vykreslených Surferem musíme posuzovat s rezervou. Negativně ji ovlivňují nízká hustota a nerovnoměrnosti vzorkovací sítě. Při velkých vzdálenostech mezi dokumentačními body s vysokými a nízkými koncentracemi - popřípadě ve směrech, kde nemáme omezení kontaminačního mraku dokumentováno vůbec – může být interpretován rozsah mraku o hodně větší, než odpovídá realitě.

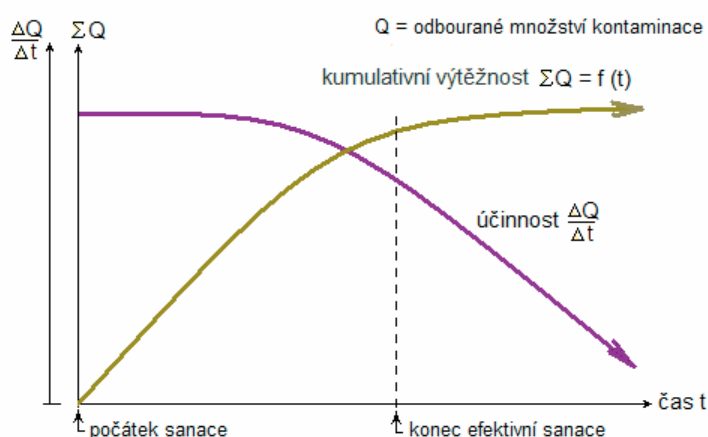
3.16. S-křivky

Pro řešení některých úloh kontaminační hydrogeologie mohou být užitečné tzv. **trendové křivky typu S**. Ty jsou vhodné zvláště k analýze a prognózám průběhu sanace v čase pomocí kumulativních křivek výtěžnosti.

Princip zobrazení průběhu sanace křivkou výtěžnosti a její charakteristický průběh ilustruje vedlejší obrázek č. 15.

Typickému průběhu výtěžnosti kontaminace odpovídá i typický vývoj účinnosti sanace v čase, jak je to ve vedlejším obrázku 15 rovněž znázorněno.

Kumulativní křivky výtěžnosti se pro účely prognóz často aproximují právě některou S-křivkou, jak je to ilustrováno na dalším obrázku č. 16.



Obr. 15. Typický časový vývoj efektivnosti sanace in-situ

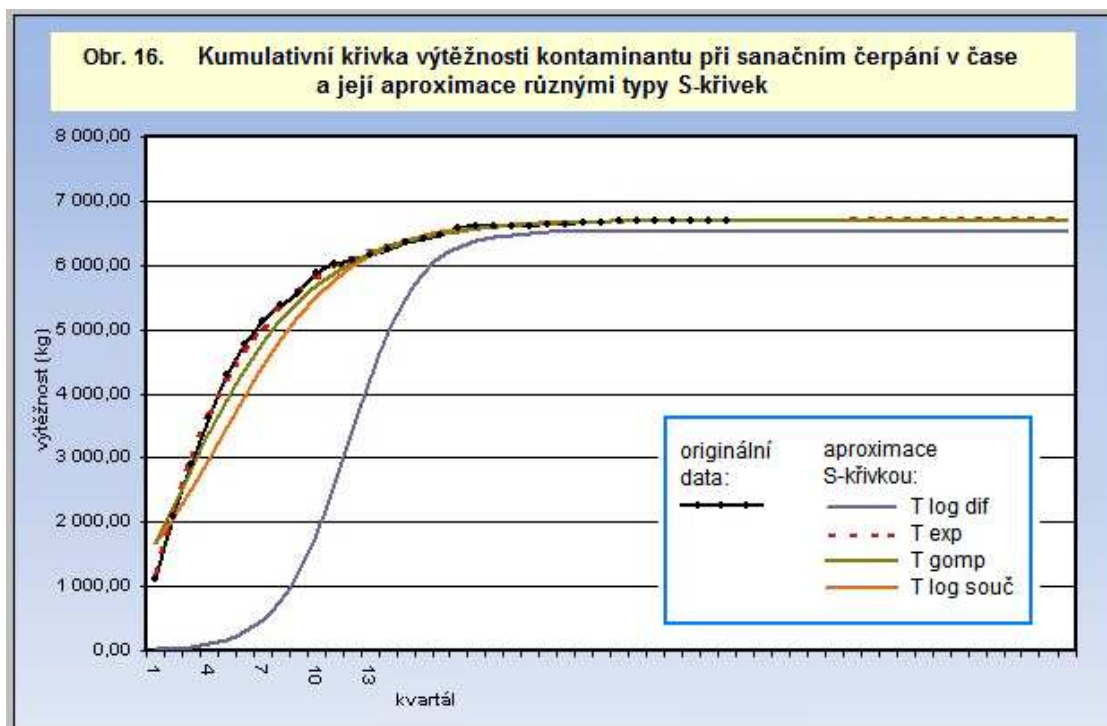
Všechny S-křivky mají charakteristický esovitý průběh, shora jsou kladně omezené. Liší se od sebe především polohou inflexního bodu. Nejčastěji používané aproximace:

- posunutá exponenciála,
- logistický trend,
- Gompertzova křivka.

Trendové křivky typu S lze s výhodou používat například pro:

- prognózování času potřebného k dosažení cílových parametrů sanačních prací v jejich průběhu,

- sledování a řízení těžby kontaminantů, kdy lze na jejich základě provádět úpravy sanačních systémů,
- zpřesňování maximálního odtěžitelného množství kontaminantu daným sanačním systémem,
- zpřesňování celkové bilance kontaminantu v průběhu sanačních prací.



3.17. Shluková analýza

Pojem shluková analýza (cluster analysis) je souhrnným názvem pro celou řadu výpočetních postupů, jejichž cílem je rozklad daného souboru dat na několik relativně homogenních podmnožin, shluků.

Rozklad množiny dat by měl být proveden takovým způsobem, aby si objekty uvnitř jednotlivých shluků byly co nejvíce podobné. Objekty patřící do různých shluků by si naopak měly být podobné co nejméně.

První monografie o shlukové analýze byla publikována roku 1939 R.C.Tyronem. Jeho původní definice zní: "Shluková analýza" je obecný logický postup formulovaný jako procedura, pomocí níž seskupujeme objektivně jedince do skupin na základě jejich podobností a rozdílností."

Pomocí shlukové analýzy tedy hledáme podobnost mezi jednotlivými údaji, přičemž se snažíme rozdělit soubor dat do skupin charakterizovaných stejnými nebo podobnými vlastnostmi.

Využití

V hydrogeologické praxi může nacházet shluková analýza široké využití, například:

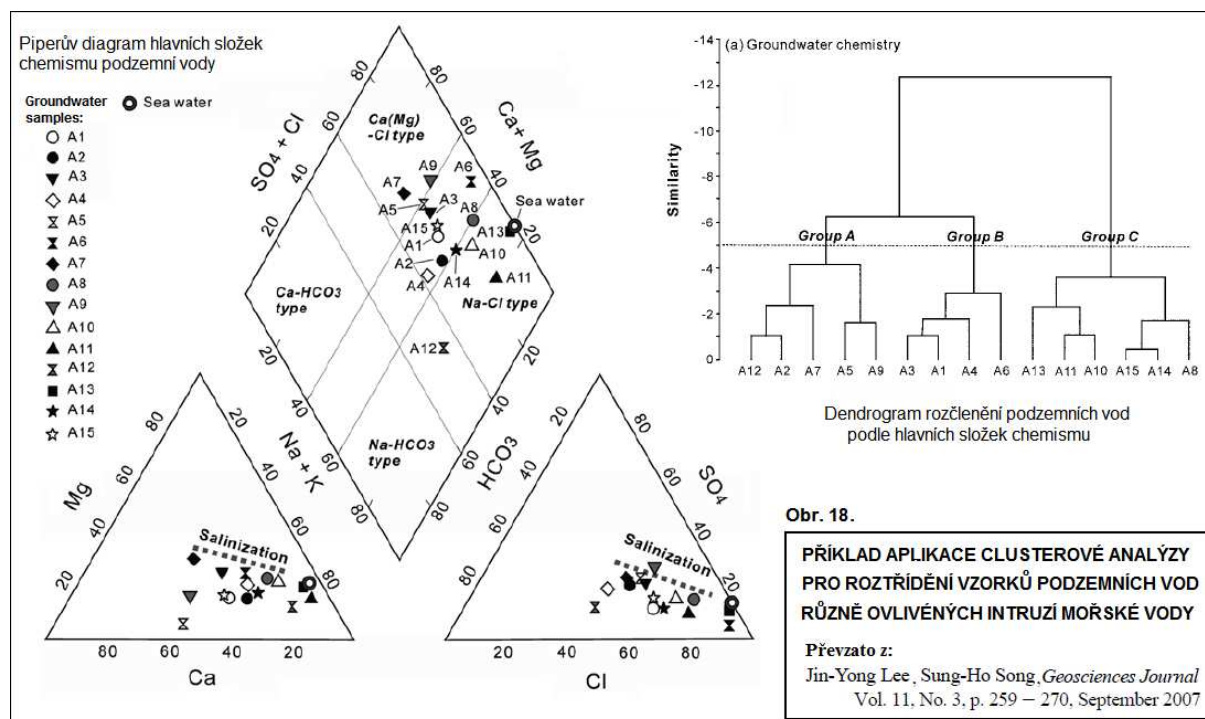
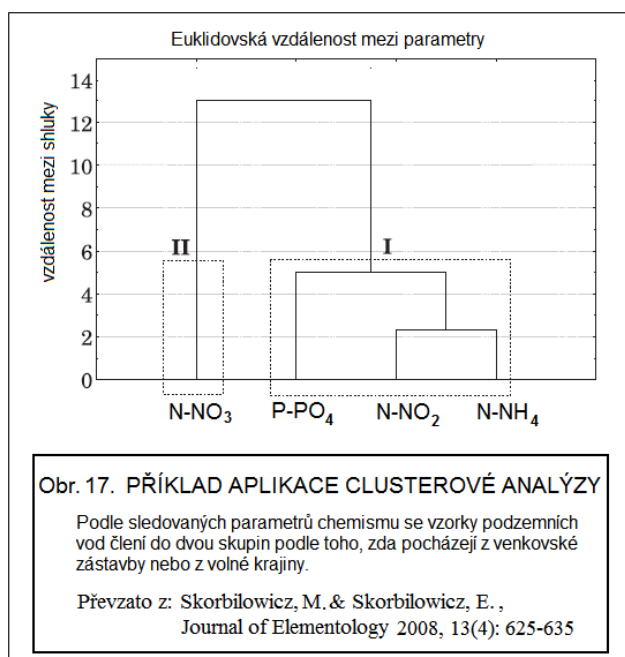
- dělení pramenů na základě chemických analýz do skupin se společnými vlastnostmi (stejného nebo podobného hydrochemického typu),

Metodická příručka Hodnocení průzkumu a sanací

- hledání podobnosti nebo odlišnosti chemického složení mezi jednotlivými ohnisky kontaminace,
- vymezování ohnisek znečištění na kontaminovaných lokalitách se složitými polymiktními mraky a hledání jejich vzájemných prostorových souvislostí.

Klasickým příkladem grafického vyjádření shlukové analýzy jsou např. Durovovy nebo Piperovy diagramy. Nejpoužívanějším způsobem vyjadřování podobnosti a shlukování/rozdělování do skupin objektů je tzv. dendrogram, který umožňuje také určovat míru podobnosti/nepodobnosti.

Dva příklady aplikace shlukové analýzy uvádějí připojené obrázky č. 17 a 18.



3.18. Rekapitulace použití statistických metod v kontaminační geologii a hydrogeologii

Primární analýza a úpravy datových souborů			
Řešený problém		Použitelná metoda	
Testy rozdělení		- vykreslení pravděpodobnostního grafu - výpočet koeficientu šikmosti nebo - Shapiro-Wilkův test (pro menší soubory do 50 vzorků) - Shapiro-Francia test (pro rozsáhlejší soubory)	
Úprava souborů s hodnotami pod mezí detekce		- náhrada nízkých hodnot poloviční hodnotou meze detekce nebo nulou (méně obvyklé) - Atchinsonova či Cohenova úprava dat (soubory s 15 až 50% dat pod mezí detekce) - rozhodnutí o použití neparametrických metod hodnocení	
Identifikace odlehých hodnot		- vykreslení a analýza pravděpodobnostního grafu - Grubbsův test - pro parametrické soubory - Dixonův test - pro neparametrické soubory	
Hodnocení průzkumů a sanací			
Řešený problém	princip postupu	médium	Použitelná metoda
Hodnocení výsledků sanace	Porovnávání se sanačními limity, odvozenými ve vztahu k ochraně zdraví či ekosystémů	zeminy podzem. voda půdní vzduch	- UCL 95 (horní mez konfidenčního intervalu) pro průměr – pro parametrické soubory - t-test jedno-výběrový - pro parametrické soubory - Chebychevův test jednostranný - pro neparametrické soubory - Wilcoxon Rank Sum test jedno-výběrový & kvantilový test pro neparametrické soubory - doplňkově pro parametrické soubory: výpočet p-hodnot
Hodnocení průzkumu – hodnocení přípustnosti určitého funkčního využívání lokality	Porovnávání s obecnými limity přípustné kontaminace, odvozenými ve vztahu k ochraně zdraví či ekosystémů		
Hodnocení průzkumu - porovnávání kontaminace lokality vůči pozadí	Porovnávání dat z lokality s požadovými koncentracemi. Pozadí může být dáno: 1) jednou hodnotou (např. nějaký limit charakterizující pozadí) nebo 2) souborem dat ze vzorkování pozadí	zeminy podzem. voda půdní vzduch	- LCL 95 pro průměr - pro parametrické soubory - t-test jedno-výběrový resp. párový - pro parametrické soubory - Chebychevův test jednostranný - pro neparametrické soubory - Wilcoxon Rank Sum test jedno-výběrový resp. párový – pro neparametrické soubory, doplněný kvantilovým testem - doplňkově pro parametrické soubory: výpočet p-hodnot
Hodnocení průzkumu - porovnávání lokality vůči pozadí	Porovnávání více než dvou souborů dat (multi-srovnávání)	podzem. voda	- ANOVA – pro parametrické soubory - Kruskal-Wallisův test - pro neparametrické soubory

Metodická příručka
Hodnocení průzkumu a sanací

Hodnocení průzkumu – kvantifikace rizika	Stanovení reprezentativní koncentrace pro vstup do výpočtu expozice	zeminy podzem. voda půdní vzduch	- UCL 95 pro průměr – pro parametrické soubory
Hodnocení využitelnosti podzemních vod pro pitné účely	Porovnávání s limity přípustných koncentrací	podzem. voda	Statistiku nelze použít - limit nesmí být překročen v žádném místě zvodně (tj. v žádném vzorku).
Monitoring podzemních vod			
Řešený problém	princip postupu	médium	Použitelná metoda
Detekční monitoring	Indikace vzniku kontaminace v areálu sledované lokality nad úroveň pozadí	podzem. voda	- porovnávání výsledků z areálu lokality s horní mezí tolerančního limitu UTL, odvozeného z požadových dat - porovnávání nového vzorkování z vrtu v areálu lokality s UTL, odvozeným z předchozích výsledků téhož vrtu - porovnávání výsledků z areálu lokality s horní mezí predikčního limitu UPL, odvozeného z požadových dat - porovnávání nového vzorkování z vrtu v areálu lokality s UPL, odvozeným z předchozích výsledků téhož vrtu
Kontrolní monitoring	Indikace vzniku kontaminace nad úroveň limitu (kritéria) přípustné kontaminace		- porovnávání obecného limitu (kritéria) s horní mezí tolerančního limitu UTL, odvozeného z výsledků dosavadního monitoringu podzemních vod v areálu lokality nebo na výstupu z ní
Hodnocení časového vývoje kontaminace (detekční, kontrolní, sanační, post-sanační monitoring)	Grafické zobrazení vývoje znečištění v čase, hodnocení dosavadních trendů a prognózování budoucího vývoje koncentrací, indikace vzniku kontaminace		Časové řady a jejich hodnocení: - klouzavé průměry - Shewartova - CUSUM metoda - exponenciální vyrovnávání časových řad - regresní analýza - pro neparametrické soubory Mann-Kendallův test
Další aplikace statistiky			
Řešený problém	princip postupu	médium	Použitelná metoda
Hledání vztahů mezi sledovanými parametry (proměnnými)	Grafické zobrazení vztahů mezi proměnnými, analýza grafů		Grafické či výpočtové prokládání přímk (křivek) korelačními grafy, výpočet koeficientu korelace
Hodnocení průzkumu - plošná interpretace výsledků vzorkování	Vykreslování map znečištění		Krigování
Rozčlenění dat do několika skupin různého charakteru	Vymezování skupin na základě příbuzných vlastností		Clusterová (shluková) analýza

Metodická příručka
Hodnocení průzkumu a sanací

Hodnocení průběhu sanace - analýza vývoje odtěžování kontaminace v čase	Grafické zobrazení výtěžnosti kontaminace v čase, analýza grafů pro prognózování dalšího vývoje, pro upřesňování bilancí, pro optimalizaci sanačních systémů		Prokládání –S-křivek grafy kumulativní výtěžnosti
Alternativní pravidla, jejichž využití pro naši praxi se nabízí ke zvážení			
Řešený problém	princip postupu	médium	Použitelná metoda
Hodnocení výsledků sanace	Porovnávání se sanačními limity obecného (nikoliv site-specific) charakteru, odvozenými ve vztahu k ochraně zdraví či ekosystémů	zeminy	<u>Pravidlo 75%/10x</u> (PDEP, 2008): 75% vzorků musí splňovat limit, přičemž v žádném ze vzorků nesmí být překročen limit více než 10ti násobně.
Hodnocení průzkumu – hodnocení přípustnosti určitého funkčního využívání lokality	Porovnávání s obecnými limity přípustné kontaminace, odvozenými ve vztahu k ochraně zdraví či ekosystémů		
Hodnocení výsledků sanace	Porovnávání se sanačními limity obecného (nikoliv site-specific) charakteru, odvozenými ve vztahu k ochraně zdraví či ekosystémů	podzemní voda	<u>Pro podzemní vody v areálu lokality pravidlo 75%/10x</u> : 75% vzorků musí splňovat limit, přičemž v žádném ze vzorků nesmí být překročen limit více než 10ti násobně. <u>Pro podzemní vody na výstupu z areálu lokality pravidlo 75%/2x</u> : 75% vzorků musí splňovat limit, přičemž v žádném ze vzorků nesmí být překročen limit více než dvojnásobně. (PDEP, 2008)
Hodnocení průzkumu – hodnocení přípustnosti určitého funkčního využívání lokality	Porovnávání s i limity přípustné kontaminace, odvozenými ve vztahu k ochraně zdraví či ekosystémů		
Hodnocení výsledků sanace	Porovnávání s obecnými sanačními limity nebo obecnými kritérii přípustné kontaminace, odvozenými pro tyto situace:	zeminy půdní vzduch podzem. voda	UCL95 pro průměr musí být pod limitem (kritériem) a žádný ze vzorků nesmí tento limit (kritérium) překročit více než dvojnásobně. (CDEQ, 1996)
Hodnocení průzkumu – hodnocení přípustnosti určitého funkčního využívání lokality	- přípustná kontaminace zemin nenasycované zóny z hlediska přestupu znečištění do podzemních vod - přípustná kontaminace zemin přípovrchové vrstvy z hlediska rizik z přímého kontaktu či z ingesce - přípustná kontaminace půdního vzduchu z hlediska rizik z inhalace těkavých látek - přípustná kontaminace podzemních vod na výstupu z lokality z hlediska ohrožení dalších receptorů po směru proudění		

4. STRATEGIE A METODIKA VZORKOVÁNÍ

4.1. Konceptuální model lokality

Prvním předpokladem pro volbu správné vzorkovací strategie a metodiky vzorkování je definování výstižného konceptuálního modelu lokality.

V rámci konceptuálního modelu musíme definovat výchozí představu o lokalitě: identifikace zdrojů a typů škodlivin, formulování konzistentní představy o možném vzniku kontaminace, o jejím charakteru, o předpokladech jejího rozložení v ploše a možnostech její další migrace v závislosti na místních přírodních i jiných podmínkách.

Základním kritériem pro úplnost a správnost koncepčního modelu lokality je, že nesmí být v rozporu s žádnými známými fakty o lokalitě, naopak musí logicky a konzistentně vysvětlovat situaci, která na lokalitě je. A to ze všech možných úhlů pohledu.

Konceptuální model lokality musíme průběžně modifikovat podle postupně získávaných výsledků.

Na základě předběžných znalostí o lokalitě a její historii se snažíme zařadit lokalitu za účelem volby vzorkovací strategie do některé z těchto základních kategorií:

- lokalita bez podezření na kontaminaci,
- podezřelá lokalita s difúzním (celoplošným) charakterem znečištění,
- podezřelá lokalita s ohniskovými zdroji kontaminace, jejichž lokalizace je známa,
- podezřelá lokalita s ohniskovými zdroji kontaminace, jejichž lokalizace není známa.

Každé kategorii odpovídá jiný druh vzorkovací strategie, jejímž cílem je ověřit správnost formulovaného konceptuálního modelu lokality. Je-li formulován chybně vstupní model, mohou být chybné i závěry průzkumu o rozsahu a míře kontaminace. Je evidentní, jak důležitou úlohu má v celém procesu právě přípravná fáze.

Rozsah i zaměření vzorkovacích prací jsou samozřejmě závislé i na cílech průzkumu. Při hodnocení rizik z kontaminace horninového prostředí musí proto konceptuální model zahrnovat i identifikaci příjemců kontaminace v závislosti na stávajícím i budoucím využívání lokality a okolí, identifikaci možných expozičních cest, identifikaci konečných recipientů migrace znečištění.

4.2. Strategie vzorkování

- Účel vzorkování: získání spolehlivé - reprezentativní - informace o zkoumaném objektu, o zkoumaném základním souboru (populaci), na základě informací ze vzorků, které jsou výběrovým souborem z tohoto základního souboru. Reprezentativnost je vyžadována jen pro ty parametry, které jsou předmětem zájmu.
- Reprezentativnost získané informace podmíněna:
 - strategií vzorkování,
 - metodikou odběru vzorku,
 - metodikou analýz.

Metodická příručka Hodnocení průzkumu a sanací

- Při volbě strategie a metodiky vzorkování musíme mít na mysli možné variability zkoumaného objektu, které jsou v našem případě v zásadě dvojího druhu:
 - variabilita horninového prostředí,
 - variabilita kontaminace.
- Volba správné strategie vzorkování předpokládá, že máme povědomost o těchto variabilitách a dokážeme zkoumané prostředí rozčlenit do menších jednotek, které jsou co nejvíce homogenní. Jen tak dokážeme správně interpretovat výsledky vzorkování a vyvozovat z nich správné závěry.
- Variabilita horninového prostředí - musíme znát pro volbu správné strategie:
 - znalost vertikálního geologického profilu,
 - znalost laterálního vývoje geologického profilu.

Nejdůležitější vlastnosti: petrografie, zrnitost, pórovitost, obsah jílové frakce, obsah organického uhlíku, výměnná iontová kapacita, sycení vodou.

- Variabilita kontaminace - musíme znát pro volbu správné strategie:
 - předběžná znalost o lokalizaci pravděpodobných míst kontaminace,
 - identifikace charakteru možné kontaminace,
 - identifikace možných migračních cest kontaminace a jejich charakter.
- Možné strategie vzorkování:
 - založené na statistických přístupech,
 - založené na odborném úsudku (někdy zvané autoritativní),
 - smluvní.
- Strategie vzorkování a hodnocení jeho výsledků musí být voleny se zřetelem k jeho konkrétnímu účelu.

V podstatě existují tyto hlavní cíle vzorkování lokality s kontaminací:

- charakteristika lokality – ověření charakteru, rozsahu a úrovně kontaminace, identifikace ohnisek znečištění,
- získání reprezentativní koncentrace škodlivin pro hodnocení rizika – pro posouzení, zda kontaminace převyšuje či nepřevyšuje přípustné koncentrační limity odpovídající stávajícímu či zamýšlenému funkčnímu využívání (je-li limit přípustné kontaminace vyšší než reprezentativní koncentrace, pak je daný způsob využívání lokality přípustný a naopak) – v tomto případě musíme u některých scénářů zohledňovat i velikost expoziční – průměrovací plochy, pro kterou se reprezentativní koncentrace stanovují,
- získání reprezentativní koncentrace škodlivin pro porovnání s pozadím,
- získání reprezentativní koncentrace škodlivin pro porovnání se sanačními limity – pro posouzení, zda bylo dosaženo cílů sanace či nikoliv (je-li sanační limit vyšší než reprezentativní koncentrace, pak byla sanace úspěšná a naopak).

U podzemních vod je předmětem zájmu navíc ještě vývoj kontaminace v čase.

4.3. Vzorkovací schémata

4.3.1. Vzorkování založené na statistických přístupech

Často nazývané náhodné vzorkování. Musí být splněna velmi přísná kritéria, aby bylo zaručeno, že jde o vzorkování skutečně náhodné.

Základní podmínky:

- správné vymezení základního souboru (populace) – musí být zaručeno, že soubor dat pochází opravdu z jedné populace,
- statisticky opravdu náhodná specifikace výběrového souboru (každé místo lokality musí mít stejnou pravděpodobnost, že bude ovzorkováno).

První z uvedených podmínek striktně vyžaduje, že data z jedné lokality nemohou být statisticky vyhodnocena jako jeden soubor, pokud jsou v něm vzorky u kterých se koncentrace neodlišují od pozadí i vzorky s vysokými koncentracemi z ohnisek plošně omezeného rozsahu. Statistické vyhodnocení dat z ohnisek kontaminace společně s ostatními daty z lokality je zásadně nekorektní z hlediska matematiky a vedlo by to k „naředování“ kontaminace (doporučení postupu pro takové případy viz konec kapitoly 3.5).

Při hodnocení kontaminace podzemních vod však lze hodnotit jako jeden soubor všechna data z jednoho průtočného profilu bez ohledu na výši kontaminace na jednotlivých vrtech profilu.

Statistická analýza dat není použitelná pro porovnávání kontaminace pitné vody s limity pro vodu pitnou ve zvodni, která je nebo mohla by být využívána pro takový účel. Kvalita vody musí vyhovovat limitům ve všech místech zvodně.

Jednoduché náhodné vzorkování

Každý element plochy má stejnou šanci být zahrnut do výběru.

Zavede se síť, každému elementu sítě se přiřadí číslo, výběr ke vzorkování losováním (generováním náhodných čísel).

Jedním z důvodů proč se náhodné vzorkování v praxi příliš nepoužívá jsou obtíže s vytyčováním vzorkovacích bodů v terénu. Významnou pomocí by zde mohlo být použití GPS.

Potřebný počet vzorků závisí na požadované spolehlivosti výsledků a dá se určit výpočtem. To však vyžaduje, aby některé ze statistických charakteristik souboru byly odhadovány předem. Existují i empirická pravidla, použitelná pro průzkum kontaminace horninového prostředí (viz kapitola 5).

Náhodné vzorkování dává dobrou informaci o střední hodnotě a o intervalu spolehlivosti, negarantuje však dostatečnou informaci o plošné distribuci kontaminace a o výskytu a ohraničení ohnisek, pokud není vzorkování opravdu dostatečně husté.

Náhodné vzorkování je tedy vhodné zejména pro homogennější areály. Jsou-li ohniska kontaminace plošně malá, pak je nemusíme při tomto způsobu vzorkování vůbec zachytit.

Na druhé straně, kdybychom se zaměřili při vzorkování na identifikaci ohnisek, nepůjde již o statisticky náhodné vzorkování.

Při statistickém zpracování nelze sdružovat oba typy vzorkování z jedné lokality do jednoho souboru. Vhodný postup: vymezit podezřelé plochy a ty ovzorkovat zvlášť - nejméně 2 vzorky z ohniska a hustota vzorkování shodná jako v ostatní (nepodezřelé) části lokality, kde

vzorkování náhodné. Známe-li podezřelou plochu předem, lze udělat návrh na stratifikované vzorkování - metodika zpracování výsledků pak umožní kombinovaně hodnotit výsledky z ohniska a ze zbytku lokality.

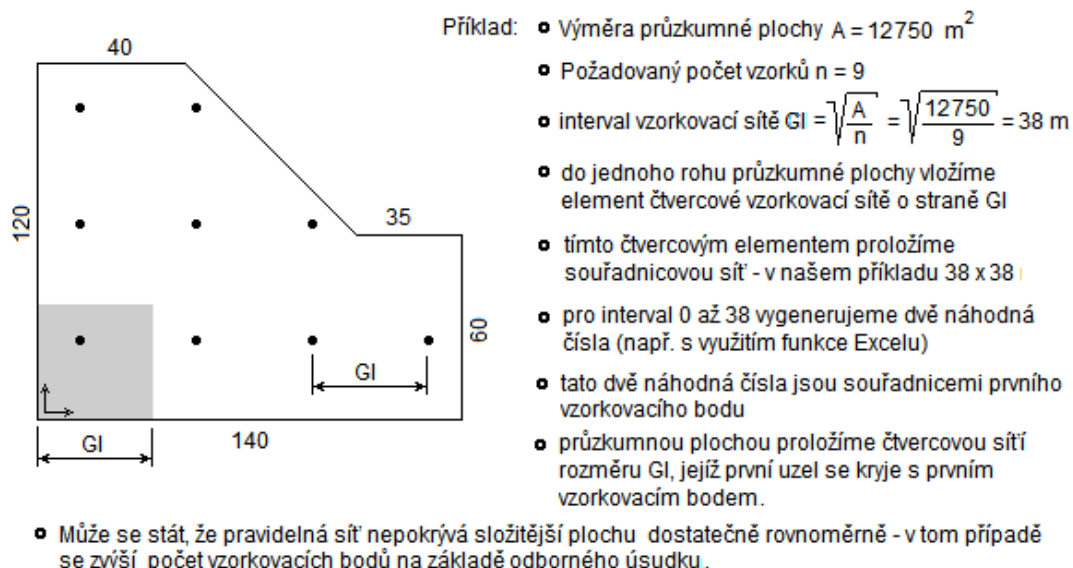
Často však lze data z úvodního o úsudkového vzorkování alespoň částečně zahrnout do souboru z následné vzorkovací etapy, ve které bylo aplikováno vzorkování pro vyhodnocení statistickými metodami.

Problematicku dodatečné stratifikace lokality podle výsledků provedeného vzorkování pojednává konec kapitoly 3.5.

Systematické náhodné vzorkování (v síti)

Území se rozdělí do elementů pravidelnou sítí, vzorkuje se v uzlech sítě. Rozměr elementů sítě (počet uzlů) se odvodí z počtu vzorků, vypočteného podle požadované úrovně intervalu spolehlivosti. Náhodným prvkem je zde volba počátečního bodu sítě. Vzorkovací síť se do plánu lokality vkládá tak, aby se jeden z uzlů překrýval s prvním vzorkovacím bodem, který je zvolen předem čistě náhodně. Postup znázorňuje připojený obrázek 19.

Obr. 19 . Systematické náhodné vzorkování ve čtvercové síti



Plocha je při tomto způsobu pokryta odběrnými body pravidelně, body jsou od sebe stejně vzdálené. Většinou se používá čtvercová síť. Méně vzorkovacích bodů vyžaduje síť trojúhelníková, která je však náročnější na vytýčení.

Analogií u lokalit protáhlého tvaru je ekvidistantní vzorkování na liniích.

Určité odchylky od systematického vzorkování v pravidelné síti (například při respektování staveb v areálu lokality) není třeba brát nijak dramaticky jako zásadní překážku aplikace statistických postupů při hodnocení vzorkování.

Systematické vzorkování je vhodné pro kontaminované areály, kde lze očekávat koncentrační gradienty. Je vhodné pro zpracování dat při použití geostatistiky.

Hustotu vzorkovací sítě možno volit z vícera zorných úhlů:

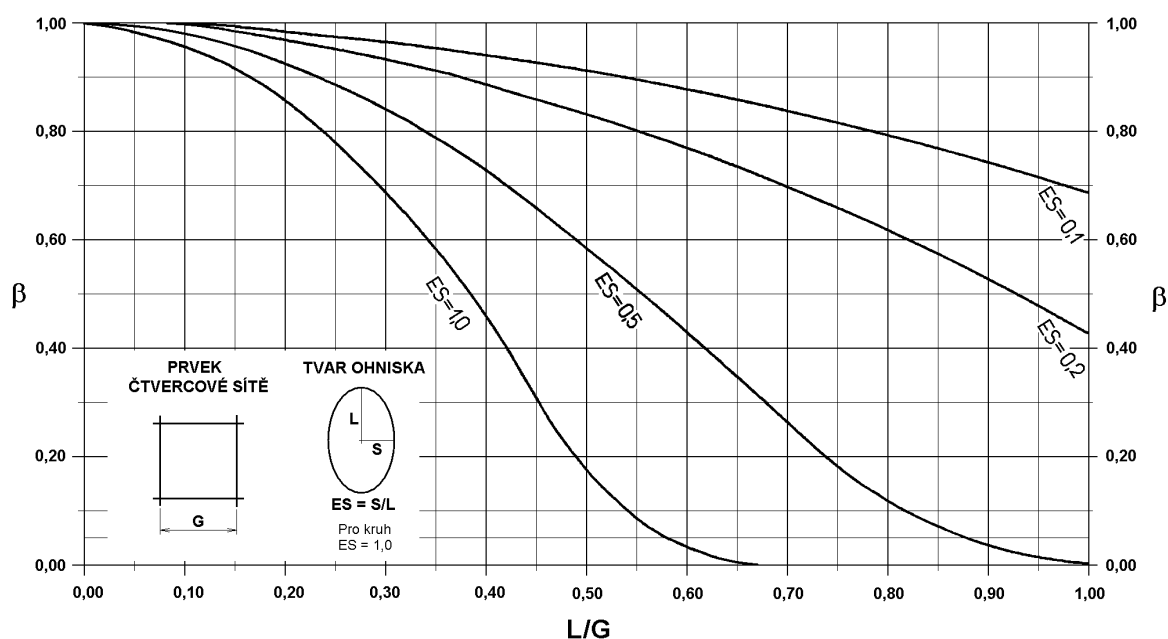
- velikost ohniska znečištění, které má být vzorkováním podchyceno (viz graf níže),

Metodická příručka Hodnocení průzkumu a sanací

- velikost expoziční plochy, pro kterou má být stanovena reprezentativní koncentrace pro účely hodnocení rizika (viz subkapitola 4.6.),
- stanovení dostatečně reprezentativní koncentrace areálu celé lokality pro srovnání s pozadím (podle statistických kritérií),
- stanovení reprezentativní koncentrace sanované plochy se sanačním limitem (podle statistických kritérií).

S pomocí dalšího obrázku č. 20 lze stanovit pravděpodobnost nepodchycení okrouhlého ohniska při daném poměru hustoty systematické vzorkovací sítě a velikosti ohniska.

Obr. 20. Pravděpodobnost β nepodchycení ohniska kruhového či eliptického tvaru při systematickém vzorkování ve čtvercové síti (dle Zirchsky a Gilbert, 1984)



Tabulka č. 4 odvozená s použitím grafu z obrázku 20 ilustruje vypovídací schopnost systematického vzorkování z hlediska garance podchycení ohnisek určité velikosti při dané hustotě sítě. Uvedena je i relativní nákladovost průzkumu.

Tab. 4. Hustota sítě vzorkování a vypovídací hodnota průzkumu kontaminace						
hustota sítě vzorkování	počet vzorků na 1 ha	Pravděpodobnost (%) nepoškození kruhové skvrny kontaminace při jejím průměru:				relativní cena průzkumu
		5 m	10 m	25 m	50 m	
m	ks					
100 m	1	99	96	80	20	1
71 m	2	98	95	60	0	2
50 m	4	96	87	20	0	4
35 m	8	93	73	0	0	8
25 m	20	82	43	0	0	20
18 m	30	76	27	0	0	30
10 m	100	52	0	0	0	100
5 m	400	0	0	0	0	400

V zásadě je rozestup sítě blízký rozměru ohniska, které nechceme vzorkováním pominout.

Údaje tabulky č. 19 o počtech vzorků na hektar a údaje o relativní ceně průzkumu se vztahují jen k průzkumu v jedné horizontální rovině. Při ověřování hloubkové distribuce kontaminace bude cena průzkumu dále vzrůstat.

Je evidentní, že by bylo neefektivní, snažit se tímto přístupem identifikovat plošně malá ohniska v areálu rozsáhlé lokality. Vyžadovalo by to velký počet vzorků.

Systematické vzorkování by mohlo vést k zásadně chybné interpretaci výsledků, když by na lokalitě existovala možnost cyklické distribuce kontaminace v ploše (viz kapitola 3.4.).

Stratifikované náhodné vzorkování

Areál je rozdělen do několika dílčích, quasi-homogenních celků a každý z nich se pak vzorkuje samostatným náhodným vzorkováním nebo náhodným vzorkováním v síti. Soubor dat z každého dílčího úseku se vyhodnocuje samostatně.

Podstatným je zde způsob rozdělení plochy do dílčích celků, který musí eliminovat určité druhy variability. Dílčí části musí být homogennější než celek, jinak metoda ztrácí smysl.

Použití tohoto přístupu je podmíněno předběžnou znalostí distribuce znečištění v zájmové ploše. Rozdělení areálu do dílčích celků může být založeno například na: předběžných indikacích o ohniscích kontaminace, topografii, geomorfologii, typu půdy, na rozdílech ve funkčním využívání území, na předpokládaném způsobu sanace.

Tento způsob vzorkování tudíž zpravidla vyžaduje předběžné vzorkovací práce (založené na odborném úsudku), za účelem co nejvýstižnějšího definování dílčích celků pro samostatné ověřování.

Metoda umožňuje přesnější ohodnocení dílčích celků a přesnější informaci o celku. Také lze pak vyhodnocovat statisticky variabilitu mezi dílčími celky.

Dílčí celky se nesmí překrývat - součet dílčích ploch se musí rovnat ploše celkové.

4.3.2. Vzorkování založené na odborném úsudku

Vzorkování založené na statistických přístupech se z finančních důvodů uplatňuje v praxi zřídka, i když náklady na reprezentativní ověření kontaminace bývají obvykle jen zlomkem nákladů na sanaci kontaminace.

Co se týče horninového prostředí, pak stratifikace vrstev, plošné i vertikální heterogenity horninového prostředí, v čase se měnící režim vlhkosti, infiltrace a proudění podzemních vod, kvalitativní omezení vzorkovacích technologií a další komplikující faktory jsou příčinou, že zajištění reprezentativního vzorkování je problematické v každém případě. Obvyklým je v tomto případě vzorkování, založené právě na odborném úsudku.

Prakticky vždy je vzorkování na základě úsudku první etapou vzorkování, kdy dosud nejsou k dispozici žádné informace o skutečné kontaminaci horninového prostředí. Vzorkovací body umísťujeme na základě předpokladů o zdrojích potenciální kontaminace, prioritou je identifikace ohnisek. Zpravidla se vrtý lokalizují nejdříve do předpokládaných ohnisek kontaminace s postupným rozšiřováním prací pro vymezení hranic kontaminačního mraku. Počet vzorků v ploše se volí na základě odborného úsudku, podle charakteru lokality a podle předpokládané kontaminace.

I u vzorkování na základě odborného úsudku je úroveň poznání lokality a reprezentativnost výsledků vzorkování závislá na vztahu počtu vzorků k velikosti studované plochy.

Hloubky vzorkování zemin se volí podle charakteru kontaminace, podle účelu průzkumu a podle charakteru migračních cest kontaminace. Vzorkování podzemních vod u mocnějších kolektorů ideálně z různých hloubek v pravidelných intervalech.

Statistické metody jsou pro hodnocení výsledků úsudkového vzorkování zásadně nepoužitelné. Hodnocení výsledků je nutno provádět pro každý vzorkovací bod jednotlivě.

Použijeme-li však na téže lokalitě v další etapě i statistické vzorkování, pak je často možné, že část dat nebo i všechna data z předchozího úsudkového vzorkování mohou být do statistických souborů zahrnuta.

Významným nástrojem pro hodnocení vzorkování i na základě úsudku však mohou být metody geostatistiky (viz. kap. 3.17.).

Nevýhody a námitky vůči reprezentativnosti: závislé na odborné erudici a zkušenosti, je individuální, nelze prověřit spolehlivost odborného odhadu a tudíž nemáme žádnou záruku objektivnosti výsledků, nelze posoudit žádným objektivním způsobem reprezentativnost vzorkování, získaný soubor neodpovídá kritériím pro statistické zpracování výsledků.

Výsledky vzorkování na základě odborného úsudku bývají východiskem pro vymezení dílčích částí celé lokality, které lze pak považovat za quasi-homogenní, s vnitřní variabilitou sledovaných parametrů výrazně nižší, než má lokalita jako celek. Tyto pak prověřujeme statistickým vzorkováním v následné průzkumné etapě.

4.3.3. Kombinace vzorkovacích strategií

V praxi se často vzorkovací strategie v rámci lokality kombinují. V úvodní etapě vzorkování je primárním cílem identifikace ohnisek kontaminace a vzorkovací strategie je založena na odborném úsudku s využitím informací které možnou lokalizaci ohnisek naznačují. Nicméně existuje riziko, že zůstanou nepodchycena ohniska, o jejichž možné existenci nemáme ani tušení. Pro snížení tohoto rizika se vzorkování doplní vzorkováním v síti, jehož hustota se volí s přihlédnutím k předpokládané velikosti a tvaru možných ohnisek (viz graf v obrázku 26).

4.3.4. Vzorkování smluvní (předpisové)

Určitý kompromis, snažící se o určitou, alespoň relativní reprezentativnost vzorkování, umožňující pak například srovnávání různých lokalit mezi sebou. Založen na empirickém stanovení pravidel, zpravidla na základě zhodnocení statistických modelů, ale s rezignací na to, aby metodika odpovídala pro konkrétní lokalitu plně statistické teorii.

Příklad: stanoven je určitý počet vzorků v závislosti na rozměrech kontaminované plochy.

4.4. Kompozitní (směsné) vzorkování

Je to možnost úspory nákladů na laboratorní práce. Z několika odebraných vzorků - v praxi nejvýše do 5 ks - se připraví jeden vzorek směsný, který se analyzuje. Směsné vzorky se připravují smícháním, kvartací a homogenizací diskrétních vzorků o stejné váze.

Pokud se použije směsné vzorkování v pravidelné síti, pak se každý čtverec sítě rozdělí na 4 stejné díly a z každého sub-čtverce se odebere jeden vzorek pro přípravu vzorku směsného.

V principu je směsné vzorkování nepoužitelné pro statistické metody interpretace dat. Lze jím ověřit jen průměr. Nelze tímto způsobem studovat odchylky a další statistické parametry kontaminace, nelze definovat intervaly spolehlivosti.

Nabízí se využití kompozitního vzorkování v kombinaci se stratifikovaným vzorkováním. Průměry z kompozitních vzorků z jednotlivých dílčích celků se pak mohou využít k výpočtu variancí a dalších statistických parametrů mezi dílčími celky. Existují metody které umožňují výpočet UCL95 pro průměr při směsném vzorkování (Edland et al., 1994; Patil et al., 1994).

Maximální možná koncentrace bude odpovídat určené koncentraci ve směsném vzorku, násobeno počtem vzorků, ze kterých byl sestaven.

Předpokladem pro použití směsného vzorkování je homogenita matrice.

Kdy směsné vzorkování zásadně používat nelze:

- pro stanovování těkavých uhlovodíků (únik těkavých složek při homogenizaci vzorku)
- pro vymezování ohnisek kontaminace v ploše
- pro zeminy s vysokým podílem jílovité frakce (problematická homogenizace vzorku)
- pro analýzy výluhů.

Směsné vzorkování se rovněž zásadně nepoužívá pro podzemní vody.

Kompozitní vzorky z plochy není vhodné používat v případech, kdy jsou ohniska kontaminace plošně malá.

4.5. Reprezentativní koncentrace

Do výpočtů pro stanovení expozice přípustného znečištění při analýze rizika a do porovnávání úrovně kontaminace určité lokality s limity či kritérii přípustného znečištění vstupují tzv. **reprezentativní koncentrace složek horninového prostředí**. Mohou to být průměry, určité kvantily konfidenční intervaly, maximální hodnoty.

V praxi se mohou vyskytovat situace, kdy reprezentativní koncentraci nelze stanovit jako jednu hodnotu odvozenou ze všech dat pro celou lokalitu a kterou by bylo možno k celé lokalitě vztahovat. Pro některé expoziční scénáře musí být reprezentativní koncentrace stanovena **se zohledněním velikosti průměrovací (expoziční) plochy**. Průměrovací plocha pro kterou stanovujeme reprezentativní koncentraci tak někdy nemusí být totožná s celou plochou zkoumané lokality. Její velikost může být odvislá od charakteru funkčního využívání, pro který se rizika hodnotí (viz též kapitola 4.6.).

Pojem reprezentativních koncentrací musíme mít ujasněn i **při posuzování výsledků sanace**, při porovnávání výsledků vzorkování se sanačními limity. Zde zpravidla reprezentativní koncentrace určujeme pro celou sanovanou plochu resp. celý objem sanovaného prostoru. Sanační limity bývají těchto druhů:

- jsou založeny na výsledcích analýzy rizika konkrétně pro danou lokalitu – stanovení přípustné koncentrace zpětným propočtem pro danou úroveň akceptovatelného rizika,
- jsou to legislativou obecně stanovené hodnoty přípustné kontaminace (zpravidla odvozené rovněž od akceptovatelné úrovně rizika, jsou stanoveny ve vztahu k určitému způsobu funkčního využívání území),
- jsou založeny na vztahu k pozadí (sanace na úroveň pozadí),
- jsou dány technologicky (vycházejí z možností aplikované sanační technologie).

Při hodnocení výsledků sanace se volí jako nulová hypotéza H_0 pro testování předpoklad, že lokalita je kontaminovaná.

Metodická příručka Hodnocení průzkumu a sanací

Rovněž je zde nutno připomenout základní zásadu, že jako jeden datový soubor nelze vyhodnocovat všechna data z lokality, pokud pocházejí z různorodých částí co se týče úrovně znečištění (data z ploch ohnisek kontaminačních mraků a data z ploch, majících jen pozadové znečištění).

Posuzování podle průměrných koncentrací nebo podle konfidenčních intervalů kolem průměru: snadno se počítá a všeobecně stačí mít k dispozici méně vzorků pro dosažení určité úrovně spolehlivosti ve srovnání s posuzováním podle jiných parametrů.

Průměry jsou vhodné pro:

- relativně homogenní lokality - nebo lokality kde technologie sanace vede k homogenizaci prostředí, takže rozptyl hodnot kolem průměru není příliš velký;
- lokality, kde jsou limity založeny na zdravotních rizicích z dlouhodobé či průměrné expozice karcinogenům a chronicky toxickým látkám;
- pro lokality, kde má jen málo vzorků hodnoty pod mezí detekce.

Použití průměrů není vhodné pro případy:

- když se uplatňuje akutní toxicita kontaminantů;
- když je velká část dat pod detekčním limitem;
- kde je kontaminace heterogenní - kde existují v ploše lokality jen plošně malá ohniska s vysokou kontaminací na mnohem větší ploše relativně čistého pozadí - při použití průměrů je výsledkem určité „naředění“ znečištění.

Podle charakteru řešeného problému a v závislosti na formulaci tzv. nulové hypotézy dosazujeme za reprezentativní koncentrace horní nebo spodní meze konfidenčních intervalů pro průměr (viz kapitola 3.12.).

Zmínku si zde zaslouží upozornění na to, že při výpočtech reprezentativních koncentrací pro hodnocení on-site expozice při analýzách humánního rizika používáme aritmetický průměr bez ohledu na skutečný charakter rozdělení souboru dat ze vzorkování. Aritmetický průměr je zde reprezentativní hodnotou ve vztahu k náhodnému pohybu potenciálního příjemce kontaminace v areálu lokality (US EPA, 1992).

Medián (=50% percentil): je alternativou k použití průměrů - není tolik ovlivněn extrémními odchylkami dat (odlehle hodnoty), variabilitou dat, existencí většího podílu nízkých hodnot pod mezí detekce (MDL). Použití mediánu nevyžaduje splnění tolika předpokladů, jako v případě průměrů.

Percentily: horní percentily je vhodné je používat v případech, kdy:

- distribuční křivky souborů dat jsou značně sešikmené,
- velký podíl dat je pod mezí detekce,
- sanační limity jsou založeny na hodnocení rizik z expozice kontaminantům s akutní toxicitou nebo když se při analýze rizik používají worst-case scénáře.

Opět, podle charakteru řešeného problému a v závislosti na formulaci tzv. nulové hypotézy dosazujeme za reprezentativní koncentrace horní nebo spodní meze konfidenčních intervalů pro daný percentil (viz kapitola 3.12.1.).

Maximální hodnoty: používají se při velmi malém počtu dat, kdy nemůžeme určit charakter rozdělení prošetřovaného souboru.

Individuální hodnoty: pokud soubor dat neodpovídá požadavkům na statistické hodnocení, musí se každá hodnota porovnávat s hodnotícím limitem (kritériem) individuálně.

Specifickým případem je situace, kdy hodnotíme přípustnost kontaminace tělesa podzemních vod z hlediska jejich využívání pro pitné účely. Zde musíme rovněž hodnotit data z každého vzorkovacího vrtu individuálně, neboť voda musí vyhovovat limitu v kterémkoliv místě zvodně.

4.6. Průměrovací - expoziční plocha a její velikost

Rozčlenění celé zkoumané lokality do dílčích celků a hustota vzorkování musí rovněž zohledňovat expoziční scénáře, podle kterých se budou provádět výpočty rizik. Do výpočtů pro stanovení expozice a do porovnávání úrovně kontaminace lokality s limity či kritérii přípustného znečištění vstupují vždy tzv. reprezentativní koncentrace složek horninového prostředí. Zpravidla to bývají průměry, mediány, určité kvantily či meze konfidenčních intervalů.

K jaké ploše se budou reprezentativní koncentrace vztahovat závisí také na expozičním scénáři, na charakteru funkčního využívání kontaminované plochy. Mohli bychom též hovořit o jednotkové expoziční ploše, expoziční jednotce. Soubor dat pro porovnávání s limitem přípustné kontaminace musí odpovídat velikosti takové expoziční jednotky. Nemusí to být vždy celá plocha zkoumané lokality

Nejlépe to osvětlí příklad:

Pro hodnocení rizik z neúmyslné ingesce kontaminovaných zemin v rozsáhlém průmyslovém areálu můžeme předpokládat náhodný pohyb zaměstnanců v rámci celé plochy. Pro stanovení reprezentativní koncentrace (třeba průměru) škodliviny v zemině můžeme v tomto případě oprávněně použít data z celé plochy lokality, bez ohledu na velikost a konkrétní pozici ohnisek znečištění.

Jiná situace však nastane tehdy, když stejná lokalita se stejnou kontaminací má být v budoucnu rozdělena na menší stavební parcely pro rodinné domky. Typická plocha jedné takové parcely je v našich podmínkách 800 m². Pokud jsou ohniska znečištění plošně omezeného rozsahu a pokud bychom použili průměru pro celou lokalitu, mohlo by to vést k závažnému podhodnocení zdravotního rizika pro obyvatele některé z parcel, na které bude výskyt ohnisek znečištění vázán. **Průměrovací plocha** musí mít v tomto případě velikost srovnatelnou s velikostí typické stavební parcely. Hustota vzorkovací sítě musí být taková, aby existovala jen zanedbatelná pravděpodobnost nepodchycení ohniska takové velikosti, která by mohla vést k významnému podhodnocení rizika pro některou z parcel.

Abychom měli jakous takous jistotu, že při průzkumu nepomineme ohnisko znečištění o velikosti srovnatelné s průměrnou stavební parcelou, měli bychom zvolit hustotu vzorkovací sítě nejméně cca 16 vzorků na hektar (krok sítě 25 m). I při této hustotě však existuje značné riziko chybné interpretace výsledků, neboť můžeme zachytit jen okrajové části takového ohniska, s jen mírně vyššími koncentracemi oproti pozadí. Žádoucí se tedy jeví zahuštění sítě na dejme tomu 36 vzorků na hektar (odpovídá kroku sítě 16,6 m).

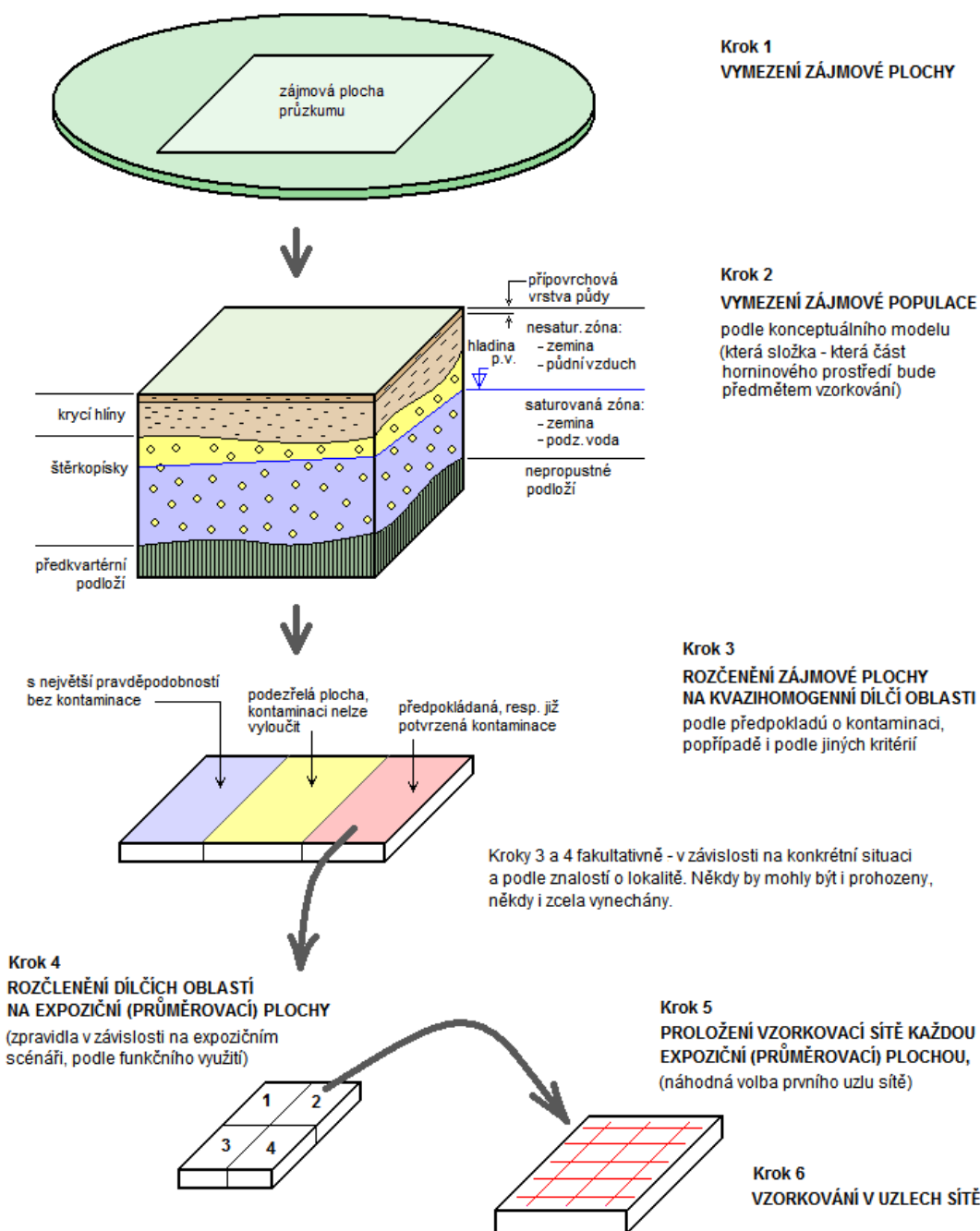
To neznamená, že u velmi rozsáhlých lokalit musí být vzorkovací síť tak vysoké hustoty, aby byla reprezentativně podchycená každá průměrovací plocha, kterých se může vejít do celé zkoumané lokality větší množství. Vhodné může být rozdělení terénních prací na etapy. Pokud nám úvodní etapa vzorkování poskytne základní představu o výskytu znečištění v ploše lokality, mohou být tyto informace vodítkem k tomu, abychom v následné etapě podrobného průzkumu rozhodli o rozdílné hustotě vzorkovací sítě různých částech lokality.

Metodická příručka Hodnocení průzkumu a sanací

Jindy zase může být naopak hustota sítě v celé ploše lokality menší díky znalosti o tom, že plošná variabilita kontaminace je nízká.

Ilustrace postupu při strategii stratifikovaného systematického vzorkování se zohledněním budoucího využívání lokality pro průměrování výsledků uvádí následující obrázek č. 21.

Obr. 21. Schéma postupu při systematickém vzorkování a zohlednění velikosti průměrovací plochy



SCHEMA POSTUPU PŘI SYSTEMATICKÉM VZORKOVÁNÍ

4.7. Vzorkování zemin ve vertikálním profilu

Vzorkování stratifikovaného profilu: v zásadě stratifikované vzorkování ve vertikálním směru, kde hranicemi dílčích celků (vrstev) jsou hranice genetické.

Vzorkování v rámci jedné uniformní vrstvy - lze volit různé přístupy, podle účelu:

- Vzorkování z náhodné hloubky: lze použít, když vrstva nemá žádnou viditelnou stratifikaci a kde potřebujeme vyhodnotit varianci v rámci celého jejího profilu. Hloubka odběru každého vzorku se určí předem výpočtem ze vztahu: $H = M \cdot RN$, kde M je mocnost vrstvy, RN náhodné číslo (interval 0 až 1).
- Stratifikované vzorkování z náhodné hloubky: podobně jako výše, ale v rámci dílčích vrstev profilu, definovaných geneticky nebo smluvně (např. z prvního metru od terénu nebo z nesaturované zóny a podobně).
- Vzorkování z definované diskrétní hloubky: např. 0,5 m pod terénem, 2 m pod terénem nebo jeden vzorek na každý metr hloubky, atp. V praxi často musíme řešit dilema v případech, kdy definovaná vzorkovací úroveň interferuje s hloubkami rozhraní vrstev. Rozhodování podle účelu vzorkování.
- Vzorkování metodou bodové brázdy – ze stanovených intervalů se odebírají vzorky, reprezentující celý interval (jde vlastně o směsný vzorek z celého intervalu).

Hloubky vzorkování závisí na účelu:

Někdy potřebujeme znát hloubkový vývoj kontaminace s hloubkou. Např. při hodnocení migrace kontaminace do podzemních vod potřebujeme znát distribuci kontaminace přes celou průchozí zónu nad hladinou podzemní vody. V tomto případě se za nejvhodnější považují diskrétní odběry po 0,5 až 1 m, nejsou-li specifické důvody pro intervaly jiné.

Pro rizika z kontaktu s kontaminací a z neúmyslné ingesce se za rozhodující považuje hloubka cca 15 cm (EPA 1994). Ve výhledu je však nutno při hodnocení rizik počítat se situací, že například při výkopových pracích při zakládání staveb či při rekonstrukcích/výstavbě podzemních sítí se mohou dostávat k povrchu i zeminy z větších hloubek. Proto se za zájmový interval pro hodnocení rizik z přímého kontaktu s kontaminací a z neúmyslné ingesce zemin uvažuje hloubka kolem 1,2 m, potenciálně až 3,5 m (MPCA, 1996).

Pro rizika z konzumace zeleniny pěstované na kontaminované půdě se považuje za důležitou hloubka do 50 cm pod terénem.

Kompozitní intervalové vzorkování: z určitého hloubkového intervalu (například 1 vzorek z každé vrstvy nebo směsný vzorek z každého metru atp.) se připraví jeden vzorek k analýze. Pro většinu aplikací je nepřijatelné připravovat kompozitní vzorek z intervalu, který by šel přes hranice vrstev profilu.

4.8. Ohnisko kontaminace

Z hlediska použití statistických přístupů pro hodnocení kontaminace lze ohnisko kontaminace definovat takto (DEQ, 2002) :

Dvě či více vzorkovacích míst která spolu sousedí na přiměřeně malých vzdálenostech a která mají významně vyšší koncentrace škodliviny oproti svému okolí, takže je to indikací že:

- jde o odlišnou statistickou populaci,
- jde o potenciální riziko, které by neměla statistická analýza dat zamaskovat.

Pokud existuje jen jeden údaj o zvýšené kontaminaci, měl by být posuzován jako indikace ohniska, které je žádoucí ověřit doplňujícím vzorkováním.

Statistické hodnocení dat z ohnisek kontaminace v jednom souboru společně s ostatními daty z lokality je zásadně nekorektní (možnost postupu v takovém případě viz konec kapitoly 3.5.).

4.9. Pozad'ové hodnoty

Pozad'ové hodnoty sledovaných kontaminantů jsou velmi důležitým parametrem při hodnocení existence a vlivu kontaminace.

Někdy se hodnoty z pozadí mohou uplatňovat i jako sanační limity.

K výběru vzorkovacího areálu pro ověření pozad'ových hodnot se používá odborný úsudek a postupuje se podle účelu (vzorky proti směru proudění podzemních vod, proti směru převládajících větrů, proti sklonu terénu...).

Rozlišujeme přírodní pozadí a pozadí se zvýšenými koncentracemi v důsledku vlivů dalších – jiných zdrojů kontaminace.

Obecně se pro charakteristiku pozadí nepoužívají statistické přístupy, většinou není k dispozici dostatečný rozsah dat. Data z lokality se porovnávají s pozadím individuálně.

Když se však již statistické metody pro porovnávání s pozadím aplikují, pak je třeba mít na zřeteli, že malý počet pozad'ových vzorků výrazně snižuje míru jistoty vyslovovaných závěrů o tom, zda lokalita vykazuje či nevykazuje zvýšené koncentrace škodliviny oproti pozadí. Používají se zde párové t-testy popřípadě Wilcoxon-Rank-Sum test (následovaný kvantilovým testem v případech, když se v datových souborech vyskytuje větší počet hodnot pod mezí detekce).

Při ověřování pozad'ových hodnot podzemních vod je zvláště nutno dbát na to, aby data souboru pro výpočet pozad'ových charakteristik byla opravdu nezávislá – musí být získána statisticky náhodným způsobem vzorkování a nesmí vykazovat trend.

Porovnávání kontaminace zemin předmětné lokality s pozadím:

Je-li pozadím jediné číslo (například regionální pozadí), pak se data z areálu lokality porovnávají s pozadím individuálně (každá hodnota zvlášť) nebo se s pozadím srovnává horní mez konfidenčního limitu (UCL) pro průměr.

Máme-li pro pozadí k dispozici soubor statisticky vyhodnotitelných dat, pak se jednotlivá data z lokality porovnávají individuálně s průměrem pozadí nebo se soubor dat z lokality porovnává s pozad'ovým souborem s využitím párových t-testů (Studentův párový test) resp. neparametrických testů (Wilcoxon Rank_sum test v kombinaci s kvantilovým testem či Chebychevův test).

Minimální počet vzorků pro dostatečnou průkaznost testů, pokud hodnotíme proti pozadí úspěšnost výsledků sanace: 10 z plochy sanované lokality, 10 z pozadí.

Porovnávání kontaminace podzemních vod s pozadím:

Rozeznáváme dvě možné situace:

- přírodní pozadí, nebo regionální kontaminace – bez trendů, bez oscilací

Metodická příručka
Hodnocení průzkumu a sanací

- pozadí ovlivněné lokálním znečištěním ze zdroje proti směru proudění od hodnocené lokality – koncentrace mohou vykazovat trendy a sezónní oscilace.

Přírodní pozadí:

- S jedinou číselnou hodnotou (limitem) z pozadí se srovnává dolní mez konfidenčního intervalu LCL pro průměr dat z areálu lokality. Použít lze i toleranční limit či predikční limit.
- Máme-li i z pozadí soubor dat, pak se použije párový t-test, popřípadě některý neparametrický test (např. Wilcoxon Rank_sum test v kombinaci s quantilovým testem).
- Minimální počet dat pro dostatečnou průkaznost testů, pokud hodnotíme proti pozadí úspěšnost výsledků sanace: 12 vzorků z pozadových vrtů a 12 vzorků z vrtů v areálu hodnocené lokality, vše odebráno při jedné vzorkovací kampani.

Pozadí ovlivněné lokálním znečištěním:

- průměrná koncentrace a flux (hmotový příron) kontaminace na výstupním profilu ze sledované lokality nesmí být vyšší než koncentrace a flux na vstupu podzemních vod do areálu lokality.
- Z minimálního počtu 8 vzorků odebraných na každém pozadovém vrtu v 8 po sobě následujících čtvrtletích se vypočte neparametrický horní toleranční limit. S tím se pak porovnává individuálně každý vzorek z areálu lokality. Jeho překročení je indikací kontaminace. Překročení UTL je podnětem k opětovnému převzorkování předmětného vrtu a výsledek tohoto převzorkování se porovnává s horní mezí predikčního limitu (UPL). Takové dvoufázové vzorkování může být účinným nástrojem

4.10. Zajištění jakosti vzorkovacích prací

QA/QC (Quality Assurance, Quality Control) je nezbytnou podmínkou zajištění reprezentativnosti vzorkovacích a analytických prací a tím i reprezentativnosti celého průzkumu a správnosti hodnocení a vyvozovaných závěrů.

Kritérium QA/QC: Když je nejistota měřené veličiny 1/3 a méně přípustné tolerance pro její použití, může být měřená veličina považována pro tento účel za bezchybnou.

Následující tabulka č. 5 uvádí základní typy kontroly vzorkovacích a analytických procedur:

Tab. 5.	analyzované sloučeniny	doporučená frekvence kontroly
Kontrola vzorkování:		
TRIP BLANK- kontrola ovlivnění vzorku při transportu	těkavé organické látky, NEL, nestabilní látky	s každou dávkou vzorků do laboratoře
STORAGE BLANK - kontrola ovlivnění vzorku během uskladnění	těkavé organické látky, nestabilní látky	1 krát týdně
FIELD BLANK - kontrola vlivu prostředí v místě vzorkování na vzorek (hlavně z ovzduší)	těkavé organické látky	5% z celkového počtu měření
EQUIPMENT BLANK	všechna prováděná stanovení	5% z celkového počtu měření
DUPLICATE FIELD - kontrola kvality vzorkování	všechna prováděná stanovení	5% z celkového počtu měření
Kontrola práce laboratoře:		
METHOD BLANK - analýzy slepých vzorků	všechna prováděná stanovení	denně

Metodická příručka Hodnocení průzkumu a sanací

DUPLICATE LABORATORY - kontrola laboratorních stanovení analýzou dělených a duplicitních vzorků	všechna prováděná stanovení	5% z celkového počtu měření
Ověření správnosti na připravených vzorcích: - FORTIFIED MATRIX, FORTIFIED BLANK	všechna prováděná stanovení	5% z celkového počtu měření
Ověření správnosti analýzou duplicitních či dělených vzorků nezávislou laboratoří	všechna prováděná stanovení	5% z celkového počtu měření

Realizace úplného rejstříku kontrol by byla mnohdy zbytečná. Někdy k nim neopravňuje celkový rozsah a význam průzkumu, některé kontroly lze vynechat podle okolností (např. při použití jednorázových vzorkovačů, nebo podle sledovaných pollutantů..., některé kontroly by byly duplicitní se systémem kontroly kvality samotné laboratoře a pod.).

4.11. Vzorkovací plán

Pro každou lokalitu musíme sestavit vzorkovací plán. Východiskem je konceptuální model lokality a konkrétní specifikace cílů průzkumu.

Vzorkovací plán jako součást projektu průzkumu kontaminace musí obsahovat:

- rozhodnutí, zda celá lokalita bude vzorkována jako celek nebo bude rozdělena do dílčích ploch, vzorkovaných a hodnocených samostatně, odůvodnění postupu
- rozhodnutí zda půjde o jednu vzorkovací kampaň nebo o více etap v závislosti na získaných výsledcích popřípadě jiných faktorech
- horizontální distribuce vzorkování (počet vrtů, hustota a tvar vzorkovací sítě),
- vertikální distribuce vzorkování (které vrstvy, v jakých hloubkách)
- typ vzorků (směsný/bodový/intervalový)
- u vzorkování vod frekvence vzorkování
- analytické parametry (co bude analyzováno)
- metodika samotného vzorkování
- metodika nakládání se vzorky
- požadavky na analytické metody a jejich přesnost
- zajištění kontroly kvality
- celkový sumář vzorkování a analýz.

5. HUSTOTA VZORKOVÁNÍ, POČTY VZORKŮ

5.1. Minimální počet vzorků pro statistické hodnocení dat

Minimální počet dat souboru pro statistické hodnocení: 9. V zásadě je to minimální počet dat, který nám zajišťuje jakž takž reprezentativní posouzení charakteru rozdělení souboru jako základního předpokladu pro korektní statistické hodnocení.

Jinak je pro stanovení počtu vzorků pro statistické hodnocení rozhodující požadovaná reprezentativnost výsledků. Obecně platí, že nejistoty ve stanovení charakteristik statistických souborů klesají s rostoucím počtem vzorků. Existují vzorce pro odvození minimálního počtu

Metodická příručka Hodnocení průzkumu a sanací

vzorků v závislosti na akceptovatelné úrovni nejistoty. Definování těchto vzorců uspokojí statistika, avšak není řešením pro praxi. Do všech takových výpočtů vstupují totiž některé statistické parametry populací které hodláme vzorkováním teprve studovat a které musíme tudíž jen odhadovat. V odborné literatuře lze nalézt i empirická pravidla, použitelná pro průzkum kontaminace horninového prostředí. Níže uvedená tabulka uvádí jedno z nich:

Tab.6.	Směrné hodnoty rozestupu sítě při systematickém statistickém vzorkování		
	Zdroj: DEQ, 2002 (lit. 5), též Sarz M., 2003 (lit. 43).		
rozměr plochy	malá plocha	střední plocha	velká plocha
	do 1000 m ²	1000 až 12000 m ²	nad 12000 m ²
rozestup sítě (m)	$R = \frac{\sqrt{A/\pi}}{2}$	$R = \frac{\sqrt{A/\pi}}{4}$	$R = \sqrt{\frac{A\pi}{SF}}$
Ve vzorcích: R = rozestup sítě, A = rozměr plochy (m ²), SF = bezrozměrný faktor roven délce delší strany plochy v m.			
orientačně: vzorci odpovídající interval rozestupu sítě	0 až 9 m	5 až 15 m	nad 9 m
celkový počet vzorků	$n = A/R^2$		

Opravdu reprezentativní podklady pro statistické hodnocení s dostatečnou úrovní jistoty zajišťuje jen taková hustota vzorků, jaká bývá v praxi z ekonomických důvodů jen stěží dosažitelná.

5.1.1. Průzkum pro charakteristiku lokality

Průzkum s tímto cílem bývá první průzkumnou etapou, při které se zpravidla statistické metody nepoužívají. Základní metodou vzorkování je vzorkování na základě odborného úsudku. Pokud se použije statistické vzorkování, pak půjde nejspíše o systematické vzorkování v pravidelné síti, přičemž hustota sítě bude stanovena v závislosti na spodní hranici velikosti ohnisek, která chceme vzorkováním podchytit.

Data pro statistické vyhodnocení výsledků musí pocházet z jedné populace, v zásadě nelze směšovat do jednoho souboru data z ohnisek znečištění se vzorky z neznečištěných částí lokality. Výsledky by byly matematicky nekorektní a vedlo by to k „naředování“ kontaminace.

5.1.2. Verifikace sanace

Pro hodnocení dosažení cílů sanace (tj. porovnávání zbytkové kontaminace se sanačními limity) se všeobecně doporučuje aplikace statistických metod pro případ lokalit středního až velkého rozsahu. Jako střední rozsah se bere velikost lokalit v rozmezí 1000 až 12 000 m² (DEQ, 2002 - lit. 5).

Výkopy

Vzorkování stěn a dna výkopů. Podle (DEQ, 2002 - lit. 5) se pro stanovení počtu vzorků se doporučuje použít vzorce pro systematické vzorkování v síti podle tabulky č. 6 v kapitole 5.1, přičemž výměra plochy pro výpočet počtu vzorků zahrnuje nejen dno ale i stěny výkopu. U malých jam musí být výsledek výpočtu případně upraven tak, aby byl odebrán nejméně 1 vzorek z každé stěny.

Připouští se dvofázový přístup k systematickému vzorkování v síti: v první etapě se odeberou vzorky ze všech bodů sítě, ale laboratorní analýzy necháme provést jen z 25%

vzorků, nejméně však 9. Výběr vzorků pro analýzy musí být proveden náhodným postupem. Teprve když by výsledek nevyhovoval, provedly by se laboratorní analýzy i z ostatních vzorků. Doporučuje se, aby už v prvním kole byly odebrány všechny vzorky a ty neanalyzované aby byly řádně uschovány, aby se nemuselo po druhé již znovu do terénu.

Pokud se prošetřovaná plocha rozděluje do více dílčích ploch pro stratifikované vzorkování, musí být minimální počet vzorků z každé dílčí plochy opět 9.

Vzorkování hromad zemin při ex-situ dekontaminaci

Má-li se použít statistický přístup, lze aplikovat postup, vycházející z postupu popsaného pro vzorkování výkopů DEQ, 2002 (lit. 5). Hromadou se proloží vzorkovací síť podle stejných zásad (podle tabulky 6 v kapitole 5.1), přidat se musí vertikální komponenta (odběr vzorků zevnitř vertikálního profilu).

Je-li hromada heterogenní, pak se rozestup sítě podřizuje tomu, jaký má být rozměr ohniska zvýšené kontaminace, které by nám nemělo uniknout.

Vzorkování zemin při sanaci in situ

Má-li se použít statistický přístup, lze aplikovat postup, vycházející z postupu popsaného pro vzorkování výkopů DEQ, 2002 (lit. 5). Sanovanou plochou se proloží vzorkovací síť podle stejných zásad - podle tabulky 6 v kapitole 5.1, přidat se musí prostorová komponenta (odběr vzorků z vertikálního profilu).

Odebrán musí být nejméně 1 vzorek z každého 1,5 m úseku vertikálního profilu sanovaného objemu až k jeho bázi. Každý takový interval 1,5 m se rozdělí na 10 stejných dílčích úseků a ke vzorkování se vybere jeden z nich náhodným postupem.

Počet vzorků pro porovnání výsledků sanace s pozadím

Minimální počet vzorků z plochy sanované lokality: 10 z pozadí a 10 ze sanované lokality.

5.1.3. Hodnocení výsledků průzkumu - porovnávání koncentrace s kritérii (limity) přípustné kontaminace

V principu se postupuje podle zásad pro verifikaci sanace. Na zřeteli však musíme mít, že pro některé expoziční scénáře (pro některé typy kontaminace a některé druhy funkčního využívání území) musí datový soubor pro hodnocení odpovídat tzv. expoziční – průměrovací ploše. Soubor dat pro porovnávání s limitem přípustné kontaminace musí pocházet z jedné takové expoziční jednotky.

Bylo řečeno, že statistické metody jsou pro hodnocení výsledků úsudkového vzorkování zásadně nepoužitelné, avšak při více-etapovém vzorkování je často možné, že část dat nebo i všechna data z předchozího úsudkového vzorkování mohou být do statistických souborů zahrnuta.

Principy takového použití dat z předchozího úsudkového vzorkování pro statistickou analýzu a porovnávání s limity přípustné kontaminace:

- Všechna data by měla být získána stejnou metodou, se stejnou mezí detekce.
- Rozděl každou z expozičních – průměrovacích ploch (expozičních jednotek) pravidelnou čtvercovou sítí na nejméně 9 stejně velkých buněk.
- V každé buňce by měla být nejméně jedna hodnota.

- Posud' pro každou buňku zvlášť, jak data z předchozí etapy úsudkového vzorkování odpovídají kritériím na statisticky nezávislá náhodná data: je-li v jedné buňce více vzorkovacích bodů, pak je žádoucí přibližně stejný řád úrovně kontaminace - nesmí dojít ke směšování dat z ohnisek s pozad'ovými daty. Data z ohnisek nesmí být do souboru zahrnuta a jejich plochy nutno hodnotit zvlášť. Jsou-li data příliš nahlučena, pak je na zvážení, zda je nenahradit jednou hodnotou, aby takový shluk neměl nepřiměřeně velkou váhu.
- Bude stačit, když nové vzorky odebereme z buněk v nichž nemáme žádnou hodnotu nebo jen hodnotu, kterou z nějakých důvodů nemůžeme převzít. Nové vzorky pak budou tvořit jeden společný soubor s výběrem dat z předchozího vzorkování.

5.2. Minimální počet vzorků při úsudkovém vzorkování

5.2.1. Úsudkové vzorkování při hodnocení cílů sanace

Místa odběru vzorků se při úsudkovém vzorkování řídí odborným úsudkem. Přednostně se vzorkování orientuje do míst s největším podezřením na zbytkovou kontaminaci nad úrovní sanačního limitu. Zpráva musí podrobně dokumentovat lokalizaci každého odběru s odůvodněním a vyhodnocením ve vztahu k sanačním limitům.

Při hodnocení vzorkování na základě odborného úsudku se srovnává se sanačním limitem každý vzorek individuálně.

Pokud jsou všechny vzorky pod sanačním limitem, sanace byla úspěšná. Jestliže některý ze vzorků vykazuje koncentrace nad úrovní sanačního limitu, může ústít vyhodnocení na základě odborného úsudku do některého těchto závěrů:

- sanaci lze považovat za úspěšnou, překročení limitu u některého ze vzorků bylo vyhodnoceno jako nesignifikantní (konstatování že překročení sanačního limitu je nevýrazné nebo ojedinělé a nadlimitní hodnota se vztahuje k celkově nevýznamnému množství zbytkového znečištění),
- je nutno doplňkové vzorkování (lokální zahuštění vzorků) pro přesnější vymezení míst s nadlimitní zbytkovou kontaminací a pro bližší posouzení situace,
- sanovanou lokalitu je nutno převzorkovat s použitím strategie statistického vzorkování,
- nebyly naplněny cíle sanace.

Úsudkové vzorkování se doporučuje jako vhodné pro malé lokality (do cca 1000 m²).

5.2.2. Úsudkové vzorkování pro výpočty rizika a/nebo pro porovnávání koncentrace s kritérii (limity) přípustné kontaminace

V principu se postupuje podle zásad pro verifikaci sanace. Na zřeteli však musíme mít, že pro některé expoziční scénáře (pro některé typy kontaminace a některé druhy funkčního využívání území) musí datový soubor pro hodnocení odpovídat tzv. expoziční – průměrovací ploše. Soubor dat pro porovnávání s limitem přípustné kontaminace musí pocházet z jedné takové expoziční jednotky.

5.3. Předpisové – smluvní - požadavky na minimální počty vzorků

Hustotu vzorkování, minimální počty vzorků a často i výběr metod statistického zpracování dat z průzkumů a monitoringu kontaminace upravují v řadě zemí oficiální předpisy či metodické pokyny.

V další části textu této příručky jsou uvedeny příklady takových návodů. Lze mít za to, že i bez nároku na závaznost mohou sloužit jako dobrá orientační pomůcka pro praxi. Příklady jsou výběrem z internetové rešerše, přičemž přednostně byly vybírány předpisy ze států, kde je metodickým aspektům průzkumů kontaminace věnována dlouhodobá a systematická pozornost. Lze odůvodněně předpokládat, že tyto návody jsou podloženy i teoretickými analýzami a představují kompromis z hlediska požadavků na reprezentativnost vzorkování. Výběr si neklade nároky na úplnost.

Minimální plošná hustota vzorkovací sítě při předběžném průzkumu lokality

Předpis MPCA, 1996 (Minnesota, lit. 32) specifikuje požadavek na minimální laterální hustotu sítě pro vzorkování zemin při předběžném průzkumu kontaminace za situace, kdy nejsou k dispozici žádné informace o možné lokalizaci ohnisek znečištění a na lokalitě se nevyskytují žádné zjevné indicie nečištění. Požadavky shrnuje následující tabulka 7:

Tab. 7	Předběžný průzkum – minimální laterální hustota pro vzorkování zemin bez předpokladů o lokalizaci ohnisek znečištění a bez zjevných známek znečištění Zdroj: MPCA, 1996 (lit. 32)
plocha lokality (m ²)	hustota vzorků (počet/ha)
do 1 ha	30 (odpovídající rozestup bodů v síti cca 18 m)
1 až 2 ha	20 (odpovídající rozestup bodů v síti cca 22 m)
2 až 15 ha	10 (odpovídající rozestup bodů v síti cca 31 m)
nad 15 ha	7 (odpovídající rozestup bodů v síti cca 38 m)
Předpokládaná místa ohnisek znečištění jsou vzorkována nad rámec základní sítě podle této tabulky a hodnocena zvlášť.	

Metodické doporučení USGS, 2003, specifikuje pro menší lokality - rozlohy do 10 ha - minimální počet 16 vzorkovacích bodů na hektar (což odpovídá hustotě sítě 25 m).

Hodnocení úspěšnosti sanace nebo přípustnosti určitého funkčního využívání lokality porovnáváním s limity, odvozenými ve vztahu k ochraně zdraví či ekosystémů

Metodický předpis PDEP, 2008 (stát Pennsylvania, lit. 42) předepisuje pro porovnávání použít UCL 95 pro průměr, avšak alternativně se připouští pravidlo „75%/10x“. Jde o ad-hoc statistické pravidlo tohoto znění:

Skutečná mediánová koncentrace škodliviny je pod sanačním limitem založeným na analýze rizika, pokud je 75% vzorků pod sanačním limitem a žádný ze vzorků nepřesahuje tento limit více než desetinásobně. Používá se tehdy, pokud sanační limit není odvozen pro konkrétní lokalitu (site-specific limit), ale jde o obecný limit.

Počet vzorků pro hodnocení podle tohoto pravidla:

- pro objemy zemin pod 100 m³ – 8 vzorků
- pro objemy zemin do 2000 m³ – nejméně 12 vzorků

Metodická příručka
Hodnocení průzkumu a sanací

- pro každých dalších 2000 m³ – nejméně dalších 12 vzorků.

Minimální počet vzorků 8 až 12 je odvozen ze simulační studie s použitím lognormální distribuce.

Pravidlo 75%/10x je inspirující v tom, že výrazně zjednodušuje statistické vyhodnocovací postupy a přitom respektuje jejich principy, je podložené. Lákavost jeho používání je dále v tom, že definuje jednoznačná a lehce srozumitelná pravidla přípustnosti nadlimitních koncentrací a to jak z hlediska jejich výše, tak četnosti.

V předpise se výslovně uvádí, že toto pravidlo se nehodí k posuzování dosažení cílů sanace, pokud jsou sanační limity odvozeny specificky pro konkrétní lokalitu. Takové limity totiž zahrnují předpoklad, že dlouhodobá expozice se vztahuje k aritmetickému průměru koncentrací v ploše předmětné lokality.

Předpis PDEP, 2008 zavádí podobné pravidlo 75%/2x i pro hodnocení kontaminace podzemních vod (viz kapitola 5.4.3).

Porovnávání kontaminace zemin nesaturované zóny s limitem (kritériem) pro zeminy z hlediska přestupu do podzemních vod

Předpis CDEQ, 1996 (Connecticut, lit. 3) vymezuje úroveň přípustné zbytkové kontaminace zemin nesaturované zóny z hlediska přestupu do podzemních vod:

Kontaminace (zbytková kontaminace) zemin je akceptovatelná, pokud oblast kontaminovaných zemin nesaturované zóny byla vymezena odběrem nejméně 20 vhodně lokalizovaných vzorků a

- a) koncentrace všech vzorků je na či pod úrovni limitu (kritéria), nebo
- b) pokud UCL 95 průměru všech vzorků je na či pod úrovni kritéria přičemž v žádném ze vzorků nebylo kritérium překročeno více než dvojnásobně.

Při odběru méně než 20 vzorků nesmí kritérium překračovat žádný vzorek.

Předpis IEPA, 1997 (Illinois, lit. 29) specifikuje minimální počet vzorků pro hodnocení přípustné (zbytkové) kontaminace z hlediska přestupu znečištění ze zemin nesaturované zóny do podzemních vod:

- plošná hustota vzorkování: 1 vzorek na 2000 m²,
- vertikálně 1 vzorek na každých 0,6 m vertikálního profilu nesaturované zóny od hloubky 5 cm pod terénem až k hladině podzemních vod.

Porovnávání kontaminace zemin s limitem (kritériem) pro zeminy z hlediska přímého kontaktu či ingesce

Předpis CDEQ, 1996 (Connecticut, lit. 3) vymezuje úroveň přípustné kontaminace zemin přepovrchové zóny z hlediska rizik z přímého kontaktu či neúmyslné ingesce:

Kontaminace (zbytková kontaminace) lokality je akceptovatelná, pokud je

- a) ve všech vzorcích koncentrace na či pod úrovni kritéria, nebo
- b) pokud UCL 95 průměru všech vzorků je na či pod úrovni kritéria přičemž v žádném ze vzorků nebylo kritérium překročeno více než dvojnásobně.

Porovnávání kontaminace půdního vzduchu s kritérii kontaminace pro půdní vzduch z hlediska inhalace

Předpis CDEQ, 1996 (Connecticut, lit. 3) vymezuje úroveň přípustné kontaminace půdního vzduchu nesaturované zóny z hlediska akceptovatelného rizika z inhalace těkavých emisí:

Kontaminace půdního vzduchu je akceptovatelná, pokud je ve všech vzorkovacích objektech sanované plochy koncentrace

- na či pod úrovni kritéria nejméně ve čtyřech po sobě následujících vzorcích odebíraných kvartálně, nebo
- pokud UCL 95 průměru všech vrtů z takového vzorkování je na či pod úrovni kritéria, přičemž v žádném ze vzorků nebylo kritérium překročeno více než dvojnásobně.

Minimální počty vzorků pro hodnocení koncentrace ve dně a stěnách sanačních výkopů

Zásady podle směrnice státu Michigan DEQ, 2002 (lit. 5) uvádí následující tabulka č. 8. Totožnou tabulku uvádí i předpis státu Minnesota - MPCA, 1996 - lit. 32):

Tab. 8		Minimální počet vzorků pro hodnocení koncentrace v zemině dna a stěn sanačních výkopů (Zdroj: DEQ, 2002, dtto MPCA, 1996)							
Dno výkopu	plocha dna (m ²)	< 50	50 – 100	100 – 150	150 - 250	250 - 400	400 – 600	600-800	800-1000
	počet vzorků	2	3	4	5	6	7	8	9
Stěny výkopu	plocha stěn (m ²)	< 50	50 - 100	100 - 150	150 - 200	200 - 300	300 – 400	> 400	
	počet vzorků celkově	4	5	6	7	8	9	1 vzorek na každých 15 m délky stěny	

Krok pravidelné vzorkovací sítě při vzorkování dna jámy podle této tabulky 8 vychází 4 až 10 m. Zachování této hustoty doporučuje tentýž předpis pro plochy dna jámy rozměrů až 12 000 m².

Tabulka č. 8 platí pro vzorkování na základě odborného úsudku. Pro větší výkopy s plochou dna nad 1000 m² se doporučuje vzorkování podle zásad pro statistické vzorkování. I u malých výkopů je absolutní minimum pro stěnové vzorky alespoň 1 vzorek z každé stěny (tj. celkem 4).

Pro hlubší výkopy se přidává pro stěny výkopů vertikální komponenta – vzorkování podle výše uvedené tabulky č. 8 pro každých 1,5 m hloubky.

Předpisy MPCA, 2008 (Minnesota, lit. 33) by mohly být vodítkem pro malé lokality při kontrole sanace lokalit s ropným znečištěním. Požaduje se jeden vzorek na každých 7 m lineární délky stěn výkopu (nejméně však 1 vzorek z každé stěny) a 1 vzorek na každých 9 m² jeho dna.

Předpis PDEP, 2008 (stát Pennsylvania, lit. 42) pro posuzování dosažení cílů sanace lokality s kontaminací ropnými látkami za situace, kdy nebyl proveden podrobný průzkum lokality a sanace je založena na vizuálním hodnocení kontaminace či na různých screeningových metodách:

Minimálně požadovaný počet vzorků:

- pro objemy výkopku do 200 m³ se odebere 5 vzorků
- další vzorek se odebere pro každých 75 m³ navíc až do 750 m³.

Metodická příručka
Hodnocení průzkumu a sanací

Odběr vzorků k analýzám se provede z míst nejpodezřelejších na kontaminaci, žádný ze vzorků nesmí překročit sanační limit.

Pokud je zastižena podzemní voda bez vizuálních známek znečištění, odeberou se navíc 2 vzorky zeminy těsně nad hladinou, nesmí vykazovat znečištění nad úroveň saturované koncentrace. Mimo to se odeberou 2 vzorky podzemní vody, které nesmí překročit pozadí.

Předpis MPCA, 2008 (stát Minnesota, lit. 33), pro vzorkování výkopů při sanaci zemin kontaminovaných ropnými látkami uvádí tabulka 9:

Tab. 9	Vzorkování výkopů při sanacích zemin kontaminovaných ropnými látkami Zdroj: MPCA, 2008
Stěny výkopů	1 vzorek na každých 7 m délky stěny, nejméně však 4 vzorky (z každé stěny alespoň jeden)
dno výkopu	1 vzorek na každých 9 m ² dna výkopu.
Před odběrem vzorku odstraň v místě odběru nejprve 30 cm zeminy ze stěny resp. dna výkopu.	

Minimální počty vzorků pro charakteristiku zemin na hromadách

Zásady podle směrnice DEQ, 2002 (Michigan, lit. 5) uvádí tabulka 10:

Tab. 10	Počty vzorků pro charakteristiku ex-situ hromad zemin, resp. odpadů Zdroj: DEQ, 2002					
objem hromady (m³)	< 20	20 – 75	75 – 400	400 – 750	750 – 1500	> 1500
počet vzorků	3 – 4	6 – 8	8 – 10	10 – 12	13 – 15	15 + 3 pro každých dalších m³

Tabulka 10 platí pro vzorkování na základě odborného úsudku. Každý vzorek se srovnává se sanačním (či jiným) limitem individuálně.

Předpisy státu Minnesota (MPCA, 2008, lit. 33) pro kontrolu dosažení sanačního limitu u hromad ex-situ dekontaminovaných zemin znečištěných ropnými látkami připouštějí podstatně menší počty vzorků než předpis předchozí. Viz následující tabulka 11. Totožnou tabulku prezentuje též předpis NDDH, 2009 (stát Severní Dakota, lit. 34):

Tab. 11	Počet vzorků pro kontrolu dosažení sanačního limitu u hromad ex-situ dekontaminovaných zemin tnečištěných ropnými látkami					
	Zdroj: MEDQ, 2008, dtto NDDH, 2009					
objem hromady (m³)	do 40 m³	40 - 400 m³	400 - 800 m³	800 – 1500 m³	1500 – 3000 m³	každých 1500 m³ nad 3000 m³
počet vzorků	1	2	3	4	5	1 další

Předpis MPCA, 1996 (stát Minnesota, lit. 32) uvádí tyto směrné hodnoty pro počet vzorků při prokazování dosažení sanačních limitů u hromady ex situ dekontaminovaných zemin obecně:

Tab. 12	Počty vzorků pro pro kontrolu dosažení sanačního limitu u hromad ex-situ dekontaminovaných zemin Zdroj: MPCA, 1996			
	objem hromady (m³)	do 400 m³	400 - 800 m³	nad 800 m³
	počet vzorků	1 vzorek na 75 m³	1 vzorek na 200 m³	1 vzorek na 400 m³
Při odběru méně než 10 vzorků, vzorky se nevyhodnocují statisticky ale každý výsledek se porovnává s limitem individuálně.				

Lokalizace a počty vzorků při sanaci zemin in situ

Podle DEQ 2002 (Michigan, lit. 5) se při použití úsudkového vzorkování pro hodnocení dosažení sanačních limitů při in-situ sanačních metodách postupuje v principu podle tabulky č. 8 a zásad pro vzorkování dna a stěn sanačních výkopů.

K tomu však musíme přidat prostorovou komponentu pro ověření sanace ve vnitřním prostoru (vertikálním profilu) sanovaného objemu, přičemž plošná hustota se vzorků volí jako pro dno jámy dle tabulky 8. Opět, vzorkování se orientuje přednostně do míst ve kterých lze očekávat vyšší zbytkovou kontaminaci. Takovými místy mohou být například centra původních ohnisek před sanací nebo místa, která jsou z různých důvodů (zóny nízké propustnosti, oblasti proudového stínu, středy rozestupů mezi sanačními objekty) hůře přístupná pro promývací médium při sanačním čerpání, při ventingu, při zasakování nutrientů.

V každém případě se odebírá **nejméně 1 vzorek na každých 1,5 m vertikálního profilu** sanované vrstvy až k její bázi, resp. až ke spodní hranici výskytu znečištění. Pokud nejsou žádné indicie pro výběr místa k přednostnímu odběru vzorku, pak se každý vzorkovací úsek délky 1,5 m se rozdělí na 10 podúseků stejné délky a vzorek se odebere z jednoho z nich na základě náhodného výběru.

Předpis MPCA, 1996 (stát Minnesota lit. 32) doporučuje použít při verifikaci in situ sanace zemin plošnou hustotu vzorkování rovněž podle stejné tabulky č. 8 pro vzorkování dna a stěn sanačních výkopů. Pro vertikální komponentu vzorkování však považuje za dostatečné rozčlenění vertikálního profilu až k bázi sanované zóny na 3 m intervaly.

Pro netěkavé kontaminanty se připouští směsné vzorky.

U sanace promýváním nesaturované zóny nebo při ventingu se požaduje, aby vzorkovací body byly v ploše lokalizovány doprostřed mezi sanačními objekty, odběr diskrétních vzorků, opět z každých 3 m hloubkového profilu.

5.4. Monitoring podzemních vod

5.4.1. Druhy monitoringu

Detekční monitoring

Periodické porovnávání koncentrací v monitorovacích vrtech na výstupu proudění podzemních vod z předmětné lokality s požadovými koncentracemi ve vrtech proti směru proudění k lokalitě nebo s limitem přípustné kontaminace.

Zpravidla se požaduje nejméně 1 vrt proti směru proudění od lokality a nejméně 3 vrty na výstupu proudění podzemních vod z předmětné lokality.

Kontrolní monitoring

Periodické vzorkování podzemních vod na výstupu jejich proudění z předmětné lokality za účelem kontroly, zda kontaminace nepřekračuje úředně stanovený přípustný limit.

V praxi se vyskytují dva typy limitů přípustné kontaminace v areálu lokality resp. na výstupu podzemních vod z ní:

- odvozené zpětným výpočtem z limitní hodnoty přípustného rizika,
- vycházející z požadových hodnot.

Sanační monitoring

Sledování koncentrací a jejich trendů v průběhu sanace podzemních vod.

Postsanační monitoring

Porovnávání koncentrací v monitorovacích vrtech na výstupu proudění podzemních vod z předemné lokality se sanačními limity. Účelem je dokladování splnění sanačních limitů.

Pozad'ový monitoring

Zpravidla poměrně krátkodobá aktivita, jejímž cílem je reprezentativní stanovení charakteristik pozad'ové kontaminace podzemních vod (pozad'ové ve vztahu k zájmové lokalitě).

5.5. Statistické metody hodnocení monitoringu podzemních vod

Rozdělení statistických metod pro vyhodnocování monitoringu podzemních vod podle principu:

- porovnávací testy (t-testy, multi-srovnávací strategie typu ANOVA, jiné porovnávací testy pro neparametrické soubory),
- použití konfidenčních, tolerančních či predikčních intervalů
- analýzy časových řad.

Formulování nulové hypotézy H_0 pro jednotlivé druhy monitoringu:

- detekční monitoring: lokalita je čistá, pokud není prokázán opak.
- postsanační monitoring: lokalita je kontaminovaná, pokud není prokázán opak
- kontrolní monitoring – nulová hypotéza v závislosti na tom, zda jsou limity založeny na pozadí nebo byly odvozeny zpětným výpočtem z úrovně přípustného rizika.

5.5.1. Požadavky na počty vzorků a frekvence vzorkování podzemních vod

Obecné požadavky na počty a frekvenci vzorků při monitoringu podzemních vod

Stanovení minimálního počtu vzorkovacích bodů a frekvence vzorkování při použití statistických metod je funkcí charakteru a účelu aplikovaného testu a požadované úrovně spolehlivosti.

Pro stanovení počtu vzorkovacích bodů (monitorovacích vrtů) je zpravidla rozhodující požadavek na dostatečné poznání a podchycení prostorové variability koncentrací.

Jako minimální frekvence vzorkování podzemních vod se obvykle považuje frekvence kvartální, tj. 4krát do roka.

Testy pro interpretaci sezónních složek kolísání chemismu obecně vyžadují vzorkování v měsíčním intervalu po dobu několika let.

Směrnice Ohio EPA, 1995 (lit 38) při detekčním monitoringu připouští pololetní frekvenci vzorkování (Ohio EPA, 1995).

Podle NYSDEC, 2002 (stát New York, lit. 37) jsou minimální požadavky na monitoring podzemních vod po směru proudění od kontaminované lokality:

- monitorovací vrtý v ploše kontaminačního mraku: vzorkování pololetně,
- vzorkování podzemních vod po směru proudění od kontaminované lokality: kvartální vzorkování po dobu nejméně 2 let,
- pozadňové vrtý (proti směru proudění podzemní vody od lokality) – vzorkování ročně.

Minimální požadavky na hustotu vzorkování při statistickém hodnocení trendů kontaminace v podzemní vodě

Doporučení AFCEE, 2007 (lit. 1) na minimální počet dat pro statistické hodnocení trendů kontaminace v podzemní vodě v areálu hodnocené lokality:

- nejméně 4 vrtý,
- nejméně 6 vzorkovacích cyklů,
- z každého vrtu nejméně 4 vzorky odebrané v průběhu šesti vzorkovacích cyklů, v žádném z vrtů nesmí být vynechány více než 2 vzorkování po sobě.

Minimální požadavky na počet vzorků podzemních vod pro aplikaci různých statistických metod

Směrnice Ohio EPA, 1995 (lit. 38), uvádí tyto minimální počty vzorků pro zajištění alespoň minimální reprezentativnosti statistického hodnocení monitoringu podzemních vod:

- neparametrická ANOVA – minimálně 3 vrtý, z každého nejméně 3 vzorky.
- parametrická ANOVA : počet vzorků minus počet vrtů musí být větší či roven 10.
- konstrukce mezí tolerančních intervalů – nejméně 8 vzorků z každého vrtu (jinak vyjdou extrémně široké)
- predikční intervaly – neméně 8, lépe 10 vzorků,
- hodnocení časových řad – nejméně 8 vzorků z každého hodnoceného vrtu.

Porovnávání kontaminace podzemní vody s limity podle analýzy rizika

Předpis PDEP, 2008 (stát Pennsylvania, lit. 42) vymezuje tyto základní zásady pro porovnávání výsledků vzorkování podzemních vod s limity, odvozenými od úrovně akceptovatelného rizika pro zdraví či ekosystémy:

Pro sledování trendů slouží časové řady. Poté, co došlo k ustálení chemismu podzemních vod (vymizení trendů), doporučuje se porovnávat se sanačním limitem hodnoty UCL 95 pro průměr individuálně v každém z monitorovacích vrtů.

Jako alternativu k UCL 95 předpis připouští aplikaci ad hoc statistického pravidla **75%/10x**: skutečná mediánová koncentrace škodliviny v areálu lokality je pod sanačním limitem založeným na analýze rizika, pokud je 75% vzorků vod z vrtů v areálu lokality pod sanačním limitem a žádný ze vzorků nepřesahuje tento limit více než desetinásobně. Používá se tehdy, pokud sanační limit není odvozen pro konkrétní lokalitu (site-specific limit), ale jde o obecný legislativní limit.

Pro monitorovací vrtý při hranici sanované lokality po směru proudění podzemních vod lze v tomto případě uplatňovat ještě další ad hoc **pravidlo 75%/2x**: v každém z vrtů bude mít

nejméně 75% vzorků koncentraci pod limitem přičemž žádná z hodnot nepřekročí tento limit více než dvojnásobně. Minimální počet vzorků: 8 z každého vrtu odebraných v 8 po sobě následujících kvartálech (lit. 63 – Pennsylvania, 2008).

Při kratších vzorkovacích obdobích se musí použít zásada, že limit nesmí překročit žádný vzorek.

Pravidlo 75%/10x resp. 75%/2x je inspirující v tom, že výrazně zjednodušuje statistické vyhodnocovací postupy a přitom respektuje jejich principy, je podložené. Lákavost jeho používání je dále v tom, že definuje jednoznačná a lehce srozumitelná pravidla přípustnosti nadlimitních koncentrací a to jak z hlediska jejich výše, tak četnosti.

Předpis PDEP, 2008 zavádí podobné pravidlo 75%/10x i pro hodnocení kontaminace zemin (viz kapitola 5.3).

Porovnávání kontaminace podzemní vody vůči pozadí

Předpis PDEP, 2008 (stát Pennsylvania, lit. 42) vymezuje tyto základní zásady pro porovnávání výsledků vzorkování podzemních vod vůči pozadí:

Pro porovnávání koncentrací v areálu sledované lokality s přírodním pozadím je požadováno odebrání nejméně 12 pozadových vzorků z jakékoliv kombinace vrtů v území proti směru proudění od hodnocené lokality a současné odebrání stejného počtu vzorků z monitorovacích vrtů v prostoru lokality. Oba soubory se porovnají pomocí vhodného párového testu (t-testu) v případě parametrických souborů nebo pro neparametrické soubory s použitím Wilcoxon-Rank-Sum testu (popřípadě doplněného ještě kvantilovým testem když se v datových souborech vyskytuje větší počet hodnot pod mezí detekce).

Pokud je již pozadí antropogenně ovlivněno (již ze zázemí migruje kontaminace), pak lze:

- porovnávat průměrné koncentrace nebo flux znečištění na vstupním a výstupním profilu (do výpočtu zahrnujeme všechna data z vrtů na profilech), nebo
- stanovíme toleranční limity z údajů vrtů před vstupem podzemních vod na lokalitu (nejméně 8 vzorků z po sobě následujících kvartálech z každého vrtu) a porovnáváme s koncentracemi na lokalitě. Prevzorkování lokality v případě překročení tolerančního limitu.

Porovnávání kontaminace podzemních vod s kritérii pro ochranu podzemní vody

Předpis CDEQ, 1996 (Connecticut, lit. 3) vymezuje tyto požadavky na hodnocení monitoringu podzemních vod kontaminované lokality ve vztahu k limitu resp. kritériu, stanovenému s ohledem na ochranu podzemních vod:

Kontaminace (zbytková kontaminace) v kontaminačním mraku je akceptovatelná,

- a) pokud je ve všech vrtech koncentrace na či pod úrovni kritéria nejméně ve čtyřech po sobě následujících vzorcích odebíraných kvartálně, nebo
- b) pokud byly všechny monitorovací objekty vzorkovány měsíčně po dobu alespoň jednoho roku (12 vzorků z každého vrtu) a UCL 95 průměru všech vrtů je na či pod úrovni kritéria přičemž v žádném ze vzorků nebylo kritérium překročeno více než dvojnásobně.

Samozřejmým předpokladem je, že lokalizace vzorkovacích bodů podchycuje kontaminaci lokality reprezentativně.

Porovnávání kontaminace podzemních vod s kritérii pro ochranu povrchových vod

Předpis CDEQ, 1996 (Connecticut, lit. 3) vymezuje tyto požadavky na hodnocení monitoringu podzemních vod kontaminované lokality ve vztahu k limitu resp. kritériu, stanovenému s ohledem na ochranu povrchových vod:

Kontaminace (zbytková kontaminace) v kontaminačním mraku je akceptovatelná,

- a) pokud je průměr ze všech monitorovacích vrtů na či pod úrovni kritéria nejméně ve čtyřech po sobě následujících vzorcích odebíraných kvartálně, nebo
- b) pokud v části mraku těsně před jeho vstupem do povrchových vod je koncentrace na či pod úrovni kritéria za předpokladu, že kontaminace v mraku nevykazují v žádném směru vzestupný trend a mrak se nerozšiřuje.

Samozřejmým předpokladem je, že lokalizace vzorkovacích bodů podchycuje kontaminaci lokality reprezentativně.

Porovnávání obsahu těkavých složek v podzemní vodě s kritérii kontaminace podzemních vod z hlediska inhalace

Předpis CDEQ, 1996 (Connecticut, lit. 3) vymezuje tyto požadavky na hodnocení monitoringu podzemních vod kontaminované lokality ve vztahu ke kritériu, stanovenému s ohledem na prevenci rizik z inhalace těkavých škodlivin, uvolňovaných z podzemních vod:

Kontaminace (zbytková kontaminace) v kontaminačním mraku je akceptovatelná,

- a) pokud je ve všech vrtech koncentrace na či pod úrovni kritéria nejméně ve čtyřech po sobě následujících vzorcích odebíraných kvartálně, nebo
- b) pokud UCL 95 průměru všech vrtů z takového vzorkování je na či pod úrovni kritéria, přičemž v žádném ze vzorků nebylo kritérium překročeno více než dvojnásobně.

Základní zásady monitoringu přírodní atenuace

Velmi moderně, instruktivně a vyčerpávajícím způsobem shrnuje problematiku monitoringu a hodnocení přírodní atenuace publikace Carey et al., 2000 (lit. 2), vydaná pod patronací environmentální agentury EPA pro Anglii a Wales.

Za minimální dobu monitoringu přírodní atenuace se podle této publikace považuje doba dvou let. Často si praxe vyžádá dobu ještě delší, aby byl k dispozici dostatek dat pro spolehlivou interpretaci výsledků a identifikaci trendů.

Frekvence vzorkování bude záviset na chování sledovaného kontaminačního mraku, na rychlosti proudění podzemních vod a migrace škodlivin, na projevech sezónního kolísání chemismu a kontaminace. Kvartální frekvence vzorkování (4 vzorky ročně) se považuje za minimální pro rychlosti proudění podzemních vod řádu desítek m/rok a vyšších. Doporučuje se vyšší frekvence v prvním roce - v úvodní fázi monitoringu - pro potvrzení platnosti konceptuálního modelu a primární identifikaci sezónních a trendových složek časové řady.

Absolutně minimální počty monitorovacích vrtů:

- minimálně jeden pozadový vrt - nad lokalitou proti směru proudění podzemních vod,
- minimálně dva vrty v kontaminačním mraku po směru proudění podzemních vod od jeho centra (při jednom vrtu bychom neměli jistotu, že sledujeme centrální proudnici mraku),

Metodická příručka
Hodnocení průzkumu a sanací

- minimálně vždy jeden vrt mezi kontaminačním mrakem a každým potenciálním receptorem kontaminace (kontrolní, tzv. sentinelový vrt pro včasnou identifikaci ohrožení receptorů).

Pokud se chceme pokoušet o početní hodnocení základních parametrů atenuace, potřebujeme dva vrty na centrální proudnici, umístěné ve vhodné vzdálenosti po směru proudění vůči sobě.

Příslušný úřad bude zpravidla vyžadovat, aby nejméně jeden monitorovací vrt byl umístěn přímo při hranici kontaminované lokality, ve směru odtoku podzemních vod z ní.

6. VÝBĚR LITERATURY

- [1] AFCEE (Air Force Centre for Environmental Excellence), 2007: AFCEE Monitoring and Remediation Optimization System Software, Appendix A.2: Statistical Trend Analysis Methods.
- [2] Carey, M. A. et al., 2000: Guidance on the Assessment and Monitoring of Natural Attenuation of Contaminants in Groundwater, WRc R&D Publication 95, Environment Agency, Bristol, UK
- [3] CDEQ (State Regulation of the Connecticut Department of Environmental Protection), 1996: State Regulations Section 22a-133k-1 through 3 concerning the remediation of polluted soil, surface water and groundwater
- [4] CL:AIRE, and The Chartered Institute of Environmental Health (CIEH), 2008: *Guidance on Comparing Soil Contamination Data with a Critical Concentration*. London, ISBN10 1-904306-79-9, ISBN13 978-1-904306-79-5.
- [5] DEQ (Michigan Department on Environmental Quality), 2002: Sampling Strategies and Statistics Training Materials for Part 201 Cleanup Criteria. Michigan Department on Environmental Quality.
- [6] EPA, 1989a: *Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities, Interim Final Guidance*. Office of Solid Waste, Waste Management Division. US Environmental Protection Agency, Wahington, D.C.
- [7] EPA 1989b: *Methods for Evaluating the Attainment of Cleanup Standards. Volume 1: Soils and Solid Media*. Office of Policy, Planning, and Evaluation, US Environmental Protection Agency, Wahington, D.C. EPA/230/02-89-042
- [8] EPA, 1992a: *Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities, Addendum to Interim Final Guidance*. Office of Solid Waste, Permits and State Programs Division. US Environmental Protection Agency, Wahington, D.C. EPA/530-93003.
- [9] EPA, 1992b: *Methods for Evaluating the Attainment of Cleanup Standards*, vol. 2: *Ground Water*. Office of Policy, Planning, and Evaluation, US Environmental Protection Agency, Wahington, D.C. EPA 230-R-92-014.
- [10] EPA, 1992c: *Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities – Addendum to Interim Final Guidance*. US EPA, Wahington, D.C.
- [11] EPA, 1992d: *Supplemental Guidance to RAGS: Calculating the Concentration Term*. US Environmental Protection Agency, Office of Solid Waste and Emergency Response, Washington, D.C., 9285.7-081
- [12] EPA, 1994: *Methods for Evaluating the Atteinment of Cleanup Standards, Volume 3: Reference-Based Standards for Soils and Solid Media*. Office of Policy, Planning, and Evaluation, US Environmental Protection Agency, Wahington, D.C. EPA 230-R-94-004.
- [13] EPA, 1994: *Statistical Training Course for Groundwater Monitoring Data Analysis*. Solid Waste and Emergency Response, U.S. EPA, Washington, D.C. EPA/530/R-93/003.

- [14] EPA, 1996: *Soil Screening Guidance: User's Guide*. US EPA, Office of Emergency and Remedial Response U.S. Environmental Protection Agency Washington, D.C. EPA/540/R-96/018.
- [15] EPA, 1996: *Soil Screening Guidance: User's Guide*. Office of Solid Waste and Emergency Response, US EPA, Washington, D.C., EPA 9355.4-23
- [16] EPA, 1997: *Technology Support Center Issue, The Lognormal Distribution in Environmental Applications*. Office of Solid Waste and Emergency Response, US EPA, Washington, D.C., EPA/600/S-97/006
- [17] EPA 1999: *Some Practical Aspects of Sample Size and Power Computations for Estimating the Mean of Positively Skewed Distributions in Environmental Applications*. US EPA, Office of Research and Development, Washington, D.C. EPA/600/s- 99/006.
- [18] EPA, 2000: *Guidance for Data Quality Assessment: Practical Methods for Data Analysis*. Office of Environmental Information, U.S. EPA, Washington, D.C. EPA/600/R-96/084.
- [19] EPA, 2006: *Data Quality Assessment: Statistical Methods for Practitioners, EPA QA/G-9S*. Office of Environmental Information, U.S. EPA, Washington, D.C., EPA/240/B-06/003.
- [20] EPOC (State of Connecticut, Department of Environmental Protection), 2006: *Guidance for groundwater monitoring for demonstrating compliance with the Connecticut remediation standard regulations*. State of Connecticut, Department of Environmental Protection, Bureau of Waste Management Remediation Division.
- [21] Gibbons, R.D. 1991: *Statistical tolerance limits for ground-water monitoring*. Groundwater 29(4), 563-570.
- [22] Gibbons, R.D., 1994: *Statistical Methods for Groundwater Monitoring*. John Wiley & Sons, New York.
- [23] Gilbert, R.O., 1987: *Statistical Methods for Environmental Pollution Monitoring*. Van Nostrand Reinhold Co., New York.
- [24] Gilbert, R.O. et al., 1996: *An Overview of Methods for Evaluating the Attainment of Cleanup Standards for Soils, Solid Media, and Groundwater, EPA Volumes 1, 2, and 3*. Prepared for US EPA under Contract DE-AC06-76RLO 1830
- [25] Hahn, G.J. & Meeker, W.Q., 1991: *Statistical Intervals: A Guide for Practitioners*. John Wiley & Sons, New York.
- [26] Helsel.D.R. &Hirsch,R.M., 1992: *Statistical Methods in Water Resources*. Studies in Environmental Science 49, Elsevier, Amsterdam.
- [27] Hayes T.D., et al., 1996: *Management of Manufactured Gas Plant Sites, Vol. 1*, Amherst Scientific Publishers, 1996, ISBN 1-884940-06-4).
- [28] IAEA (International Atomic Energy Agency), 1994: *Soil sampling for environmental contaminants.*, Viena, IAEA-TECDOC-1415.
- [29] IEPA (Illinois Environmental Protection Agency), 1997: *Tiered Approach to Corrective Action Objectives (TACO)*, Fact Sheet 10: Compliance with Remediation Objectives.

- [30] Jewel, Ch. M. et al., 1993: *Site investigation and monitoring sampling techniques for contaminated sites and potential waste disposal sites*. Geotechnical Management of Waste and Contamination, Balkema Rotterdam, ISBN 9054103078.
- [31] McNichols, R. J. & Davis, C.B., 1988: *Statistical issues and problems in ground water detection monitoring at hazardous waste facilities*. Groundwat. Monitor. Review 8(4), 135-150.
- [32] MPCA (Minnesota Pollution Control Agency), 1996: Draft Guidelines Risk Based Site Characterization and Sampling Guidance
- [33] MPCA (Minnesota Pollution Control Agency), 2008: Excavation of Petroleum Contaminated Soil and Tank Removal Sampling – Guidance Document 3-01
- [34] NDDH (North Dakota Department of Health), 2009: Guideline 7 – Land Treatment of Petroleum Contaminated Soil
- [35] NJDep (New Jersey Department of Environmental Protection), 1995: *Compliance Averaging*. Site Remediation News, Vol 7, No 2, Article 08.
- [36] NJDep (New Jersey Department of Environmental Quality), 2009: N.J.A.C. 7:26E Technical Requirements for Site Remediation ("Tech Rules").
- [37] NYSDEC (New York State Department of Environmental Conservation), 2002: Technical Guidance for site investigation and Remediation
- [38] Ohio EPA, 1995: *Technical Guidance Manual for Hydrogeologic Investigation and Ground Water Monitoring, Chapter 13, Statistics for Ground Water Quality Comparison*.
- [39] Ott, W.R. (1990) *A physical explanation of the lognormality of pollutant concentrations*. J. Air Waste Management Ass. 40, 1378-1383
- [40] PDEP (Pennsylvania Department of Environmental Protection) 2002: Guidance for attainment Demonstration with Statistical Methods 253-0300-100.
- [41] PDEP (Pennsylvania, Department of Environmental Protection), 2001: *Groundwater Monitoring Guidance Manual*. Commonwealth of Pennsylvania, Department of Environmental Protection. Document Number 383-3000-001.
- [42] PDEP (Pennsylvania, Department of Environmental Protection), 2002, revised 2008: *Land Recycling Program Technical Guidance Manual, Section IV.B - General Guidance, Guidance for Attainment Demonstration with Statistical Methods*. Commonwealth of Pennsylvania, Department of Environmental Protection, Document Number 253-0300-100.
- [43] Sarz, Martin, M., 2003 (2nd edition): *Site Assessment and Remediation Handbook*, Lewis Publishers, ISBN 0-56670-577-0
- [44] TEPA (Texas Environmental Protection Agency), 1998: *Use of Statistics for Determining Soil/ Groundwater Cleanup Levels under the Risk Reduction Rules*.
- [45] UK Environment Agency, 2000: *Secondary Model Procedure for the Development of Appropriate Soil Sampling Strategies for Land Contamination*. R&D Technical Report P5-066/TR. Bristol, ISBN: 1 85705 577 2.
- [46] UK EPA, 2003: *Guidance on Monitoring Landfill Leachate, Groundwater and Surface Water*. Environment Agency, Bristol, ISBN 1-84432-159-2

- [47] Victoria EPA, 2009: *Industrial Waste Resource Guidelines, Soil Sampling*. EPA Victoria, Publication IWRG702-June 2009.
- [48] WEBER, E. F., 1995.: *Statistical methods for assessing groundwater compliance and cleanup: a regulatory perspective*. Groundwater Quality: Remediation and Protection (Proceedings of the Prague Conference, May 1995). IAHS Publ. no. 225.
- [49] Zirchsky, J. and Gilbert, R.O., 1984: *Detecting Hot Spots at Hazardous Waste Sites*, Chemical Engineering, 91: 97 – 100.

Výběr statistických programů, bezúplatně dostupných na Internetu:

- **Zaitun Time series** - freeware program pro řešení časových řad (<http://www.zaitunsoftware.com/>) – trendová analýza, dekompozice časových řad, klouzavé průměry, exponenciální vyrovnávání, lineární regrese, korelogramy, neuronové sítě.
- **KyPlot** – freeware statistický program, velmi podobný Excelu. Dostupný na adrese: (<http://www.freeststatistics.info/stat.php>). Na této stránce jsou k dispozici další statistické a matematické programy včetně doplňků pro Excel. Program obsahuje velký počet statistických metod včetně řešení shlukové analýzy.
- **PAST** – freeware statistický program pro paleontologii, nicméně výborně použitelný také v hydrogeologii. Dostupný na adrese: (<http://folk.uio.no/ohammer/past/>). Program obsahuje velké množství statistických metod (klasické statistické metody, regrese, časové řady, shlukovou analýzu – clustering atd.).
- **Dplot Jr** – doplněk pro Excel. Umožňuje vytvářet různé typy grafů v Excelu včetně trojúhelníkových grafů. Plná verze je komerční (Dplot), freewarová verze je nazvaná DPlot Jr. Ta plně dostačuje pro běžnou hydrogeologickou praxi. Instaluje se jako doplněk pro Excel. Dostupný na adrese: (<http://www.dplot.com/index.htm>).
- **IDAMS** – freewarový statistický program dostupný na adrese (http://portal.unesco.org/ci/en/ev.php-URL_ID=2070&URL_DO=DO_TOPIC&URL_SECTION=201.html). Program šířený prostřednictvím UNESCO pod názvem WINIdams. Je zaměřený především na časové řady.

Širokou škálou běžnějších statistických operací disponuje i MS Excel.