

Univerzita Palackého v Olomouci

Přírodovědecká fakulta

Katedra geoinformatiky

**GEOSTATISTICKÉ ANALÝZY ÚČINNOSTI
NOVÝCH ENVIROMENTÁLNĚ ŠETRNÝCH
TECHNOLOGIÍ PRO ČIŠTĚNÍ VODY A PŮD**

Magisterská práce

Bc. Karel MACKŮ

Vedoucí práce Mgr. Pavel TUČEK, PhD.

Olomouc 2015

Geoinformatika

ANOTACE

Tato diplomová práce řeší komplexní geoinformatický úkol aplikovaný na problematiku kontaminace podzemních vod. Hlavním cílem je zpracování dostupných dat a provedení celkové analýzy situace na zájmové oblasti případové studie se zaměřením na predikci, vzájemné vztahy a prostorové charakteristiky jevů. Práce byla řešena na dvou ukázkových lokalitách, které se lišily technologiemi dekontaminace a zaměřením analýzy – v první lokalitě byl kladen důraz na atributovou složku dat, v druhé lokalitě na prostorovou složku. Během řešení byly otestovány různé geostatistické přístupy ke zpracování dat, jakými jsou například základní popisné statistické charakteristiky, korelační výpočty, testování statistických hypotéz, lineární a nelineární modely pro časovou predikci, shluková analýza nebo prostorové interpolace. Důraz byl kladen na vhodnou grafickou vizualizaci výsledků v podobě grafů nebo datových náhledů. Zjištěné výstupy práce byly řádně okomentovány. Závěr práce hodnotí kvalitu používaných dekontaminačních technologií a stav zkoumané lokality v čase posledních dostupných dat. Byla zjištěna silná účinnost čištění podzemní vody a dobrá funkčnost celého systému. Pomocí predikovaných hodnot byl stanoven předpokládaný výhled do blízké budoucnosti. Zároveň práce ukázala vhodnost zkoumaných metod pro daný typ úkolu, kde se některé velice osvědčily, a jiné byly pro tento typ dat spíše nevhodné.

KLÍČOVÁ SLOVA

podzemní voda; znečištění; geostatistika; dekontaminace

Počet stran práce: 55

Počet příloh: 5 (z toho 2 volné)

ANOTATION

This diploma thesis deals with complex geoinformatics task applied on issue of groundwater contamination. Main goal is processing of available data, performance of comprehensive analysis of situation in case area with a focus on future prediction, relationships between phenomena and spatial distribution and characteristics. Thesis solved two sample locations, that differed in decontamination technologies and orientation of used analysis – first location was concentrate on attribute part of data and second on spatial component. Various geostatistical approaches were applied on data processing, e.g. descriptive statistical characteristics, correlation calculations, statistical hypothesis testing, linear and nonlinear models for time prediction, clustering or spatial interpolation. Emphasis was placed on a suitable graphic visualization of the results, as graphs or data views. Detected outputs of the analysis has been properly commented. Finally, the work evaluate quality of the used decontamination technologies and state of sample location at the time of the latest available data. Strong groundwater decontamination efficiency and good functionality of entire system has been detected. Using the predicted values, predictions to near future have been determined. Thesis also demonstrated appropriateness of different research methods for this type of tasks – some of them were very suitable, another seemed to be inappropriate.

KEYWORDS

Groundwater; pollution; geostatistics; decontamination

Number of pages: 55

Number of appendixes: 5

Prohlašuji, že

- diplomovou práci včetně příloh, jsem vypracoval samostatně a uvedl jsem všechny použité podklady a literaturu.

- jsem si vědom, že na moji diplomovou práci se plně vztahuje zákon č.121/2000 Sb. - autorský zákon, zejména § 35 – využití díla v rámci občanských a náboženských obřadů, v rámci školních představení a využití díla školního a § 60 – školní dílo,

- beru na vědomí, že Univerzita Palackého v Olomouci (dále UP Olomouc) má právo nevydělečně, ke své vnitřní potřebě, diplomovou práci užívat (§ 35 odst. 3),

- souhlasím, aby jeden výtisk diplomové práce byl uložen v Knihovně UP k prezenčnímu nahlédnutí,

- souhlasím, že údaje o mé diplomové práci budou zveřejněny ve Studijním informačním systému UP,

- v případě zájmu UP Olomouc uzavřu licenční smlouvu s oprávněním užít výsledky a výstupy mé diplomové práce v rozsahu § 12 odst. 4 autorského zákona,

- použít výsledky a výstupy mé diplomové práce nebo poskytnout licenci k jejímu využití mohu jen se souhlasem UP Olomouc, která je oprávněna v takovém případě ode mne požadovat přiměřený příspěvek na úhradu nákladů, které byly UP Olomouc na vytvoření díla vynaloženy (až do jejich skutečné výše).

V Olomouci dne

Bc. Karel MACKŮ

Děkuji vedoucímu práce Mgr. Pavlu Tučkovi, PhD. za odborné konzultace a podněty k vypracování práce. Za poskytnuté rady a materiály děkuji pracovníkům firmy Aquatest a.s.

Vložený originál **zadání** bakalářské/magisterské práce (s podpisy vedoucího katedry, vedoucího práce a razítkem katedry). Ve druhém výtisku práce je vevázána fotokopie zadání.

OBSAH

ÚVOD	8
1 CÍLE PRÁCE	9
2 METODY A POSTUPY ZPRACOVÁNÍ	10
3 PROBLEMATIKA PODPOVRCHOVÝCH VOD	12
3.1 Vznik podzemní vody	12
3.2 Druhy podzemních vod	12
3.2.1 Voda provzdušněného pásma	13
3.2.2 Voda nasyceného pásma	13
3.3 Ochrana a znečištění podzemních vod	15
4 PŘÍPADOVÁ STUDIE / ŘEŠENÁ LOKALITA	16
4.1 Podzemní těsnící stěna	18
4.2 Mrak M5.....	20
5 PODZEMNÍ TĚSNÍCÍ STĚNA	22
5.1 Základní popisné charakteristiky.....	22
5.2 Korelační matice	24
5.3 Změny na reakčních branách.....	29
5.3.1 Grafy závislostí látek	29
5.3.2 Shluková analýza	30
5.3.3 Fyzikálně-chemické parametry	32
5.4 Nelineární regrese.....	34
5.4.1 Beta regrese.....	34
5.4.2 Logistická regrese	36
6 MRAK M5	40
6.1 Regresní model.....	40
6.2 Prostorové hodnocení	42
6.2.1 Popisné charakteristiky	43
6.2.2 Kriging	44
6.2.3 Prostorová analýza vrtů	46
6.2.4 Vizualizace v animaci	49
6.2.5 Rozložení látek v prostoru	50
7 VÝSLEDKY	52
7.1 Výsledky analýzy PTS	52
7.2 Výsledky analýzy M5.....	53
8 DISKUZE	54
9 ZÁVĚR	55
POUŽITÁ LITERATURA A INFORMAČNÍ ZDROJE	
PŘÍLOHY	

ÚVOD

Problematika pitné vody je v mnoha částech světa vysoce aktuálním tématem. V naší společnosti je její přítomnost považována za samozřejmost, avšak tento přístup není správný. Voda je drahocenný přírodní zdroj, bez kterého by nemohl svět tak, jak jej známe dnes vůbec existovat.

S rozvojem průmyslu a technologického pokroku obecně je spojen silný dopad na okolní životní prostředí. Vedlejší produkty výroby, nesledované úniky chemikálií či chemické havárie – všechny tyto faktory mohou výrazně kontaminovat půdu a podzemní vodu. V důsledku podzemního proudění se tyto chemikálie mohou dále šířit a ohrožovat okolní ekosystémy či zdroje pitné vody. Dnes je již téměř samozřejmostí recyklace vody v čističkách, nicméně stále dochází k akutním situacím výrazných lokálních znečištění, které je třeba řešit odbornými a náročnými sanačními pracemi.

Různé technologie mohou mít různou účinnost a vhodnost v rozdílných podmínkách. Je proto dobré určitým způsobem hodnotit efektivitu a kvalitu používaných technologií. Geoinformatika nabízí rozličné nástroje datové analýzy na prostorové či atributové úrovni. V rámci této práce budou pomocí různých přístupů analyzována data znečištění podzemních vod a zkoumána dlouhodobá účinnost používaných technologií.

Diplomová práce byla řešena v rámci projektu číslo: TE01020218 s názvem „Ekologicky šetrné nanotechnologie a biotechnologie pro čištění vod a půd“, který byl řešen Regionálním centrem pokročilých technologií a nanomateriálů.

1 CÍLE PRÁCE

Cílem diplomové práce je vypracovat komplexní geoinformatický úkol v podobě zpracování dat s tematikou znečištění podzemních vod. Hlavní důraz je kladen na úpravu, zpracování a analýzu dat, vhodnou vizualizaci a celkovou interpretaci výsledků. Z vybrané lokality byla získána data od zúčastněné firmy, která se problematikou zabývá, takže sběr dat mezi cíle práce nepatří, avšak jejich vhodná úprava a následné zpracování ano. Na vybraných lokalitách budou zkoumány vztahy mezi sledovanými atributy a hledání vzájemných souvislostí. Důraz je kladen také na predikci vývoje celého systému do budoucna, na základě kterého lze vyvozovat důsledky o životnosti a dlouhodobé efektivitě používaných technologií.

Jelikož se jedná o práci z geoinformatiky, důležitá je také prostorová složka všech sledovaných jevů. Je nutno lokalizovat centra největšího znečištění a sledovat proměnlivost jevu v prostoru a čase.

Výsledkem práce bude vyhodnocení situace na zájmovém území, popis účinnosti použitých technologií čištění vody a hlavně přehled vybraných analytických metod a výsledků jimi dosažených, porovnání jejich vhodnosti pro tento typ úkolů a dat. Na základě zhodnocení metod lze potom navrhnout optimální postupy pro řešení podobné problematiky.

K diplomové práci budou vytvořeny internetové stránky, a také poster graficky prezentující celou práci.

2 METODY A POSTUPY ZPRACOVÁNÍ

Použitá data

Na základě stáže u firmy Aquatest a.s. byl k dispozici rozsáhlý dataset, který bylo nutno zpracovat a provést na něm veškeré vhodné operace tak, aby byl problém komplexně zanalyzován a šlo na základě výsledků prohlásit tvrzení o stavu lokality.

K dispozici byly původně tři datové sady, zaměřující se na tyto oblasti:

- Data z 10 reakčních bran v mraku M2
- Data z vrtů v mraku M2
- Data z vrtů v mraku M5

Každá datová sada je zpracována ve formátu .xls a rozdělena do několika tematických listů: nejprve je uveden přehled dostupných vrtů, jejich název a zeměpisná souřadnice v systému S-JTSK, včetně výškové souřadnice. Druhým tématem je hladina podzemní vody, dále fyzikální parametry odebraného vzorku, nejdůležitější část – přehled kontaminátů a jejich hodnoty, a jako poslední - list hodnot obsažených anorganických látek. U všech záznamů na všech listech je vždy uvedeno datum a čas odběru vzorku a následně hodnoty koncentrace sledovaných látek v jednotkách miligramů na litr roztoku.

Každá z datových sad obsahovala velké množství atributů, z nichž nebyly všechny důležité a nebo neobsahovaly žádná data. Již během konzultace s odborníky byly tyto atributy protříděny tak, aby byly zachovány jen ty, na kterých bude dále prováděna analýza. Přehled těchto atributů uvádí tabulka:

Tab. 1 Přehled datových sad

DATOVÝ LIST	ATRIBUTY
Přehled vrtů	souřadnice X, Y, Z
Hladina podzemní vody	datum a místo odběru, hladina podzemní vody m n. m.
Fyzikálně-chemické parametry	datum a místo odběru, pH, konduktivita, oxidačně-redukční potenciál
Chlorované uhlovodíky	datum a místo odběru, chlorované ethyleny, methyleny a jejich deriváty
Anorganické látky	datum a místo odběru, chloridy, sírany, dusičnany, hydrogenuhličitan, vápník, železo

Tato struktura dat byla zachována ve všech zkoumaných oblastech, za účelem jednotného zpracování a přístupu.

Posledním krokem předzpracování dat bylo rozdělení celé sady do dílčích částí, na základě potřeby pro jednotlivé analýzy. Tento úkon zahrnoval například rozdělení datasetu týkajícího se reakčních bran do dílčích kusů pro každou bránu, výběr dat patřící do vstupu a výstupu, seskupení dat podle stejného data, výběr maximálního počtu dat z vrtů ve stejném časovém období pro prostorovou interpolaci a podobně. Jelikož původní data byla

velice rozsáhlá, vybrat ke každé analýze pouze potřebné údaje byl úkon časově velice náročný.

Použité programy

Při řešení práce bylo využito jen malé množství programů. Základní tíha předzpracování dat a jejich přípravy pro další analýzy a vizualizace ležela na produktu Microsoft Excel. Statistické analýzy byly prováděny v programu RStudio, kde kromě základní sady balíčků, které tento program nabízí, byl dále využíván balík *cor2cor* pro výpočet parciálních korelačních koeficientů, balík *betareg* umožňující vytvořit regresní model a balík *ggplot2*, který nabízí silné nástroje pro kvalitnější grafické výstupy. Mapové výstupy, GIS operace byly prováděny v ArcGIS 10.1, kde bylo mimo standartní nastavení použito extenze *Geostatistical analyst*, která nabízí přehledného průvodce pro prostorové interpolace. Pro kvalitnější grafické výstupy byl částečně využit Adobe Illustrator, některé výsledky jsou prezentovány jako animace v Google Earth.

Postup zpracování

Během řešení práce proběhlo několik dílčích fází. Na počátku stála pouze teoretická rešerše problematiky podzemních vod a jejich znečištění. Tento teoretický úvod sloužil spíše jako doplňující informace ke všeobecnému teoretickému přehledu o problematice.

Důležitým bodem pro vývoj práce byla stáž u firmy Aquatest a.s., kde proběhlo bližší uvedení do problému, seznámení se s lokalitou jak na teoretické úrovni, tak osobní návštěvou, a hlavně získání dat potřebných pro analýzu, jejich popis, a základní výběr důležitých atributů. S potřebnými daty se mohl rozběhnout proces zpracování celé analýzy. Data bylo nutné upravit do dílčích datasetů, které odpovídaly jednotlivým oblastem a tématům, vybrat z nich jen významné atributy a správně je strukturovat, aby práce s nimi byla co nejjednodušší. Na základě dat mohly být také voleny další metody zpracování.

Následoval tedy proces vykonání jednotlivých dílčích analýz, které jsou detailně popsány v dalších kapitolách. Výsledky byly co nejpřehledněji shrnuty do tabulek, grafů, či datových náhledů.

V posledním kroku proběhl finální proces spojování výsledků do závěrečného hodnocení situace řešené problematiky. Byla zhodnocena vhodnost jednotlivých použitých metod, a vybrán obecný postup pro analýzy znečištění podzemních vod, založené na podobných datech a principech.

3 PROBLEMATIKA PODPOVRCHOVÝCH VOD

Hydrologie podzemních či podpovrchových vod je vědní dílčí disciplína hydrologie zabývající se podzemní vodou jako součástí oběhu vody v přírodě a vztahy mezi ní a horninným prostředím, v němž se pohybuje a v něm nabývá svých fyzikálních a chemických vlastností (Netopil, 1970). Je zde zahrnuta veškerá voda nacházející se pod zemským povrchem. Můžeme dále rozlišovat tři základní skupiny (přednáškové texty z hydrologie):

- podpovrchová voda: voda ve všech skupenstvích nacházející se v zemské kůře
- podzemní voda: podpovrchová voda v kapalném skupenství. Zabývá se jí hydrogeologie.
- půdní voda: voda obsažená v půdě bez ohledu na skupenství, nevytváří souvislou hladinu. Zabývá se jí hydropedologie.

3.1 Vznik podzemní vody

Netopil klasifikuje podzemní vodu podle jejího vzniku na dva základní druhy, a to podle prostředí, ve kterém se tvoří. **Juvenilní** podzemní voda vzniká hluboko v zemské kůře, kde se z podzemních hmot uvolňují plyny včetně vodíku a kyslíku, jejichž molekuly se mohou při působení vysokého tlaku slučovat a tak vzniká vodní pára. Ve vyšších chladnějších vrstvách zemské kůry pára kondenzuje a může se mísit s vodou přicházející ze zemského povrchu. V oblastech se silnou endogenní aktivitou, např. sopečné oblasti, může juvenilní voda pronikat až na povrch ve své plynné podobě například ve formě gejzírů či součástí horkých minerálních pramenů. Juvenilní voda tvoří poměrně malou část podzemních vod (www.geotech.fce.vutbr.cz/).

Druhou skupinou je voda **vadózní**, která se dostává do půdy vsakováním povrchových srážek. Proniká do hornin z povrchu země a pohybuje se v mělkých vrstvách zemské kůry. Tvoří největší část zásob podzemní vody a je stálou složkou koloběhu vody v přírodě (Netopil, 1970). Existují dvě hlavní cesty, jak se voda z povrchu dostává pod zem. Buď proniká jako vodní pára z přízemních vrstev atmosféry do dutin a pórů hornin, kde se následně ochlazuje a kondenzuje (kondenzační teorie). Tento způsob je významný převážně v suchých oblastech, které jsou charakteristické nízkým množstvím atmosférických srážek, zatímco v humidních oblastech význam klesá. Druhá cesta vody je vsakováním spadlých srážek nebo povrchové tekoucí vody do propustných půd, zemin a hornin (infiltrační teorie). Vsakování vody ovlivňuje spousta činitelů, ať už z klimatických či vegetačních či pedologických hledisek. Hledí se na množství, druh, trvání srážek; teplotu, tlak vzduchu a výpar; sklon a drsnost povrchu, druh a typ půd, pórovitost hornin nebo stupeň nasycení půd a hornin vodou.

3.2 Druhy podzemních vod

Podmínkou existence podzemní vody je výskyt volných dutin, kde se voda může ukládat. V půdách tyto prostory nazýváme průliny, v pevných horninách pukliny. Jsou-li prostory zčásti naplněné vodou, pak se jedná o pásmo provzdušněné. Pod ním se zpravidla

nachází pásmo nasycení (jinak také zvodnělá vrstva, horizont), kde jsou prostory zcela vyplněny vodou (Netopil, 1970).

3.2.1 Voda provzdušněného pásma

Voda přítomná v tomto pásmu se také někdy označuje jako půdní vláha. Rozdělujeme zde několik druhů (www.is.muni.cz):

Vodní pára: voda v plynném skupenství vyplňující volné prostory v hornině. Vzniká výparem v prostředí o vyšší teplotě a při ochlazení může zpětně kondenzovat.

Adsorpční voda: tvoří na povrchu zrn nebo puklin nesouvislé nebo souvislé blanky o tloušťce 5 až 15 molekulárních vrstev, poutané k jejich povrchu fyzikálními silami přitažlivost. Množství adsorpční vody v hornině je velmi proměnlivé, závisí na množství vodních par. Změny množství jsou nejsilnější ve svrchní mělké vrstvě, kde je velký vliv kořenového systému rostlin.

Kapilární voda: vyplňuje póry menší než 1 mm a pukliny menší než 0,25 mm, v nichž je k povrchu pevné hmoty poutána kapilární silou. Trvale se kapilární voda vyskytuje v jemnozrnných sypkých horninách nad hladinou podzemní vody, kde vytváří pásmo kapilárního zdvihu (Netopil, 1970). To se pohybuje společně s hladinou podzemní vody. kapilární zdvih má v přírodě velký význam, hlavně v suchých oblastech, kde umožňuje rostlinám přístup k vodě.

Vsakující voda: voda pronikající ze zemského povrchu do vrstev půd a hornin. Účinkem gravitace se od povrchu pohybuje póry do hlubších vrstev, kde se její část váže na povrch horninných částic jako adhezni či kapilární voda. Hloubka jejího pronikání závisí jak na množství vody, které dopadne na zemský povrch, tak na její spotřebě směrem k jiným druhům půdní vody (www.is.muni.cz).

Půdní led: vzniká z vodní páry při teplotách pod 0 °C, která se na podchlazených stěnách půdních puklin usazuje a díky přítomnosti adsorpční a kapilární vody se ledové krystalky zvětšují. Jejich zvětšováním narůstá tlak na okolní stěny a může tak docházet k jejich narušování.

3.2.2 Voda nasyceného pásma

Výskyt podzemní vody nasyceného pásma je vázán na existenci dutin, které jsou zcela zaplněny vodou. V těchto dutinách se může voda volně pohybovat, její pohyb je vyvolán výhradně gravitační silou (Netopil, 1970). Netopil také podle hloubky výskytu vody rozděluje 3 hlavní pásma:

- **pásmo svrchní:** zde probíhá intenzivní výměna vody podzemní a povrchové. Voda je zde převážně sladká, slabě mineralizovaná.
- **střední pásmo:** zpomalené výměny vody, která podmiňuje silnější mineralizaci vody při jejím dlouhém pobytu v horninovém prostředí o vyšší teplotě.
- **spodní pásmo:** velmi zpomalená výměna vody, charakteristická silnou mineralizací. Vyskytují se v něm často vody slané s vysokým obsahem chloridů.

Hranice a struktura výše uvedeného dělení je obtížné jednoznačně definovat, jsou totiž silně závislé na reliéfu a geologické stavbě území. Hranice mezi prvním a druhým pásmem se může nacházet v hloubce od několika desítek metrů až po řády kilometrů (v hornatých oblastech).

Z hlediska pohybu vody je nutno dále rozlišovat vodu **průlinovou** a **puklinovou**. Jako průliny označujeme póry, mezery mezi zrny zpevněných i nezpevněných sedimentů. Pukliny jsou dutiny různého původu nacházející se ve zpevněných horninách (přednáškové texty z fyzické geografie).

Průlinová voda se v horninách pohybuje pomocí filtračního proudění, což je pohyb velmi pomalý, řádově v desítkách centimetrů za den v jemnozrnných horninách, v řádech metrů v hrubozrnných horninách. Pohyb vyvolaný gravitací je silně zpomalován odporem půdních částic. Procesem filtrace se voda zbavuje některých látek, které by se mohly dostat do hlubších vrstev, dochází tedy k přirozenému procesu čištění. Pomalým pohybem horninou může docházet k rozpouštění a tak se voda obohacuje a mineralizuje. Průlinová voda vytváří plošně rozsáhlejší **zvodnělé vrstvy** se souvislou vodní hladinou. Vertikální rozsah zvodnělé vrstvy je dán hloubkou podložní nepropustné vrstvy a polohou hladiny vody - mocnost zvodnělé vrstvy. Nachází-li se nad hladinou podzemní vody další propustná vrstva, voda se v ní může volně pohybovat a potom hovoříme o hladině volné. Hladina ohraničená nepropustnou vrstvou ze spodního i vrchního směru, bez možnosti pohybu, se nazývá napjatá vodní hladina. Dojde-li však k prokopání nebo provrtání krycí vrstvy, vystoupí v otvoru do vyšší polohy zvané výstupná výška (též tlačná nebo piezometrická výška). Ustálí-li se hladina pod úroveň terénu, je výstupná výška negativní, vystoupí-li však ve vrtu nad povrch terénu nebo vytéká-li voda z otvoru, je výstupná výška pozitivní (www.is.muni.cz).

Zvláštním druhem podzemní vody s napjatou hladinou je **voda artézská**. Jedná se o vodu ležící v hlubších zvodněných vrstvách, kde je pod velkým tlakem a v případě narušení svrchních vrstev vytéká či vystřikuje na zemský povrch.

Puklinová voda se pohybuje podle zákonů platných pro proudění vody, a pukliny vyplňuje z části nebo zcela. Při zaplnění puklin drobnými úlomky hornin či jemnými částicemi se voda pohybuje podobně jako voda průlinová podle zákonů filtrace. Puklinová voda se ve volných prostorách nachází jen krátkou dobu, proto nemůže účinně působit jako rozpouštědlo ani nějak výrazně měnit své chemické vlastnosti. Vycházejí-li pukliny až na zemský povrch, může tudy pronikat do spodních vrstev znečištěná voda a dále se šířit. Pomocí rozsáhlých puklinových systémů se voda může šířit na velké vzdálenosti i ve velkých hloubkách.

Zvláštním případem puklinové vody je voda **krasová**. Vyskytuje se na území krasovitých vápenců, dolomitů a jejich sutí. Voda zde vtéká pod zem ne jen úzkými puklinami, ale také širšími vstupy jako jsou závrtky a ponory. V podzemí pak tvoří složité systémy různě pospojovaných toků.

3.3 Ochrana a znečištění podzemních vod

Pitná voda je nezbytná pro každodenní fungování a celkové přežití života na zemi. V našich zeměpisných podmínkách si spousta lidí tento fakt a vzácnost vody možná ani neuvědomuje, avšak pitná voda není samozřejmost. Proto je nutné omezit špatné zacházení s vodními zdroji, ochránit je před znečištěním a případně umět vodu čistit či recyklovat.

Svým objemem a zásobami převyšuje podzemní voda tu povrchovou, a v mnohých oblastech je to jediný zdroj vláh. Celkové zásoby jsou odhadovány na 10 miliard krychlových metrů, což je výrazně více, než je roční objem všech srážek (www.geology.cz). Nadměrné čerpání podzemní vody může způsobit několik negativních jevů, jako je pokles stability povrchu, lokální vyschnutí či rozsáhlé šíření kontaminace.

Ochrana podzemní vody je stanovena v zákonech č. 138/1973 Sb. o vodách /vodní zákon/ a zákona FS ČSFR č. 17/1992 Sb., o životním prostředí. Na ty navazují další konkrétní zákony. Zejména při odvodňování staveb a podzemních objektů musí být vyřešen problém likvidace odpadních vod. Vody nelze volně vypouštět do povrchových toků nebo nechat zasakovat do horninového prostředí. Kvalitativní limity pro vypouštěné vody jsou dány nařízením vlády ČR č. 82/1999 Sb. a pro posuzování potenciálního ohrožení podzemních a povrchových vod používaných stavebními materiály je nařízení vlády ČR č. 81/1999 Sb., kterým se stanoví jejich limity vyluhovatelnosti (www.geologie.vsb.cz). Ochrana vody může být buď **preventivní** (slouží k zabránění kontaminace pomocí různých legislativních opatření) a **reparativní** (sanace již existujících znečištění). To je dále děleno na pasivní a aktivní, podle systému čištění. V této diplomové práci budou řešeny oba z těchto způsobů.

Na našem území existuje několik firem, které se zabývají tematikou čištění podzemních vod a sanačních prací. Příkladem takových firem je například Aquatest a.s. nebo OPV s.r.o.

4 PŘÍPADOVÁ STUDIE / ŘEŠENÁ LOKALITA

V rámci řešení diplomové práce byla navázána spolupráce s firmou Aquatest, a.s. Tato firma nabízí rozsáhlé služby v oblasti hydrologických prací - sanační práce, řešení úkolů v oblasti vodního hospodářství, vyhledávání vodních zdrojů, hydrogeologický průzkum, ekologické posudky a služby, laboratorní analýzy, a nebo také odborné dekontaminace podzemních vod. U liberecké divize firmy Aquatest, a.s. se konala týdenní stáž, kde pod vedením RNDr. Petra Kvapila, PhD. proběhlo bližší seznámení s celkovou problematikou dekontaminačních prací ve vybraných lokalitách. Z databáze firmy byla poskytnuta potřebná data týkající se studovaných lokalit, včetně patřičného výkladu pro porozumění významu jednotlivých atributů těchto datových sad.

Diplomová práce se detailně zabývá vybranými částmi průmyslového areálu Spolchemie, s.r.o. v Ústí nad Labem. Na území tohoto objektu se již několik let provádí monitoring a testování různých čistících postupů, a proto je z této oblasti dostupné dostačující množství dat pro provádění analýz účinnosti.

Společnost Spolchemie je činná již od roku 1856, a od té doby se její areál v Ústí nad Labem postupně rozrůstá a výroba rozšiřuje. V současné době se zaměřuje především na výrobu pryskyřic, rozpouštědel a sloučenin chlóru. Z důvodu dřívějšího neekologického zacházení s chemikáliemi a nevhodné likvidace vedlejších produktů výroby, došlo v oblasti k silnému znečištění podzemní vody. Pro odstranění této nebezpečné ekologické zátěže byly podniknuty patřičné kroky, ze kterých je nejvýznamnější zbudování podzemní těsnící stěny.

Ve zkoumaných oblastech areálu Spolchemie, kterými se tato práce zabývá, se mimo jiné vyskytují chlorované ethyleny a methyleny, jejichž koncentrace a výskyt byly podrobněji analyzovány. Přehled zkratk a názvů zkoumaných látek uvádí tabulka:

Tab. 2 Přehled kontaminátů

Chlorované methyleny	Chlormethan
	Dichlormethan
	Trichlormethan
	Tetrachlormethan
Chlorované ethyleny	Vinylchlorid (VC)
	1.1 dichlorethylen (1.1-DCE)
	1.2. cis / trans-dichlorethylen (1.2-c/t-DCE)
	Trichlorethylen (TCE)
	Tetrachlorethylen (PCE)

V datech jsou tyto hodnoty uvedeny jednak zvlášť, jednak jako součty v sumách.

Suma CIM (chlorovaných methylenů): chlormethan + dichlormethan + trichlormethan + tetrachlormethan.

Suma CIU dle rozhodnutí (s limitem 1 550 µg/l, dále jen suma dle rozhodnutí): tetrachlormethan + 1.1 dichlorethylen + suma 1.2 dichlorethylenu + trichlorethylen + tetrachlorethylen.

Suma CIU dle nového rozhodnutí (s limitem 2 000 µg/l, dále jen nová suma): suma chlormethanů + vinylchlorid + suma dichlorethylenů + trichlorethylen + tetrachlorethylen.

Chlorované uhlovodíky

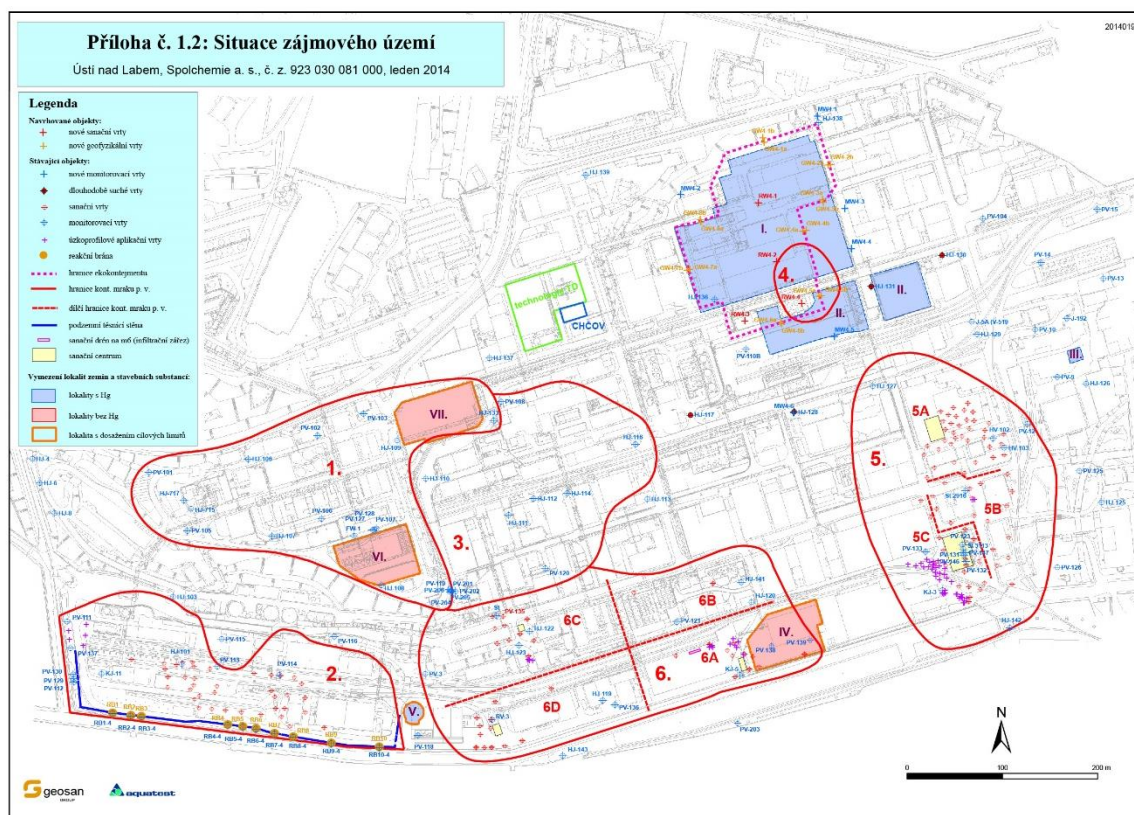
Jak uvádí Bartoň (2007), chlorované uhlovodíky jsou látky syntetického původu, které jsou toxičtější a nebezpečnější než ropné látky. Jsou charakteristické svými nežádoucími vlivy, jako je toxicita, odolnost nebo schopnost ukládat se v živých organismech. Jejich výskyt vždy souvisí s antropogenní činností, jako například v oblasti areálu Spolchemie. Komerčně nejužívanějšími byly trichlorethylen a tetrachlorethylen, které se díky tomu také staly poměrně běžnou kontaminací podzemní vody (Stejskal, 2014). Tyto látky se potom redukčními procesy dále rozkládají na nižší formy, jako jsou trans/cis dichlorethyleny. Stejskal popisuje chlorované ethyleny jako bezbarvé kapaliny s charakteristickým sladkým pachem podobným chloroformu. Výjimkou je pouze vinylchlorid, který se v přírodních podmínkách vyskytuje pouze v plynné fázi, je bezbarvý, zápach je podobný jako u ostatních CIU. Dále uvádí, že jejich výskyt je typický hlavně pro průmyslové oblasti, kde byly chlorované uhlovodíky předmětem výroby a docházelo k nekontrolovaným únikům při špatné manipulaci, transportu či neodbornému ukládání na skládkách. Chlorované uhlovodíky jsou využívány jako odmašťovadla a rozpouštědla, velké využití proto nacházejí v kovozpracujících závodech, strojírenství nebo čistírnách (Bartoň, 2007).

Stejskal dále popisuje toxicitu chlorovaných uhlovodíků, která se liší způsobem vniknutí do těla, dle typu látky a nebo délky vystavení jejímu působení. U **vinylchloridu** byly prokázány karcinogenní účinky na organismus, a také možnost vlivu a narušení lidského DNA, což může být jedna z příčin rakoviny. Dále může způsobovat nekrózu jater. Do těla se dostane prostým vdechnutím, kontaktu s kůží a nebo vnitřně, např. v podobě kontaminované vody. **1.2 dichlorethylen** způsobuje při vyšších koncentracích poruchy nervového systému, poškození jater a oběhového systému. Větší toxicita byla zaznamenána u 1,2-trans-dichlorethylenu než u 1,2-cis-dichlorethylenu. **Trichlorethylen** postihuje nervovou soustavu v podobě slabosti, ospalosti nebo bolesti hlavy. Také má negativní účinek na játra a trávicí soustavu podobně jako ostatní chlorethyleny. **Tetrachlorethylen** se do těla dostává kontaminovanou vodou/potravou, náchylnost při kontaktu s kůží není tak velká. Snadno se ukládá v tukových buňkách. Má podobné účinky na nervovou soustavu a játra jako TCE, dráždí kůži, oči a sliznice, avšak až při vyšších koncentracích, např. při laboratorních experimentech.

Z výše uvedených informací je patrné, že negativní vliv chlorovaných uhlovodíků je velký a proto o významu sanačních prací nelze pochybovat.

Celá lokalita Spolchemie je rozdělena do šesti řešených dílčích oblastí - tzv. mraků. Tyto mraky se liší především ve zjištěných kontaminátech, v použitých technologiích čištění a v době řešení daného mraku. Pro tuto diplomovou práci je stěžejní situace a

informace především z mraku 2, kde byla za účelem zabránění odtoku znečištěné podzemní vody směrem na jih z areálu továrny vybudována podzemní těsnicí stěna (dále jen PTS). Druhou zkoumanou oblastí je mrak 5.



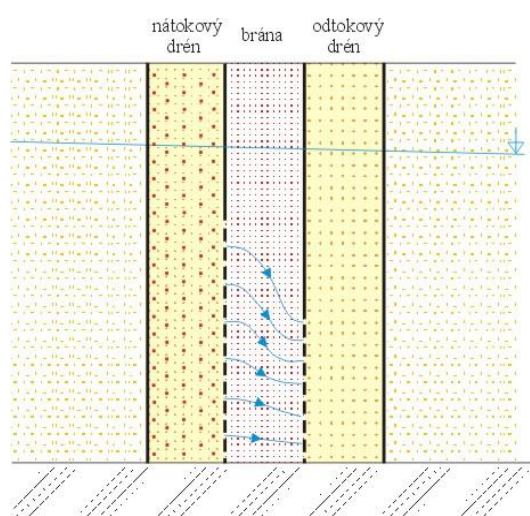
Obr. 1 Areál Splochemie s.r.o. (zdroj: Aquatest a.s.)

4.1 Podzemní těsnicí stěna

Jak uvádí technická zpráva firmy Aquatest, a.s., podzemní těsnicí stěna je nepropustná bariéra usazená do nepropustného podloží, obsahující 10 prostupů, tzv. reakčních bran. Poloha PTS byla vhodně stanovena na základě analýzy hydrologického modelu proudění podzemních vod tak, aby zabránila odtoku kontaminované podzemní vody mimo oblast továrny do blízké řeky Bíliny, kde by způsobovala environmentální nebezpečí. Snahou bylo zajistit plnění stanovených limitů pro maximální přípustnou koncentraci chlorovaných uhlovodíků ve vodě opouštějící areál Splochemie. Tato hodnota byla stanovena v Rozhodnutí KÚ Ústeckého kraje č.j. 159/ZPZ/05/SC-15 ze dne 29.3.2005 na hodnotu 1 550 $\mu\text{g/l}$ (suma všech chlorovaných uhlovodíků). Později byla stanovena nová suma na celkovou hodnotu 2 000 $\mu\text{g/l}$.

Konstrukce PTS začala již v roce 2006 jako spolupráce firmy Aquatest a.s. a Geosan Group a.s. Hodnota průměrné hloubky stěny je přibližně 13 metrů pod povrchem okolního terénu - po analýze hladiny podzemní vody bylo takto zajištěno, aby byla hloubka dostatečná a nedocházelo k odtoku vody pod stěnou. Šířka stěny je 60 cm, a zbudována byla při celkové délce 461 metrů, čímž pokrývá jihozápadní okrajovou část areálu továrny v oblasti mraku 2.

V PTS je vytvořeno celkem 10 reaktivních bran, které společně tvoří hlavní čistící prvek celé PTS. Bránou protéká znečištěná voda ve směru přirozeného směru průtoku a dochází v ní k dekontaminaci. Reaktivní brána má podobu dutého válce o průměru 160 nebo 110 cm, a výšky shodné s hloubkou celé PTS. Tento válec je ze dvou stran perforován tak, aby do něj mohla přitékat a následně z něj odtékat voda - nátoková strana je perforována cca 2/3 – 3/4 ode dna roury a na výtokové straně cca 1/2 ode dna roury, voda tedy protéká přirozeným spádem. Podzemní voda nejprve přitéká perforací výpažnice a protéká nátokovým drénem, který je tvořený směsicí z 90 % štěrku a z 10 % litinovými pilinami. Odsud voda protéká přímo do hlavní reakční části celé brány, která je tvořená samotnými litinovými třískami. V této části dochází k čištění vody, která pak samospádem odtéká do odtokového vsakovacího drénu, který je již vyplněn pouze štěrkem, a dále pryč z reaktivní brány.

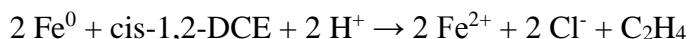
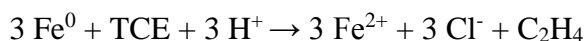


Obr. 2 Průřez reaktivní branou (zdroj: Aquatest a.s.)

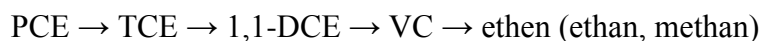
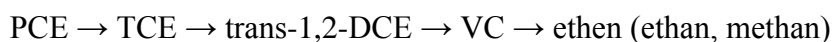
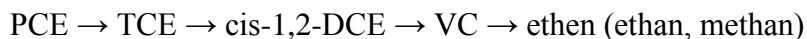
Důležitým prvkem reaktivních bran jsou monitorovací trubky, díky kterým lze provádět sledování hladin podzemní vody a odběr vzorků. V každé braně je celkem 5 takových trubek - 2 v nátokovém drénu, 2 přímo v reakční braně a 1 v odtokovém drénu. V datových sadách jsou vybrány vždy 2 hodnoty ke každé braně – jedna je měřena v monitorovacích trubkách na vstupu, druhá na výstupu. Díky těmto dvěma měřeními lze pozorovat rozdíly, které se projeví při průtoku branou.

Kolem celé PTS byly provedeny vhodné povrchové a terénní úpravy tak, aby stavba nijak nenarušovala okolní prostředí. V současné době se nad stavbou nachází komunikace, na povrchu jsou patrné pouze kryty umožňující přístup ke reaktivním branám a oděru vzorků z monitorovacích trubek.

Samotný princip čištění vody při průtoku reaktivní branou popisuje Kvapil (2007) jako novou generaci sanačních technologií. Jedná se o pasivní odstraňování kontaminátů pomocí železa (viz. litinové třísky sloužící jako náplň reaktivních bran), k jehož oxidaci dochází a současně redukuje nežádoucí deriváty uhlovodíků. Chemický zápis procesu vypadá následovně:



Kvapil (2007) dále uvádí, že v průběhu reakce (dechlorace) uhlovodíků dochází k postupné ztrátě atomů chlóru z molekuly CIU a jejich náhradě atomy vodíku. Vznikají při tom chloridové ionty a několik produktů s menším počtem atomu chlóru popsanych v těchto rozpadových řadách:



Všechny tyto chlorované uhlovodíky jsou měřeny a zaznamenány v dostupných datech, takže lze hodnotit, v jakém stádiu procesu odbourávání se nachází. K procesu redukce dochází v anaerobním prostředí, které se vyznačuje nízkou hodnotou oxidačně-redukčního potenciálu (Eh). Ten je jedním ze sledovaných atributů v datech, v sekci fyzikálně-chemických parametrů. S klesající hodnotou Eh by se teoreticky měla zvyšovat redukce polutantů, a ubývat jich. Redukční podmínky by měly nastat při Eh menším než 50 mV.

Od 18.1. do 16.5.2007 proběhla pilotní zkouška celého systému na reaktivní bráně číslo 4. Zkouška byla složena z chemického monitoringu, karotážního měření a stopovací zkoušky. Monitoring změn chemismu vod probíhal v týdenních intervalech za účelem stanovení obsahu rozpuštěných látek a sledování fyzikálně-chemických parametrů. Dále se sledovala hladina podzemní vody a to ne pouze v oblasti PTS, ale v celém areálu továrny. Karotážní měření mělo za úkol prozkoumat proudění vody v oblasti PTS a jejich reakčních bran, jeho výsledky dále potvrzovala tzv. stopovací zkouška.

Pilotní zkouška měla prokázat funkčnost celého systému tak, aby na základě jejich výsledků bylo možno pokračovat v monitoringu na celé PTS. Jak je uvedeno ve zprávě hodnotící výsledky pilotní zkoušky na RB-4 od doktora Kvapila (2007), celý navržený systém se zdál být jako vyhovující a přinášející dostatečnou účinnost.

PTS byla navržena k funkční životnosti 10 let. Tato doba se pomalu chýlí ke konci a proto je jedním z úkolů diplomové práce zjistit, zda-li za uplynulé období funkčnost postupně klesla, a nebo je-li naopak možné očekávat dobré výsledky i za odhadovanou hranici životnosti.

4.2 Mrak M5

Mrak M5 se nachází ve východní části areálu Spolchemie, a je to druhá ze zkoumaných lokalit, kterými se tato práce zabývá. Mrak se dělí na 3 dílčí části A, B, C. Kontaminace i použité metody jsou ve všech částech stejné, rozdělení je pouze pro potřeby celého projektu. Diplomová práce detailněji studuje pouze část A, ostatními se nezabývá.

Pro mrak 5 platí stejná datová struktura, zkoumané prvky a limitní hodnoty znečištění jako v případě PTS. Podle informační zprávy Stejskala (2014) jsou hlavním polutantem v mraku 5A chlorované methany, ve východní části pak chlorované etheny. Ve zprávě je také popsána metoda čištění v této oblasti: Princip odstranění kontaminantu (CIU) z podzemní vody na mraku M5 je založen na plošné aplikaci in-situ sanačních metod

(oxidace pomocí Fenolova činidla, reduktivní dehalogenace, aplikace nanoželeza), které jsou podporovány sanačním čerpáním kontaminovaných podzemních vod metodou „pump and treat“ (Stejskal, 2014). Voda je čerpána z vrtů, prochází procesem čištění, a po dosažení požadovaných limitů koncentrace polutantů je zasakována zpět do země. Tato metody odčerpání kontaminovaných vod byla později rozšířena o přímou aplikaci redukčních činidel, čímž se intenzita celkového odčerpávání snížila. Pro potřeby diplomové práce byla zpracována data přímo z vrtů pro prostorovou analýzu koncentrací škodlivin, a také celkové sumy procházející čistícím systémem, na kterých bylo predikováno pomocí metod lineární regrese, kdy dojde k celkovému vyčištění, respektive dosažení minimálních požadovaných hodnot koncentrace.

5 PODZEMNÍ TĚSNÍCÍ STĚNA

Jako první byla analyzována data z reakčních bran v podzemní těsnící stěně. Jelikož má PTS liniový charakter a reakční brány jsou na ní zhruba pravidelně rozmístěny taktéž v linii, nenabízí se ze příliš možností na prostorové analýzy typu interpolace, jako tomu bude dále v mraku 5. Důraz zde byl proto kladen na analýzu procesů a chování v samotných reakčních branách. Jedním z požadavků bylo určení hrubého odhadu účinnosti dekontaminačních procesů do budoucna, dále zkoumání možného překročení limitů na jednotlivých branách, a nebo změny v anorganických látkách, nebo fyzikálně chemických parametrech vody.

5.1 Základní popisné charakteristiky

Při analýze datového souboru je vždy vhodné nejprve provést výpočet základních popisných statistik, které dají rychlý prvotní přehled o celé problematice. Z těchto charakteristik bylo zjištěno minimum, maximum, průměrná hodnota, směrodatná odchylka a variační rozpětí souboru, kterým byla data vždy z jedné brány. Hodnoty byly vypočteny pro sumy obou limitů (1 550 i 2 000 µg/l).

Tab. 3 Základní statistiky polutantů na reakčních branách

	Suma CIU dle rozhodnutí				Nová suma CIU			
	Min	Max	Průměr	SMOD	Min	Max	Průměr	SMOD
RB1-1	0,0	4134,9	990,6	1403,2	0,0	4625,9	1098,4	1609,5
RB2-1	2,5	131,2	23,5	33,3	2,7	131,4	26,3	35,9
RB3-1	3,8	21640,0	1169,1	4825,2	4,1	22950,0	1243,0	5256,8
RB4-1	32,2	25914,1	3898,4	4406,3	33,4	26964,1	4127,4	4609,4
RB5-1	21,9	10157,3	2004,1	2871,5	31,4	10649,3	2167,3	3038,7
RB6-1	37,4	3983,3	874,5	1006,1	40,4	4027,3	1065,7	976,2
RB7-1	12,2	440,8	151,3	125,0	13,0	764,5	218,3	220,4
RB8-1	9,2	958,5	103,3	212,6	16,5	970,7	123,1	217,6
RB9-1	2,6	752,2	300,4	237,4	2,8	1027,8	394,6	312,3
RB10-1	7,5	188,5	42,5	49,2	7,5	265,8	56,6	72,0
RB1-4	0,0	159,6	13,5	36,0	0,0	160,7	13,8	37,3
RB2-4	0,3	176,9	12,4	40,0	0,3	178,1	12,5	41,5
RB3-4	0,4	111,3	9,6	24,9	0,4	112,8	9,8	25,9
RB4-4	1,5	16052,4	1800,2	2562,1	1,5	17030,5	1927,0	2738,8
RB5-4	5,8	1399,8	102,1	315,8	7,8	1500,8	119,8	347,3
RB6-4	5,9	322,1	53,7	90,2	10,6	338,6	59,5	94,3
RB7-4	1,4	68,1	12,8	19,4	1,4	114,9	16,3	29,2
RB8-4	0,4	188,4	16,2	42,4	0,4	189,7	17,4	44,5
RB9-4	0,8	118,8	10,9	26,5	0,8	118,8	11,5	27,2
RB10-4	1,1	51,4	6,5	11,4	1,1	52,0	7,0	11,8

V tabulce jsou vstupní hodnoty kódovány jako RBX-1, hodnoty na výstupu RBX-4. Toto označení je převzato z původních dat a bude používáno v celé práci. U většiny bran bylo k dispozici 18 záznamů odběrů chlorovaných uhlovodíků v časovém intervalu duben 2008 až březen 2014. Výraznou výjimkou byla brána 4, kde bylo záznamů podstatně více,

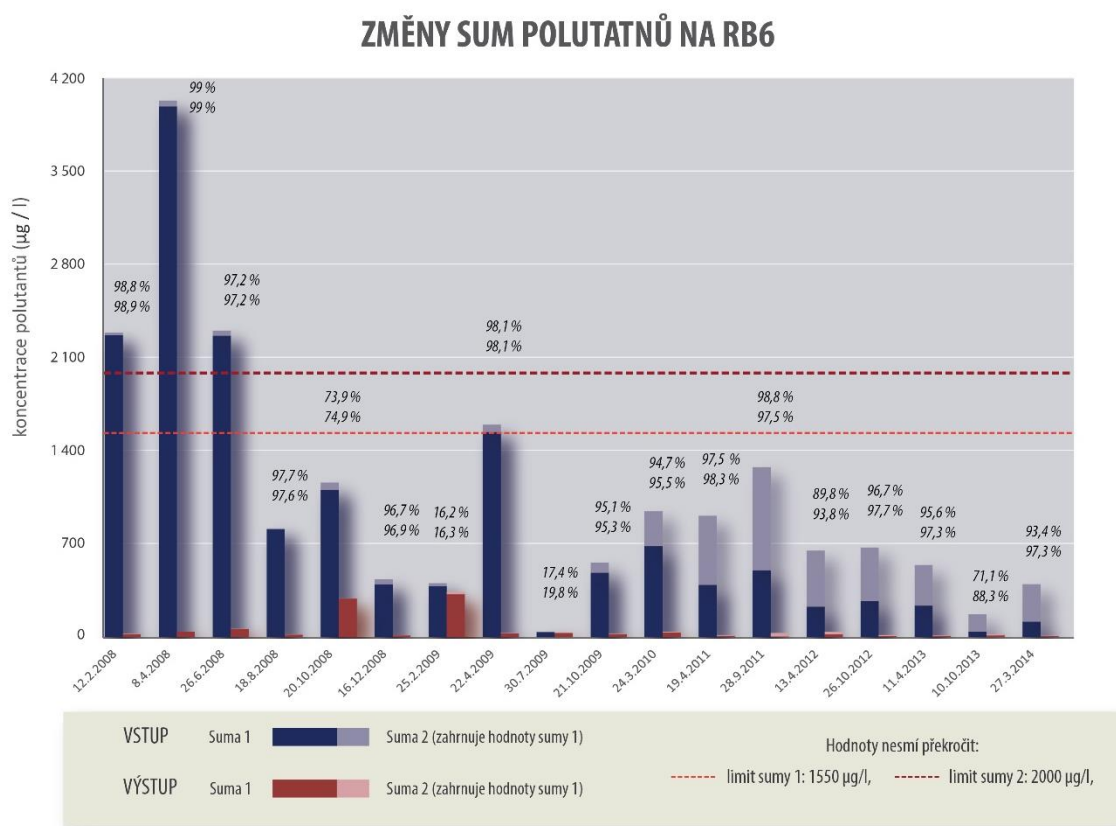
a to z důvodu pilotní zkoušky systému, která zde probíhala už od roku 2007 a během které byla časová intenzita monitoringu vyšší.

Z tohoto základního přehledu je patrné, že vyšší koncentrace znečištění se nachází ve střední části PTS, a to na vstupu bran 3-6. Směrodatná odchylka vypovídá o velkém rozptylu hodnot v rámci měření, tedy mezi jednotlivými lety, a to opět na branách 3-6. Zajímavá je situace na bráně 2, která se prostorově nachází v blízkosti bran 1 a 3, avšak svými nízkými hodnotami mezi ně příliš nezapadá.

Z hlediska zastoupení jednotlivých polutantů v sumách převládá výrazně 1,2-cis-dichlorethylen, jehož celková suma koncentrací na všech branách byla téměř 250 000 µg/l. Další vysoké koncentrace byly zaznamenány u vinylchloridu, trichlorethylenu a tetrachlorethylenu se součtem koncentrací 16 000, 24 000 a 14 000 µg/l.

Při pohledu na hodnoty výstupu je patrné, že kromě brány 4 byly vždy dodržovány žádoucí limity, a to pro obě dvě kategorie. Kromě prostředních bran 4 a 5 jsou všechny hodnoty relativně nízké a vyrovnané. Na základě těchto popisných statistik lze tedy soudit, že ohnisko znečištění se nachází přibližně ve středu PTS mezi bránami 3 až 6.

Avšak pro celkové hodnocení by tato tabulka byla příliš zgeneralizovaná, proto bylo navrženo grafické zpracování, které detailně popisuje všechny měření na každé bráně, přehledně zobrazuje doby, kdy byly překročeny stanovené limity a je doplněno také o údaj efektivity čistícího procesu ve sledovaných obdobích.



Obr. 3 Časový vývoj znečištění na RB6

Na těchto grafech lze přehledně pozorovat velké rozdíly v rámci bran, a i v časovém vývoji na každé reakční bráně. Celkově nízké hodnoty byly zaznamenány na branách č. 2, 3, 7, 8, 9 a 10. Hodnoty zde průběžně kolísají s náhodnými výraznými výkyvy, které však i na vstupu málo kdy překračují dané limity. Výrazně odlišné jsou hodnoty na branách 4, 5 a 6, které po většinu sledovaného období převyšují znečištěním ostatní pozice. Nejvyšší hodnoty byly měřeny hlavně z počátku, cca do poloviny roku 2009, poté již byly zaznamenávány hodnoty nižší. Zvláštní jsou údaje z RB1, která se nachází mimo hlavní centrum bran 4, 5 a 6, a jsou na ní naměřeny vysoké hodnoty. Od ostatních se liší také časovým vývojem- zatímco u zbytku bran je trend spíše klesající, na RB1 spíše rostoucí. Tyto informace tedy indikují nový zdroj znečištění na západní části PTS.

Důležitá je informace, že hodnoty na výstupu z bran nepřekročily žádný ze stanovených limitů, kromě několika záznamů na bráně 4, a to výrazně v období pilotní zkoušky systému, v pozdějších letech bylo překročení limitů už jen nepatrné.

V měřeních se ojediněle vyskytovaly záznamy, kde byly hodnoty na výstupu výrazně vyšší než na vstupu, a to řádově až v desetinásobcích. Tyto hodnoty byly klasifikovány jako chybné záznamy a z datasetu odstraněny.

5.2 Korelační matice

Obecně je často řešen úkol, který se nezabývá pouze jedním základním statistickým souborem, nýbrž závislost či podmíněně změny mezi dvěma statistickými soubory. Také v případě PTS je jedna z pokládaných otázek, jak spolu vzájemně souvisí jednotlivé atributy. Je jich zde měřená celá řada, a proto je vhodné prozkoumat, jestli mezi některými z nich existuje určitý stupeň vazby. K tomuto je nejvhodnější využít korelací.

Podle Brázdila (1981) můžeme vyjádřit proměnlivost znaků dvou souborů pomocí jejich odchylek od průměrů.:

$$d_{x_i} = x_i - \bar{x}, \quad d_{y_i} = y_i - \bar{y}$$

Společným výrazem odchylek d_{x_i} a d_{y_i} je jejich součin $(x_i - \bar{x})(y_i - \bar{y})$.

Upravením o sumu rozsahu výběrů dostaneme výraz jejich proměnlivosti v podobě tzv. kovariance:

$$s_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n}$$

Kovariance vychází v absolutních hodnotách, což omezuje její použití například při porovnávání více kovariancí mezi sebou. Proto kovarianci převádíme na relativní míru těsnosti vztahu dvou výběrů tím, že ji dělíme součinem směrodatných odchylek obou výběrů (Brázdil, 1981).

$$s_{xy} = \frac{\frac{1}{n} \sum(x_i - \bar{x})(y_i - \bar{y})}{s_x s_y} = \frac{\frac{1}{n} \sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum(x_i - \bar{x})^2} \sqrt{\frac{1}{n} \sum(y_i - \bar{y})^2}}$$

Tento zápis nazýváme jako Pearsonův korelační koeficient. Jelikož neber v potaz měrné jednotky a hodnoty jsou relativní, lze mezi sebou srovnávat různé sledované dvojice.

Koeficient může nabývat hodnot na intervalu $(-1,1)$. Podle jeho výsledku hodnotíme míru závislosti mezi dvěma statistickými výběry. Záporná korelace značí, že při růstu hodnot jednoho výběru klesají hodnoty druhého, a kladná korelace zase říká, že při růstu hodnot jednoho výběru rostou také hodnoty výběru druhého. Podle zařazení do intervalu dělíme korelaci na (přednáškové texty ze statistiky):

- nízkou na intervalu $\langle -0,3;0,3 \rangle$
- mírnou na intervalu $\langle -0,5;-0,3 \rangle \cup \langle 0,3;0,5 \rangle$
- výraznou na intervalu $\langle -0,7;-0,5 \rangle \cup \langle 0,5;0,7 \rangle$,
- velkou na intervalu $\langle -0,9;-0,7 \rangle \cup \langle 0,7;0,9 \rangle$
- velmi vysokou na intervalu $\langle -1;0,9 \rangle \cup \langle 0,9;1 \rangle$

Data z PTS byla hodnocena několika způsoby. Korelace byly počítány pro každou bránu na všech sledovaných atributech, tedy všechny záznamy polutantů a pěti sledovaných anorganických látek. Výpočty proběhly na hodnotách ze vstupu do reakčních bran a na rozdílech mezi vstupem a výstupem. Tato úprava dat byla zvolna proto, abych bylo dobře pozorovatelné, jak se ovlivňují vzájemné změny mezi jednotlivými látkami. Poslední výpočet proběhl také na hodnotách z výstupu z reakčních bran, avšak tyto čísla už nejsou příliš zajímavá a důležitá, neboť po výstupu z PTS už se vzorky dál nesledují a nic se s nimi neděje. Navíc z důvodu naprosté redukce některých látek a tudíž jejich nulových hodnot, nemohly být v některých případech korelace ani vypočteny.

Pro vybrané látky byly tedy vytvořeny korelační matice, dále zvýrazněny barevnými škálami pro přehlednější rozlišení silný/slabých a kladných/záporných vztahů. Takto prezentované výsledky jsou přehledné pro sledování a vzájemné porovnávání mezi sebou.

Při analýze vztahu datového souboru, který se skládá z více než dvou sledovaných složek, je často do korelačního vztahu taky zahrnut vliv ostatních proměnných. Proto je potřeba upravit metodu tak, aby byla odhalena čistá závislost pouze dvou konkrétních proměnných. Vliv ostatních proměnných musí být potom vyloučen uměle, a to jejich zkonstantněním. Takto změněný výpočet se nazývá **parciální korelační koeficient**.

$$\rho(X, Y|Z \dots Q) = \rho_{XY.Z\dots Q}$$

Zápis je interpretován jako parciální koeficient mezi X, Y při vyloučení vlivu proměnných Z až Q. Pokud jsou veličiny, figurující v předchozích rovnicích, jen slabě lineárně závislé, nejsou velké rozdíly mezi jednoduchými a dílčími korelačními koeficienty. Naopak, existuje-li mezi těmito veličinami silná lineární závislost, jsou velké rozdíly i mezi hodnotami jednoduchých a dílčích korelačních koeficientů (www.iastat.vse.cz).

Výpočet parciální korelace vychází z klasické korelační matice, kde se počítá parciální korelace 1. řádu (jedna vyloučená proměnná) až do *n-tého* řádu (*n* vyloučených proměnných). Parciální koeficient 1. řádu lze popsat jako:

$$R_{X,Y(Z)} = \frac{R_{X,Y} - R_{X,Z}R_{Y,Z}}{\sqrt{(1 - R_{X,Z}^2)(1 - R_{Y,Z}^2)}}$$

kde popisujeme koeficient prvního řádu vztahu mezi proměnnými X, Y s vyloučením vlivu Z. Jednotlivé koeficienty R značí párový koeficient mezi dvěma proměnnými.

Výpočty byly provedeny v RStudios na upravených datech z Excel. Stačí, aby data byla ve formátu jednotlivých atributů ve sloupcích za sebou, a z nich už je příkazem *cor* vypočtena korelační matice. Celý programový kód pro výpočet matic, včetně matic parciálních korelačních koeficientů vypadá následovně:

```
attach(RB3)
x <- RB3[1:12]
y <- RB3[1:12]
korelace <- cor(x, y)
View(korelace)
partial = cor2pcor(korelace)
View(partial)
write.table(partial, file="vystupRB3.txt")
detach(RB3)
```

Kromě základního nastavení je nutno nainstalovat balík *corpcor*.

1	1,000																	1 VC
2	-0,914	1,000																2 1.2-c-DCE
3	0,728	0,530	1,000															3 1.2-t-DCE
4	0,118	0,358	-0,028	1,000														4 TCE
5	-0,076	-0,035	-0,047	-0,363	1,000													5 PCE
6	0,968	0,904	-0,725	-0,151	0,152	1,000												6 Suma CIM
7	-0,919	-0,993	0,529	0,294	-0,152	0,918	1,000											7 Suma CIU dle Rozhodnutí
8	-0,048	0,147	0,241	-0,514	-0,072	-0,029	0,126	1,000										8 Chloridy
9	-0,375	-0,320	0,561	0,503	0,476	0,355	-0,260	0,210	1,000									9 Dusičnany
10	-0,520	-0,572	0,591	0,166	-0,375	0,681	-0,610	0,106	-0,173	1,000								10 Sířany
11	-0,836	-0,733	0,953	0,194	0,125	0,839	-0,718	-0,078	-0,588	-0,654	1,000							11 Hydrogenuhlíčitany
12	0,492	0,555	-0,664	-0,384	0,015	-0,609	0,550	-0,134	0,539	0,762	0,747	1,000						12 Vápník
	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12						

Obr. 4 Ukázka korelační matice

Bylo tedy vypočítáno 10 matic sledující hodnoty na vstupu do reakčních bran a podávající informaci o tom, jestli jsou výskyty jednotlivých látek mezi sebou podmíněny. Další 10 matic zaměřené na vztahy mezi změnami hodnot popisují ovlivněné chování při úbytku / přírůstku některých látek. Kromě toho byla vytvořena ještě jedna skupina korelačních matic, a ta sleduje závislosti polutantů a anorganických látek, na fyzikálně-chemických parametrech, kterými jsou konduktivita, pH a oxidačně-redukční potenciál. Poslední ze zmíněných je nejdůležitější, neboť nízké hodnoty oxidačně-redukčního potenciálu podporují redukcí polutantů. Bohužel data fyzikálně-chemických parametrů a ostatní sledované látky se špatně časově kryla, tudíž nebylo možné spárovat a použít větší množství záznamů než 6. Je potřeba si uvědomit, že v takovém případě jsou výsledné hodnoty ne zcela přesné a vypovídající, jako by tomu bylo u většího výběrového vzorku.

Vzniklé korelační matice bylo mezi sebou nutné porovnat a vyhodnotit. Vzhledem k jejich velkému množství byla práce velmi nepřehledná, a proto byly upraveny do výsledné tabulky tak, že všechny buňky matice byly uspořádány do řádků pod sebe, a každý sloupec pak odpovídal jedné reakční bráně. Takto byl celý výsledek mnohem přehlednější a umožňoval lepší orientaci. Ukázka takového uspořádání je zde:

Tab. 4 Korelace mezi rozdíly atributů

		RB1	RB2	RB3	RB4	RB5	RB6	RB7	RB8	RB9	RB10	
VC	1.2-c-DCE	-0,059	0,849	0,764	-0,409	-0,036	-0,133	-0,021	0,728	0,130	0,092	
	1.2-t-DCE	0,709	0,484	NA	0,077	0,871	0,334	0,560	NA	-0,341	0,238	
	TCE	-0,108	-0,944	0,666	-0,420	0,367	0,302	-0,644	0,859	0,327	-0,789	
	PCE	-0,088	-0,502	-0,701	-0,414	-0,855	0,383	0,301	-0,438	-0,224	0,788	
	Suma CIM	-0,013	-0,870	0,974	0,128	-0,835	-0,366	-0,833	-0,373	0,319	0,241	
	Suma dle Roz.	-0,067	0,923	0,744	0,410	-0,216	-0,118	0,815	0,692	0,356	0,705	
	Chloridy	0,494	-0,926	0,905	0,494	0,890	0,603	0,822	-0,045	-0,856	0,493	
	Dusičnany	0,578	0,976	0,916	-0,401	-0,279	0,425	0,651	-0,156	0,276	-0,107	
	Sířany	0,022	0,897	0,688	0,586	0,075	-0,405	0,384	0,340	-0,633	-0,519	
	Hydrogenuhlíčitany	0,126	-0,960	-0,872	0,711	-0,951	0,420	-0,332	-0,817	0,476	0,071	
Vápník	0,439	-0,920	-0,877	-0,648	0,474	0,238	0,374	0,050	0,733	0,718		
1.2-c-DCE	1.2-t-DCE	0,693	-0,364	NA	-0,668	-0,245	-0,838	0,708	NA	0,408	0,111	
	TCE	-0,418	0,918	-0,913	-1,000	0,627	0,868	0,706	-0,666	0,098	0,577	
	PCE	0,967	0,104	0,983	-0,999	-0,323	0,854	-0,909	0,507	-0,036	0,382	
	Suma CIM	-0,939	0,985	-0,814	0,268	-0,076	-0,785	-0,290	0,803	0,373	-0,681	
	Suma dle Roz.	-1,000	-0,974	-0,999	1,000	-0,981	-0,999	0,464	-0,976	-0,757	-0,626	
	Chloridy	-0,736	0,895	-0,934	0,015	-0,224	0,525	-0,069	-0,551	-0,163	-0,434	
	Dusičnany	-0,599	-0,869	-0,908	-0,002	-0,598	-0,576	0,375	0,264	-0,404	0,222	
	Sířany	0,828	-0,743	-0,426	0,202	0,446	0,284	0,875	-0,784	0,027	0,234	
	Hydrogenuhlíčitany	0,990	0,900	0,886	0,547	-0,208	0,035	-0,875	0,476	0,089	0,509	
	Vápník	-0,741	0,723	0,693	-0,358	0,421	0,535	0,814	0,535	0,100	-0,644	
1.2-t-DCE	TCE	0,473	0,592	NA	-0,656	0,046	0,557	-0,165	NA	-0,174	0,026	
	PCE	-0,590	0,714	NA	-0,668	0,581	0,503	0,556	NA	0,193	-0,100	
	Suma CIM	0,676	0,456	NA	0,089	0,734	-0,410	0,719	NA	-0,512	-0,057	
	Suma dle Roz.	0,698	-0,457	NA	0,667	-0,080	-0,850	-0,792	NA	0,248	0,018	
	Chloridy	0,203	0,619	NA	-0,416	-0,938	0,076	-0,295	NA	-0,203	-0,187	
	Dusičnany	0,015	-0,454	NA	0,392	-0,102	-0,791	-0,456	NA	0,666	0,321	
	Sířany	-0,549	-0,675	NA	-0,245	0,262	0,619	-0,810	NA	0,116	-0,080	
	Hydrogenuhlíčitany	-0,733	0,502	NA	0,521	0,735	-0,354	0,848	NA	-0,283	-0,196	
	Vápník	0,141	0,656	NA	0,070	-0,093	0,182	-0,861	NA	0,119	-0,238	
	TCE	PCE	0,439	-0,436	0,943	-0,998	0,728	-0,995	0,891	0,368	0,884	0,329
Suma CIM		-0,416	-0,959	-0,696	0,266	0,260	0,965	-0,368	0,443	-0,715	0,713	
Suma dle Roz.		-0,417	0,977	-0,917	1,000	0,684	0,852	0,186	-0,677	-0,551	0,758	
Chloridy		-0,429	-0,989	-0,832	0,027	0,035	-0,850	0,730	-0,110	0,251	0,455	
Dusičnany		-0,326	0,958	-0,828	-0,009	0,939	0,202	0,151	-0,006	0,088	-0,289	
Sířany		0,277	0,941	-0,588	0,213	-0,873	0,208	-0,338	-0,482	0,798	-0,557	
Hydrogenuhlíčitany		0,425	-0,958	0,779	0,547	0,617	-0,501	0,442	0,781	-0,788	-0,416	
Vápník		-0,194	-0,932	0,774	-0,367	-0,927	-0,729	-0,385	0,188	-0,339	0,883	
PCE		Suma CIM	0,878	-0,197	0,766	0,283	-0,751	0,974	-0,066	-0,018	0,769	0,310
		Suma dle Roz.	0,963	0,279	0,979	0,999	-0,490	0,837	0,218	0,316	0,446	-0,473
	Chloridy	0,802	-0,475	0,902	0,009	0,628	-0,877	-0,431	-0,155	-0,004	-0,171	
	Dusičnany	0,755	0,466	0,883	-0,001	-0,658	0,149	0,125	0,172	-0,346	-0,143	
	Sířany	-0,741	0,644	0,434	0,197	0,441	0,243	0,674	0,431	-0,818	0,267	
	Hydrogenuhlíčitany	-0,926	-0,427	-0,858	0,554	-0,960	-0,527	-0,772	-0,110	0,868	-0,516	
	Vápník	0,758	-0,679	-0,684	-0,356	0,809	-0,744	0,719	0,087	0,365	-0,255	
	Suma CIM	Suma dle Roz.	-0,941	0,990	-0,792	-0,269	-0,239	-0,766	0,917	0,889	-0,093	-0,322
Chloridy		-0,731	-0,947	-0,946	0,026	0,866	0,879	0,704	0,841	0,220	-0,321	
Dusičnany		-0,466	0,901	-0,965	-0,059	-0,168	-0,044	0,699	-0,445	0,403	0,487	
Sířany		0,864	0,830	-0,591	-0,064	-0,154	-0,336	0,584	0,842	0,768	0,383	
Hydrogenuhlíčitany		0,952	-0,910	0,933	-0,143	-0,785	0,596	-0,652	-0,390	-0,842	0,853	
Vápník		-0,705	-0,804	0,821	0,134	0,346	0,751	0,701	-0,783	-0,404	-0,682	
Suma dle Roz.	Chloridy	-0,732	0,968	-0,920	-0,015	-0,050	0,499	-0,616	-0,658	0,119	-0,560	
	Dusičnany	-0,589	-0,949	-0,892	0,002	-0,644	-0,596	-0,781	0,318	-0,387	0,058	
	Sířany	0,831	-0,872	-0,412	-0,202	0,449	0,314	-0,758	-0,780	0,494	0,407	
	Hydrogenuhlíčitany	0,991	0,940	0,870	-0,547	-0,382	0,004	0,776	0,507	-0,366	-0,013	
	Vápník	-0,739	0,858	0,678	0,358	0,513	0,517	-0,766	0,634	-0,120	-0,789	
Chloridy	Dusičnany	-0,782	0,952	-0,981	0,758	-0,098	-0,254	-0,516	0,359	0,379	0,370	
	Sířany	0,653	0,959	-0,501	-0,876	0,341	0,632	-0,329	-0,642	-0,365	0,767	
	Hydrogenuhlíčitany	0,693	-0,925	0,976	0,052	0,748	-0,806	0,131	-0,114	0,232	0,274	
	Vápník	-0,819	-0,944	0,770	0,889	-0,141	-0,662	-0,132	0,968	0,543	-0,502	
Dusičnany	Sířany	0,302	-0,923	-0,529	0,707	0,886	0,804	-0,606	0,314	-0,191	-0,377	
	Hydrogenuhlíčitany	0,506	0,945	0,977	-0,091	-0,544	-0,641	0,595	-0,253	0,367	-0,605	
	Vápník	-0,640	0,935	0,792	-0,650	0,926	-0,053	-0,547	-0,182	0,151	0,374	
Sířany	Hydrogenuhlíčitany	-0,846	0,862	0,425	-0,118	0,333	0,930	0,898	0,350	0,968	-0,395	
	Vápník	0,817	0,993	0,931	0,872	-0,829	0,374	-0,804	0,627	0,687	0,306	
Hydrogenuhlíčitany	Vápník	0,703	-0,879	-0,702	0,301	0,701	-0,563	0,959	0,151	-0,667	0,363	

Na první pohled je výsledek stále velmi nepřehledný, a nejsou v něm patrné žádné pravidelnosti nebo vzory. Proto bylo pro začátek zvoleno testování shody jednotlivých bran, a to pomocí metody statistického testování hypotéz. Nejprve je nutno stanovit hypotézy:

H_0 – mezi jednotlivými branami není rozdíl

H_a – mezi jednotlivými branami je rozdíl

Dalším krokem je vybrání vhodného statistického testu. Volba závisí především na rozdělení hodnot výběrového souboru. Jak bylo zjištěno Shapiro-testem, v případě korelačních matic nebylo ve většině případů dosaženo normálního rozdělení. Z tohoto důvodu nebyl použitý klasický t-test, ale jeho neparametrická obdoba – Wilcoxonův test.

Wilcoxonův párový test se používá v případě, kdy data neodpovídají normálnímu rozdělení. Porovnává dvě měření jednoho statistického souboru a testuje hypotézu rovnosti jejich distribučních funkcí. Princip testu je založen na práci s rozdíly mezi páry měření: tyto rozdíly se seřadí od nejnižšího po nejvyšší (kromě nulových hodnot) a přiřadí se jim pořadí. Následně se testuje hypotéza, že součet kladných i záporných hodnot by měl být roven 0. V případě softwarového řešení v RStudiosu je test vykonán příkazem `wilcox.test()`.

Vzájemné testování bylo shrnuto do další matice, která vyjadřuje výsledky Wilcoxonova testu v podobě hodnoty p-value mezi jednotlivými branami. Je-li p-value větší než 0,05, pak nelze zamítnout nulovou hypotézu o podobnosti bran ve prospěch alternativní.

Tab. 5 Výsledek testování podobnosti bran

RB1	x									
RB2	0,932	x								
RB3	0,983	0,926	x							
RB4	0,945	0,828	0,956	x						
RB5	0,951	0,771	0,779	0,883	x					
RB6	0,898	0,51	0,954	0,954	0,908	x				
RB7	0,778	0,989	0,678	0,734	0,563	0,774	x			
RB8	0,772	0,811	0,837	0,754	0,466	0,779	0,811	x		
RB9	0,835	0,893	0,863	0,696	0,833	0,961	0,818	0,759	x	
RB10	0,838	0,803	0,99	0,873	0,878	0,87	0,591	0,779	0,609	x
	RB1	RB2	RB3	RB4	RB5	RB6	RB7	RB8	RB9	RB10

Testování hypotéz Wilcoxonovým testem tedy jasně nepotvrdilo rozdílnost či podobnost na branách (ať už v případě hodnot rozdílů či hodnot na vstupu), ačkoliv vizuální pohled na korelační matice indikuje rozdílnost. Byl vyzkoušen také neparametrický Kruskal-Wallisův test, který zjistí rozdíly ve skupinách dat, nicméně výsledek byl stejný jako při použití Wilcoxonova testu.

Tyto výsledky byly pravděpodobně způsobeny charakterem testovaných hodnot. Jelikož se jednalo pouze o čísla na intervalu (-1,1) s malými rozdíly na úrovni desetin a

setin, nebyly rozdíly dostatečně patrné. Závěrem je tedy zjištění, že statistické testy hypotéz jsou na zjišťování rozdílů v korelačních maticích nevhodné.

Celkové výsledky zjištěné pomocí korelací jsou velice nevyvážené a nejednoznačné. Korelační koeficienty se na jednotlivých branách dost liší, a proto o nich nelze pronést jednoznačné tvrzení. Celkově převládají silné záporné závislosti (s hodnotami vyššími než $\pm 0,5$).

Podobně nejednoznačná byla taky analýza vztahů mezi polutanty, anorganickými látkami fyzikálně-chemickými vlastnostmi. Hlavní očekávané vztahy byly především v podobě kladné závislosti mezi oxidačně-redukčním potenciálem a chlorovanými uhlovodíky, a záporné závislosti mezi pH a chlorovanými uhlovodíky. Tato tvrzení taktéž nebylo možné u všech bran potvrdit, protože hodnoty značně kolísaly. Jak už bylo ale zmíněno dříve, u těchto korelací bylo velmi malé množství měření, takže výpočet není dostatečně přesný.

5.3 Změny na reakčních branách

5.3.1 Grafy závislostí látek

Další sledovanou skupinou dat byly naměřené koncentrace anorganických látek ve vzorcích vody. Na základě předchozí konzultace s firmy Aquatest a.s., byly vybrány ke sledování tyto látky: chloridy, dusičnany, sírany, hydrogenuhličitany, vápník a železo.

Ke zobrazení změn koncentrací látek byla vybrána grafická vizualizace v podobě bodového grafu. Jak konstatují Plašil a Vlach (2007), grafy sice nejsou žádné exaktní vědecké nástroje, avšak jsou velice efektivní v odhalování základních vzorů v datech a jejich chování. V tomto případě byly na osu x vyneseny hodnoty koncentrací na vstupu a na osu y koncentrace na výstupu. Jednoduchou grafickou vizualizací tak bylo možno snadno zjistit, zda látek při průchodu branou spíše ubývá nebo přibývá, jestli tvoří sobě blízké shluky, a nebo jestli se jejich chování na jednotlivých branách liší.

Tab. 6 Základní statistiky anorganických látek

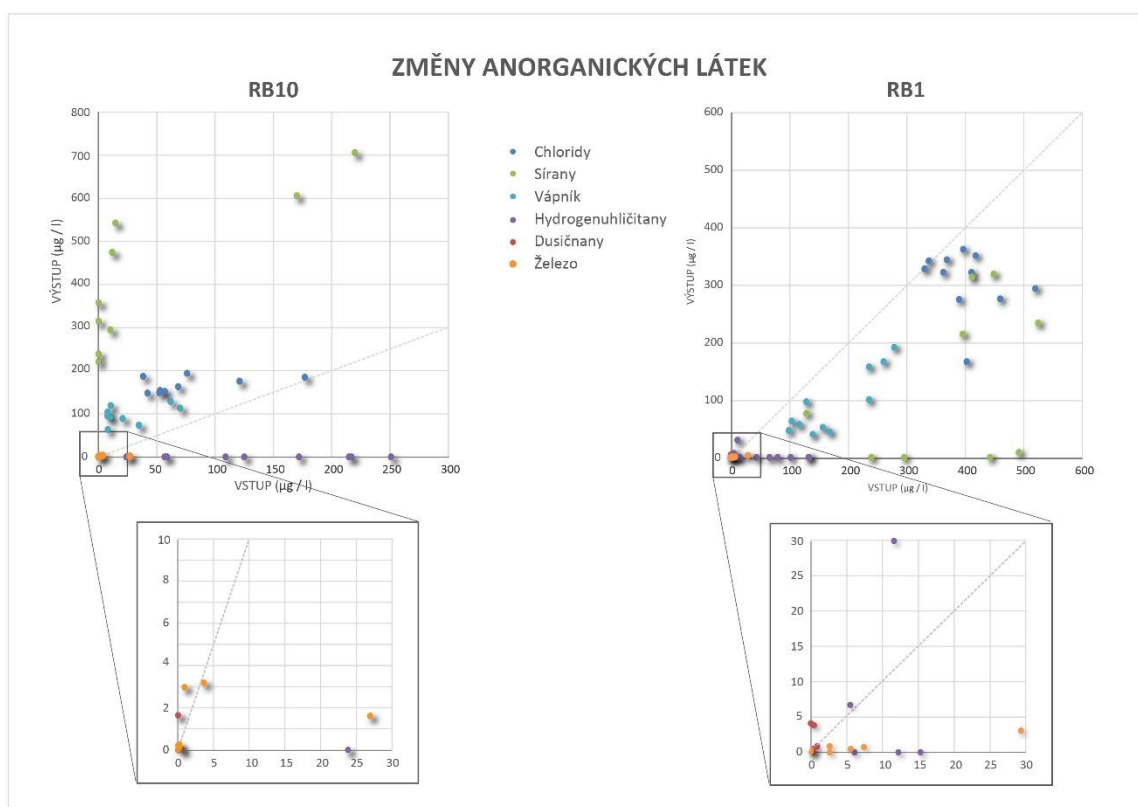
	VSTUP				VÝSTUP			
	Min	Max	Průměr	SD	Min	Max	Průměr	SD
Chloridy	36,2	1740,0	320,4	240,6	0,0	788,0	215,5	138,5
Dusičnany	0,0	24,6	1,8	3,6	0,0	33,2	2,0	4,6
Sírany	0,0	1290,0	352,1	281,3	5,0	706,0	216,9	147,9
Hydrogenuhličitany	0,0	445,0	118,1	112,8	0,0	188,0	12,5	30,0
Vápník	7,6	597,0	180,0	112,0	0,0	389,0	93,7	76,4
Železo	5,0	224,0	10,6	23,4	0,0	93,2	13,2	22,8

Byly vytvořeny 2 skupiny grafů: první zobrazuje vždy všechny anorganické látky na dané braně a druhá pouze hodnoty jedné látky ze všech bran ve všech časech. Takto se očekává zjištění případných patrných rozdílů mezi jednotlivými branami.

Už ze základních popisných charakteristik vyplývá, že celkově anorganických látek během procesu čištění ubývá. Výjimkou je železo a dusičnany, kde průměrné hodnoty

narůstají. K největší redukci dochází u hydrogenuhličitanů, kde se výstupní hodnoty nachází velice často na nule. Obecně platí pravidlo, že rozptyl v rámci jedné látky je velice široký, netvoří se žádné výrazné shluky, které by vypovídaly o sjednoceném charakteristickém chování. Rozdíly mezi jednotlivými látkami jsou také velice velké. Vápník a chloridy jako jediné nebyly v žádném ze záznamů zredukovány úplně. Podle Kvapila (2007) za redukčních podmínek, jaké panují v PTS, klesá množství látek obsahující uhlík, síru a dusík. Tyto tvrzení se nepotvrdilo na branách 1 a 8, kde docházelo k nárůstu síranů. Úbytek dusičnanů a hydrogenuhličitanů v reálu probíhal. Zajímavé bylo sledovat situaci na bráně 4, kde bylo z důvodu dřívější pilotní zkoušky nejvíce dat: mezi lety 2007-2009 zde nejprve obsah všech látek spíše narůstal, a až v posledních letech začal pokles. U chloridů se očekával nárůst koncentrací, jakožto důsledek rozkladu chlorovaných uhlovodíků, avšak kromě bran 1, 9 a některých měření na RB4 byl naopak pozorován spíš pokles řádově o stovky miligramů na litr.

Při porovnání grafu pro jednotlivé brány je lépe patrné jak se mezi sebou liší. Oproti ostatním je zde dobře patrný větší nárůst koncentrací na branách 1, 8 a 9. Na první pohled je taky poznat kolísání rozsahu hodnot od brány k bráně.



Obr. 5 Porovnání změn látek na RB1 a RB10

5.3.2 Shluková analýza

Na datech změn anorganických látek byla provedena shluková analýza za účelem identifikovat skupiny, které jsou datově podobné a zjistit, do jaké míry jsou tyto shluky totožné s jednotlivými reakčními bránami a tudíž jaký je mezi branami rozdíl.

Meloun a Militký (2004) popisují shlukovou analýzu jako metodu zabývající se mírou podobnosti objektů s velkým množstvím naměřených hodnot, které jsou klasifikovány do shluků, tedy skupin objektů, které jsou si mezi sebou maximálně podobné a zároveň se co nejvíce liší od ostatních. Dle použité metody rozlišujeme dvě hlavní skupiny, a to shlukování hierarchické a nehierarchické.

V **hierarchickém** shlukování jsou data utvářeny do stromové struktury, postupně od obecné roviny do roviny konkrétního shluku. Postupné spojování ve shluky lze vizualizovat vývojovým stromem, tzv. dendrogramem.

Hierarchické aglomerativní shlukování je na počátku tvořeno samostatnými objekty, které se postupně spojují na základě největší podobnosti s dalšími objekty a vytváří shluky. Tento proces je opakován, dokud všechny prvky netvoří jeden shluk. Oproti tomu stojí divizní shlukování, kdy je jeden velký soubor postupně rozdělován na dílčí části, dokud nejsou všechny prvky samostatné.

Nehierarchické shlukování se snaží data optimálně rozdělit podle určitých kritérií (např. s nejmenší chybou) do předem určených shluků. Sledujeme zde dva hlavní typy, a to metodu optimalizační, která hledá ideální rozklad přesouváním prvků mezi shluky, tak aby bylo dosaženo zadaných kritérií, a metodu analýzy módů, hledající rozklad do shluku, z hlediska chápání shluků jako oblasti se zvýšenou koncentrací objektů v m-rozměrném prostoru znaků (Meloun a Militký 2004).

V práci byly vyzkoušeny dvě rozdílné metody shlukování, a to hierarchické shlukování Wardovou metodou, a nehierarchická K-means metoda. Programový kód obou metod je popsán v příloze. U metod byl nastavený cílený počet shluků na 10, a ve výsledku byly tyto shluky porovnány s reakčními bránami, jestli jsou si podobné či nikoliv. Výsledky nejlépe shrnují kontingenční tabulky:

Tab. 7 Počet záznamů v jednotlivých shlucích a branách při hierarchické metodě

	shluk 1	shluk 2	shluk 3	shluk 4	shluk 5	shluk 6	shluk 7	shluk 8	shluk 9	shluk 10	Celkem
RB1		10									10
RB2		4	3	1							8
RB3		6	2		3						11
RB4	20	5			1	3	1	2			32
RB5	1						2	3			6
RB6	1	3					1	3			8
RB7								7			7
RB8	2	1		1		4					8
RB9									1	6	7
RB10	3										3
Celkem	27	29	5	2	4	7	4	15	1	6	100

Z tabulky 6 vyplývá, že na branách 1, 7, 9 a 10 byl identifikován nízký počet shluků (1 nebo 2), lze tedy tvrdit, že tyto brány jsou určitým způsobem z hlediska změn anorganických látek homogenní, jejich chování je specifické. To však ještě neznamená, že

se liší od ostatních, protože jeden shluk může zahrnovat více bran. Příslušnost více bran k jednomu shluku může znamenat jejich podobnost. Ideální stav by byl, kdyby byla jedna brána spojena s jedním shlukem. K tomu se pouze blíží brána 9. U většího počtu shluků na bránu a pak lze vyvodit nejasné a necharakteristické změny.

Další kontingenční tabulka stejným způsobem popisuje výsledky dosažené pomocí metody K-means. Je vidět, že metody poskytují velice podobný výsledek. Dle tabulek je také vidět, že obě metody mají velice podobný výsledek. Lze identifikovat 3 dvojice shluků, které jsou velmi podobné v obou metodách, a to shluky 1, 2 a 5 z hierarchické metody v páru se shluky 8, 10 a 9 z K-means metody. Nízký počet shluků spadá stejně jako u hierarchické metody na brány 1, 7, 9, 10, a navíc také 8.

Tab. 8 Počet záznamů v jednotlivých shlucích a branách při metodě K-means

	shluk 1	shluk 2	shluk 3	shluk 4	shluk 5	shluk 6	shluk 7	shluk 8	shluk 9	shluk 10	Celkem
RB1								4		6	10
RB2					3		1			4	8
RB3					1				4	6	11
RB4	3			1		3	2	19	1	3	32
RB5				2			3	1			6
RB6	1				1		3			3	8
RB7							7				7
RB8						5		3			8
RB9		6	1								7
RB10	3										3
Celkem	7	6	1	3	5	8	16	27	5	22	100

Shluková analýza ukázala, že vzniklé shluky se většinou výrazně liší od určených reakčních bran. Z toho vyplývá, že chování anorganických látek na branách je pestré a nelze je jednoznačně odlišit mezi sebou. Prolínáním mezi shluky se také identifikují brány, které mají mezi sebou určitou podobnost. Mezi shluky samotnými je taky velký rozdíl, hlavně v počtu zahrnutých pozorování. Zvolené metody se jeví jako relativně podobné.

5.3.3 Fyzikálně-chemické parametry

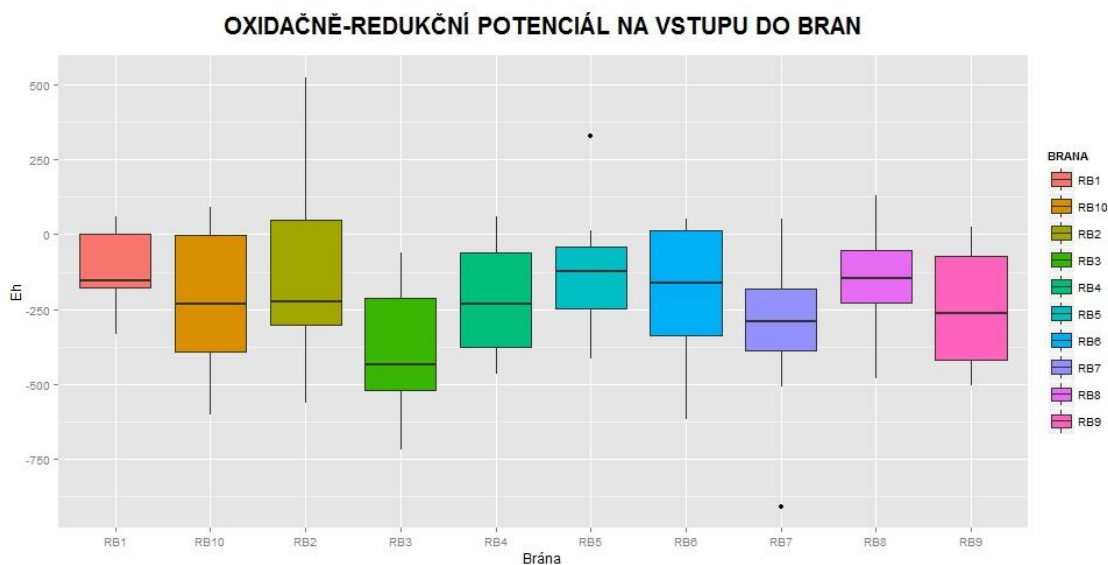
Jelikož se při korelačních výpočtech nepotvrdily očekávané tvrzení o fyzikálně-chemických parametrech, respektive výsledky byly nejednoznačné, byl proveden další rozbor těchto vlastností, a to v podobě jednouchého grafického sledování.

V dispozici bylo bohužel pouze 9 měření z let 2008 až 2014. V měření z 2013 byla nalezena chyba neznámého původu, kde pro všechny brány na vstupu i výstupu byla stejná hodnota 217 mV. Toto vadné měření bylo odstraněno.

Zpracovány byly celkem 4 datasey: hodnoty oxidačně-redukčního potenciálu, a hodnoty pH, obojí vždy pro vstup a výstup z brány. Jak je již uvedeno v kapitole 4.1, nízké hodnoty oxidačně-redukčního potenciálu vypovídají o příznivých podmínkách pro odstraňování polutantů, konkrétně by mělo být redukční prostředí navozováno při hodnotách menších než -50mV. Zároveň je očekáván růst hodnot pH (Kvapil, 2007). Na

základě těchto předpokladů byly sledovány data na vstupu a výstupu z bran, a bylo pozorováno, jestli probíhají očekávané procesy.

Nejprve byly data porovnána pomocí skupinových boxplotů. Boxplot je obecně známý grafický statistický nástroj pro vizualizaci dat pomocí kvartilů. Hlavní krabicová část je vymezená 1. a 3. kvartilem, střední linie odpovídá hodnotě mediánu. Navazující úsečky označují minimální a maximální hodnoty, ze kterých jsou však odstraněny outliers, tedy extrémní hodnoty, které nespádají do intervalu $\pm 1,5$ x mezikvartilové rozpětí. Graf tedy popisuje střední hodnotu, rozptyl hodnot, ale také informaci o symetrii a koncentraci dat. Díky tomu je efektivním nástrojem pro porovnání dat.



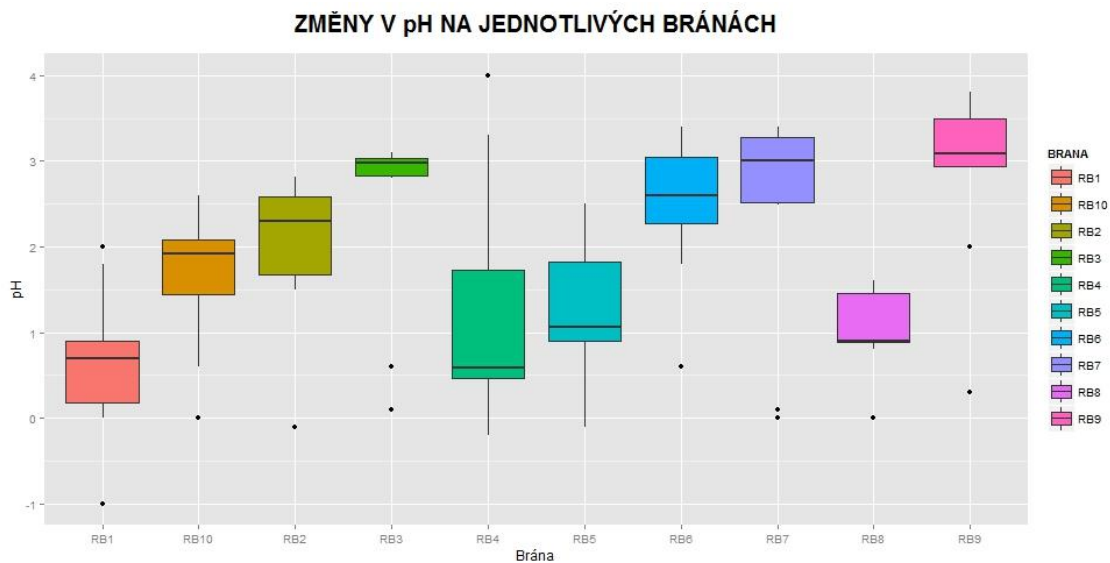
Obr. 6 Boxplot vstupních hodnot

V RStudiu byly vytvořeny boxploty pro vstup a výstup všech bran, a dále liniové grafy pro porovnání časového vývoje. Je vidět, že hodnoty Eh byly na výstupu nižší než na vstupu, potvrzuje to také tabulka rozdílů. Samozřejmě hodnoty jednotlivých bran se liší, viz boxplot, ale celkový trend je dodržen a očekávané stavy potvrzeny. Jak kolísají hodnoty v průběhu času je vidět na liniových grafech. U všech bran na vstupu i výstupu byl pozorován nárůst hodnot Eh v druhé polovině roku 2008, poté zase následný pokles. Celkový trend je stoupající, což může indikovat postupné slábnutí redukčních procesů.

Druhou sledovanou vlastností je pH (kyselost či zásaditost roztoku), u kterého se očekával nárůst na výstupu. Tato teorie byla potvrzena, byly nalezeny pouze 3 záznamy, kde hodnota pH lehce poklesla. U ostatních měření byl vždy pozorován nárůst, a to až o 4 jednotky. V datech je opět sledována velká rozdílnost mezi branami (hlavně na vstupu), při výstupních hodnotách už rozdíly nejsou tak velké. Pomocí boxplotu (obr. 7) lze sledovat změny na bránách, je patrný celkový nárůst hodnot a také rozdílná velikost tohoto růstu na jednotlivých bránách.

Zajímavé by bylo sledovat fyzikálně chemické parametry také během průchodu branou na všech 5 monitorovacích trubkách, avšak z těch jsou k dispozici data jen za jediné měření

v roce 2008, a dále údaje z karotážního měření v roce 2007. Tyto data už však byla analyzována ve zprávě o pilotní zkoušce doktora Kvapila.



Obr. 7 Boxplot změn pH

5.4 Nelineární regrese

Jeden z požadavků na analýzu dat PTS se zabývá předpovědí účinnosti stěny do budoucna. Původní efektivní období PTS bylo navrženo a předpokládáno na 10 let, tudíž se pomalu blíží ke konci. Je proto vhodné zjistit, jestli se během sledovaného období efektivita nějakým způsobem výrazně měnila, nebo jestli jsou brány stále kvalitně funkční a dá se s nimi počítat i do budoucna.

Jako predikci označujeme schopnost předpovědět následující hodnotu z určité posloupnosti dat, na základě analýzy hodnot předchozích a sestavení určitého matematického modelu. Aby bylo možno hodnoty predikovat do budoucna, je nutné nejprve nalézt vztah mezi existujícími daty, a to mezi závisle a nezávisle proměnnou. K tomu nejlépe slouží odvětví statistiky označované jako regresní analýzy. Pomocí té odhadujeme hodnoty určité náhodné veličiny (závisle proměnné) na základě znalosti jiných veličin (nezávisle proměnných, regresorů).

V této kapitole jsou použity dva druhy regrese pro rozbor dat. Nejprve je zkoumán celý datový celek pomocí beta regrese a jsou hledány ty atributy (nezávisle proměnné), které mají největší význam na sledovaný jev (závisle proměnnou). Dále je využito logistické regrese k předpovědi vývoje stavu jevu v budoucnosti. Pro data, která se nachází na uzavřeném intervalu nebo mají alternativní rozdělení (0/1) jsou metody nelineární regrese vhodnější, a jejich křivky, ne přímky, často data prokládají mnohem lépe a přesněji.

5.4.1 Beta regrese

Při zkoumání chování vícedimenzionálních dat, jako je tomu v případě záznamů z reakčních bran, je vhodné zanalyzovat, jak na sebe jednotlivé složky působí, respektive

jak působí na jeden hlavní sledovaný jev. Tím je v tomto případě úspěšnost redukci nežádoucích látek. Prvním krokem této analýzy je tedy stanovení, v jaké podobě bude tato úspěšnost vyjádřena. Jako ideální se jeví poměrové vyjádření změn na vstupu vůči výstupu, aneb kolik procent ze všech látek naměřených na vstupu bude na výstupu odstraněno. Dále by šlo tento jev kvantifikovat pouze na dichotomní hodnoty 0 a 1, například podle toho, jestli došlo k úplné redukci polutantů, nebo jestli byl / nebyl překročen stanovený limit. Procentuální vyjádření se jeví jako vhodnější kvůli své větší přesnosti než jev 0/1.

Procenta byla vypočítána pro hodnoty sumy dle Rozhodnutí, která byla stanovena dříve a její dodržování je klíčové. Tyto hodnoty závisle proměnné byly do datové sady přidány jako sloupec JEV. Celý dataset byl upraven do vhodné podoby, aby mohl být v programu RStudio použit k výpočtu. Nejprve bylo experimentováno s vytvářením modelů pro každou rekční bránu zvlášť. Tento přístup se setkal s neúspěchem, protože malý počet záznamů a charakter dat v mnoha případech pro výpočet modelu nestačil. Proto byla zvolena cesta, ve které nebyl brán ohled na identifikaci reakční brány, a všechny data byla sloučena do jednoho modelu jako celek. Tento krok je odůvodnitelný předpokladem, že chování bran z hlediska efektivity by nemělo nijak být ovlivněno jejich polohou, nýbrž pouze hodnotami a ty jsou i při sloučení zachovány.

Ferrari a Cribari-Neto (2004) doporučují beta regresi pro situace, kde jsou analyzované hodnoty spojité na intervalu (0,1). Jedním ze vstupních předpokladů je dodržení beta rozdělení. To lze zapsat v podobě následujícího vzorce:

$$f(y; \alpha, \beta) = \frac{y^{\alpha-1}(y-1)^{\beta-1}}{B(\alpha, \beta)}$$

Pro $y \in (0,1)$; α a β jsou kladné parametry. Beta funkce $B(\alpha, \beta)$ je potom popsána jako:

$$B(\alpha, \beta) = \int_0^1 y^{\alpha-1} (y-1)^{\beta-1} dy$$

Pro potřeby práce bylo využito softwarové řešení v RStudios. Pro výpočet beta regrese je nezbytný balík *betareg*, který obsahuje stejnojmennou funkci pro výpočet modelu. Touto funkcí je pak definován přepis:

$$g(\mu) = x_i' \beta$$

kde $\beta = (\beta_1, \dots, \beta_k)'$ je vektor neznámých regresních parametrů, které zjišťujeme a $x_i = (x_{i1}, \dots, x_{ik})'$ je vektor regresorů. Za g lze dosadit různé funkce, například logistickou.

Samotný zápis v RStudios je jednoduchý. Nejprve je nutno upravit data do přijatelné podoby, což spočívá především v odstranění chybějících záznamů, se kterými model nemůže být spočítán. Byly vytvořeny celkově dva modely: jeden zahrnoval všechny sledované atributy, a druhý byl zaměřen pouze na polutanty bez anorganických látek. Ukázkový zápis v RStudios pro celkový model je představen zde:

```
> model.reg <- betareg(UCIN ~ VC + HPV + X1_1_D + X1_2_C + X1_2_T + TCE
+ PCE + SUM + SUM_CL + SUM_R + CL + N + S + HCO + CA +FE)
> summary(model.reg)
```

Výsledkem tohoto výpočtu je vyhodnocení vlivu jednotlivých parametrů, tzn. jakým způsobem ovlivňují hodnotu závislého jevu, kterým je v tomto případě efektivita čištění, skryta zde pod atributem UCIN. Například má-li koeficient pro vinylchlorid (VC) hodnotu 0,0029, znamená to, že při změně hodnoty VC o jednotku se změní hodnota závislého jevu, tedy účinnosti, 0,0029krát.

Důležitým ukazatelem je výsledné p-value každého atributu, které popisuje, jaká je jeho významnost v rámci celého modelu. Jako významné jsou brány hodnoty s p-value menší než 0,05. Takto označeny byly vinylchlorid, suma polutantů, suma chlorovaných methylenů, hydrogenuhličitany, a již s menší hodnotou významnosti také vápník.

V druhém modelu, který nezahrnoval anorganické látky, byla zjištěna výrazná významnost stejných atributů, tedy vinylchlorid, suma polutantů a suma chlorovaných methylenů. Z tohoto zjištění vyplývá, že vliv anorganických látek na celkovou účinnost čištění je zanedbatelný. Jejich změny jsou tedy spíše důsledkem procesu čištění a nikoliv příčina změn v účinnosti. Tím se potvrdily předpoklady, že anorganické látky se sice během redukčních procesů mění, ale výrazně je neovlivňují. Zajímavým zjištěním byl fakt, že se neprokázal signifikantní vliv hladiny podzemní vody (HPV) na úspěšnost čištění. Tento parametr byl dříve diskutován a očekávala se jeho významnost, avšak regresní model tuto hypotézu zamítl. Z tohoto důvodu nebyla hladina podzemní vody v práci dále výrazně zkoumána.

Na základě výsledků lze celý model popsat následujícím vztahem:

$$\begin{aligned} \text{Úspěšnost čistícího procesu} = & 1,375 + 0,002 * \text{vinylchlorid} - 0,037 * 1,2 \text{ dichlorethylen} \\ & + 0,01 * 1,2\text{-cis-dichlorethylen} - 0,046 * 1,2\text{-trans-dichlorethylen} + 0,009 * \\ & \text{trichlorethylen} + 0,007 * \text{perchlorethylen} - 0,002 * \text{suma CIU} + 0,007 * \text{suma} \\ & \text{chlorovaných methylenů} - 0,008 * \text{suma dle Rozhodnutí} - 0,001 * \text{chloridy} + 0,013 * \\ & \text{dusičnany} + 0,001 * \text{sírany} - 0,002 * \text{hydrogenuhličitany} + 0,003 * \text{vápník} - 0,004 * \\ & \text{železo} \end{aligned}$$

V případě odstranění anorganických látek má vztah tuto podobu:

$$\begin{aligned} \text{Úspěšnost čistícího procesu} = & 1,335 + 0,003 * \text{vinylchlorid} + 0,170 * 1,2 \\ & \text{dichlorethylen} + 0,012 * 1,2\text{-cis-dichlorethylen} - 0,107 * 1,2\text{-trans-dichlorethylen} + 0,011 \\ & * \text{trichlorethylen} + 0,008 * \text{perchlorethylen} - 0,002 * \text{suma CIU} + 0,007 * \text{suma} \\ & \text{chlorovaných methylenů} - 0,009 * \text{suma dle Rozhodnutí} \end{aligned}$$

Z obou zápisů je patrné, že modely si jsou velmi podobné, více se liší jen v případě 1.2 dichlorethylenu a 1,2-trans-dichlorethylenu, a to v řádech desetin.

5.4.2 Logistická regrese

Další použitou metodou je logistická regrese. Ta je stejně jako beta regrese vhodná pro modelování závislostí, jejichž hodnota se nachází na intervalu (0,1). Tento typ byl zvolen hlavně z důvodu průběhu logistické křivky, která dobře kopíruje zkoumaná data a je mnohem vhodnější a přesnější než běžný lineární model.

Tento typ regrese byl navržen jako alternativa pro případy, kdy vysvětlovaná proměnná nabývá binárních hodnot (Melou, Militký, 2004). Stejně jako u ostatních typů regresí,

sledovaná proměnná je vyjadřována jako vztah s nezávisle proměnnými, rozdíl je pouze v hodnotách závisle proměnné, kterých může nabývat.

Hosmer a Lewenshow (200) definují vztah mezi závislou a nezávislou proměnnou vycházející z obecného přepisu pro regresi, využívající ovšem tzv. logistickou transformaci.

$$\beta_0 + x \cdot \beta = \log \frac{p(x)}{1 - p(x)}$$

Logistická transformace je definována jako porovnání pravděpodobnosti jevu, který nastal vůči jevu, který nenastal. Tato transformace má mnoho pozitivních vlastností lineárních modelů. Zápis samotného logistického modelu potom vypadá následovně:

$$p(x) = \frac{e^{\beta_0 + x\beta}}{1 + e^{\beta_0 + x\beta}}$$

Stejně jako u beta regrese bylo nutno data nejprve upravit, aby byla vhodná pro výpočty v RStudio. První krok bylo odstranění chybějících hodnot. Dále bylo nutno přidat do dat informace o závisle proměnné, kterou byla opět procentuální účinnost čištění polutantů. Nejprve bylo experimentováno pouze s hodnotami 0/1. Hodnota 1 byla zvolena v časech, kdy účinnost čištění byla vyšší než 95 % (jev vyčištění nastal), hodnota 0 zvolena v časech efektivitu nižší než 95 % (vyčištění nenastalo). Pokusy s tímto přístupem však nepřinesly ideální výsledky, neboť rozdělením pouze na 2 hodnoty byly data velmi zkresleny. Proto bylo nově pracováno s konkrétní hodnotou účinnosti vyjádřenou přímo procentuálně. Tato data byla mnohem přesnější a výsledná křivka mnohem lépe vystihovala skutečná data.

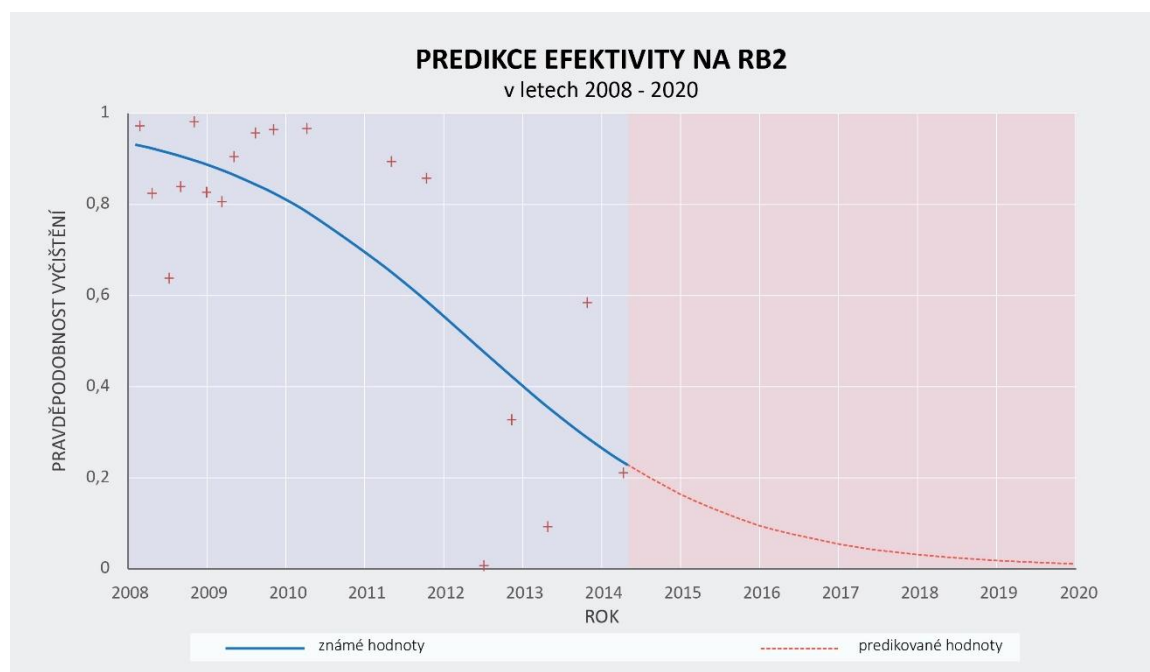
Hlavní problém v úpravě dat ležel na určení času. Jelikož měla být regrese použita pro predikci do budoucna a sledování časového vývoje, bylo na rozdíl od modelu beta regrese nezbytné čas v datech zahrnout. Při prvotním importu dat do RStudio bylo zjištěno, že v datum v běžném formátu, v jakém je používáno např. v Microsoft Excel, není pro RStudio schůdné. Jako nejjednodušší řešení tohoto problém se ukázalo převést datum již při zpracování dat v Excel z datového formátu Datum na Číslo. Tím bylo datum zaznamenáno v podobě celých čísel, čímž se sice ztratila přehlednost, avšak tento problém byl později při upravování grafických výstupů odstraněn.

Takto připravená data pro každou reakční bránu byla naimportována do RStudio ve formátu .csv, a mohl být zahájen výpočet logistických regresí. V prostředí RStudio byla použito příkazu *glm*, který slouží pro tvorbu zobecněného lineárního modelu. GLM umožňuje na základě *n* historických měření předpovídat systematickou složku do budoucna pomocí zvolených vysvětlujících veličin, a zároveň při tom zachovat náhodnost sledovaného děje. Funkce *glm* v RStudio zahrnuje běžně používané modely, a nastavením parametru *family = family=quasibinomial(logit)*, dostaneme z modelu logistickou regresi (www.data.princeton.edu).

Celkový postup byl trochu složitější než u beta regrese, bylo totiž nutno nejen vytvořit logistickou regresi na měřených hodnotách, ale vypočítat také predikci za hranici známých hodnot. Jako nejzazší bod, do kterého byla prováděna predikce, byl stanoven 1. leden 2020.

Predikce do vzdálenější budoucnosti se nejevila jako důležitá, protože takový odhad už by byl velmi přibližný a nevyovídající.

Nejprve byl tedy spočítán regresní model pomocí příkazu *gml*. Aby bylo možno vykreslit závislost účinnosti na datu, do modelu vstoupily pouze tyto dva prvky. Dále byly připraven vektor dat, pro které měla být spočítána predikce. Tento vektor zahrnoval jednak původní časové úseky, ve kterých probíhalo měření, a dále časovou posloupnost od 1. 1. 2015 do 1. 1. 2020, vždy s půlročním intervalem, aby byla zaručena rovnoměrnost v datech pro vykreslení křivky. Pomocí příkazu *predict* byla v RStudiosu spočítána pravděpodobnost úspěchu čištění pro tento nový vektor dat. Podstata grafické vyjádření pak spočívala v překrytí dvou barevně odlišených logistických křivek přes sebe – spodní vrstva znázorňovala predikované hodnoty, a přes ni byly překresleny známé hodnoty (avšak jen do posledního známého měření v březnu 2014). Kompletní programový kód je v příloze diplomové práce.



Obr. 8 Predikce pomocí logistické regrese

Jelikož vyplotovaný výsledek z RStudia není graficky příliš uspokojivý, byl tento export vložen do Adobe Illustrator a zde dále upraven a překreslen pro dosažení lepších grafických výsledků (viz. obrázek 8).

Ve výsledku bylo vytvořeno 10 takových vizualizací pro všechny brány. Analýza pomocí logistické regrese přinesla zajímavé výsledky. Dalo by se očekávat, že účinnost systému reakčních bran bude postupně klesat, avšak dle této analýzy tomu tak vždy není. Celkově lze rozdělit výsledky do tří skupin: u pěti bran je patrný v průběžný pokles účinnosti – jedná se brány 2, 4, 5, 7 a 10., přičemž nejprudší pokles je viditelný na bráně 2. Rostoucí účinnost můžeme sledovat u bran 1, 8, 9 a 10, obzvláště u 1 se v posledních pěti letech drží účinnost na velmi vysoké úrovni a neklesla pod 95 %. Na bráně 10 hodnoty kolísají, celkový trend je sice rostoucí, ale ne příliš důrazně. Zvláštní skupinu tvoří brány

3 a 6: průběh výsledné křivky je sice mírně rostoucí, ale je patrné že tento trend způsobuje pár extrémních (nízkých) hodnot, u brány 6 v roce 2009, a u brány 3 je těchto hodnot více především mezi lety 2008 – 2009, a potom dále v letech 2011 a 2013. Kdyby byly tyto outliery odstraněny, trend by se v obou případech změnil na velmi pozvolna klesající. Avšak z důvodu ne příliš velkého počtu měření nebyl tento krok proveden.

Pomocí logistické regrese byly tedy předpověděny a zhodnoceny trendy na reakčních branách. Je však nutné si uvědomit, že tyto predikce jsou pouze orientační a založené na co nejvěrnějším proložení skutečně naměřených hodnot. Nelze je tedy brát za jednoznačně spolehlivé, realita může být ovlivněna spoustou dalších vlivů, které regrese nedokáže zohlednit.

6 MRAK M5

Mrak 5 (část A) je druhou řešenu lokalitou v rámci práce. Bylo zde použito méně hodnotících metod než u PTS, a jejich výběr byl díky charakteru dat jiný. Opět bylo využito regresních modelů za účelem predikce vývoje do budoucna, a dále byly více zkoumány prostorové aspekty znečištění. Pomocí vhodných metod lze zkoumat, jak se v čase a prostoru koncentrace polutantů vyvíjí.

6.1 Regresní model

Stejně jako u PTS, i v oblasti mraku 5 byla poptávka po odhadu vývoje do budoucna. K dispozici ze byla data z jednotlivých vrtů, a nebo koncentrace polutantů získané jako sumy ze všech čerpaných vrtů, které vstupují do čistícího systému a následně jsou zasakovány zpět. Provádět predikci na každém vrtu zvlášť by nebylo ideální z důvodu jednak většího počtu vrtů a hlavně kvůli časové nestejnorodosti v odběrech. Proto byla pro predikování stavu vybrána součtová data z čerpaných vrtů. Nemělo však význam pracovat zde podobně jako u PTS s hodnotami účinnosti čištění, protože v případě mraku 5 je efektivita mnohem vyšší a přes 90 % všech záznamů dodržuje stanovené limity s efektivitou vyšší než 95 %. Taková data by tedy pro model byla hodně zkreslená a v podstatě ani nemá význam něco takového vytvořit, protože předpověděné hodnoty by se nijak výrazně nelišily od těch měřených. Navíc podstata čistící technologie na mraku 5 zaručuje mnohem jistější a účinnější systém čištění než u PTS.

Proto byla pro predikci vybrána data celkových koncentrací polutantů vstupující ze všech odčerpávaných vrtů do čistícího systému, tedy ze sumy dle rozhodnutí, a pro zajímavost také z nové sumy rozhodnutí. Zde se na rozdíl od PTS jednalo o absolutní hodnoty, které se nepohybovaly na žádném stanoveném intervalu, nýbrž náhodně kolísaly a jejich rozptyl byl velký. Z tohoto důvodu nebylo možné použít logistickou či beta regresi pohybující se na uzavřených intervalech, jako tomu bylo v případě PTS. Jako vhodná alternativa se nabízí obyčejný lineární regresní model.

Lineární regrese představuje nejjednodušší z regresních modelů. Závislá a nezávislá proměnná jsou popsány funkcí s parametry, a jsou-li tyto parametry lineární, je tento typ regrese označován také za lineární. Vztah je potom popsán rovnicí:

$$y = \alpha + \beta x + \varepsilon$$

kde y je závisle proměnná, x nezávisle proměnná, α a β jsou lineární parametry a ε popisuje chybovost, tedy rozdíl mezi hodnotou predikovanou a naměřenou, tzv. reziduum. Jak uvádí Brázdil (1981), regresní přímka je určena metodou nejmenších čtverců, která má požadavek splnění takové podmínky, že její průběh musí přimykát ke známým datům tak, aby součet čtverců vzdálenosti bodů pole od přímky musí být minimální. Tato podmínka je popsána vztahem:

$$S_{rez} = \sum_{i=1}^{i=n} \varepsilon_i^2 = \sum_{i=1}^{i=n} (y_i - y'_i)^2$$

Dále je podle Brázdila (1981) vhodné konstruovat intervaly spolehlivosti, a to z důvodu, že rovnice se může při použití různých výběrů ze souboru lišit. Výpočtem intervalu spolehlivosti tedy určujeme pro vybraná x interval, ve kterém se budou hodnoty y s určitou pravděpodobností nacházet.

Praktické zpracování nabízí v poměrně jednoduché sadě příkazů RStudio. Jediná úprava vstupních dat spočívala v konverzi informace o datumu z datového formátu Datum na Číslo. V případě zachování datumu nedokáže RStudio s takovými čísly pracovat. Jestliže se místo data zobrazuje na ose X celé číslo v rádech desetitisíců, není to příliš vypovídající, avšak drobnými grafickými úpravami na výsledném obrázku lze snadno dosáhnout uspokojivého výsledku. Programový kód pro výpočet regrese je velmi snadný:

```
> lm.out <- lm(SUM ~ DAT)
> lm.out
> summary(lm.out)
> plot(lm.out)
```

Poslední příkaz `plot` vytvoří sadu grafů, hodnotící kvalitu lineárního modelu. První plot nazvaný Residuals vs. Fitted hodnotí rozdíly mezi skutečnými hodnotami a hodnotami vypočítanými modelem. Normal QQ plot popisuje, jak se data modelu blíží k normálnímu rozdělení, tedy stavu ideálnímu pro modelování. Plot Residuals vs. Leverage vykresluje hodnoty, které mohou mít na celkovou kvalitu modelu negativní vliv. Tyto hodnoty jsou jakýmiś outliersy a je doporučováno je odstranit (www.wv2.coastal.edu). V tomto případě byly takto označeny 3 hodnoty.

Celý lineární model pro obě dvě sumy byl tedy vypočten ještě jednou, a tentokrát bez výše zmíněných označených outlierů. Z průběžných hodnotících plotů bylo pozorováno zlepšení, stejně tak v koeficientech obsažených ve shrnutí modelu příkazem `summary()`. Vzhledem k charakteru dat a jejich velkému rozsahu a kolísavosti je jasné, že lineární regrese bude pouze hrubým přiblížením skutečnosti, avšak využití navržených postupů lze dosáhnout přesnějších výsledků.

Z těchto modelů lze tedy odvodit rovnici pro křivku popisující celkový trend, a touto cestou pak spočítat hodnotu, kdy by mělo dojít k celkovému poklesu koncentrací polutantů pod stanovené limity:

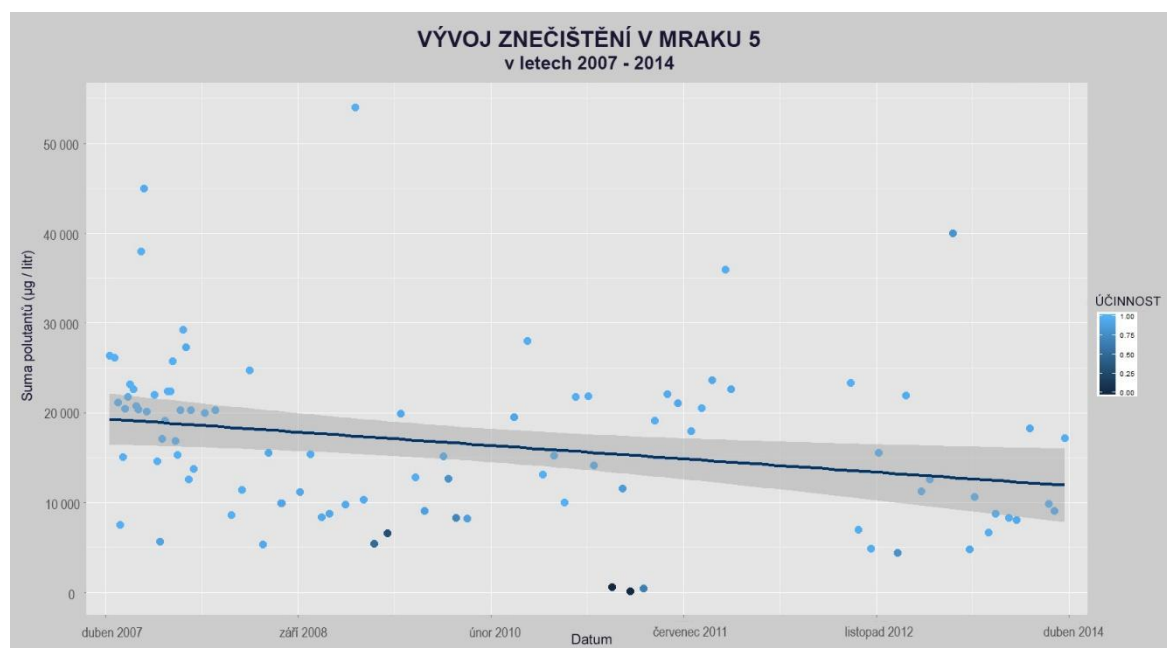
$$\text{Suma dle Rozhodnutí} = -3,22 * \text{datum} + 141\,710,95$$

$$\text{Nová suma dle Rozhodnutí} = -2,92 * \text{datum} + 133\,931,13$$

Koeficienty rovnice vypadají na první pohled monstrózně, je to však dáno převedením datumu na celé číslo, které je vysoké. Tyto rovnice je už velice snadné upravit tak, aby bylo odhadnuto datum, kdy dojde k dosažení stanovených limitů pro znečištění. Pro sumu dle Rozhodnutí by mohlo být dosaženo limitu 1 550 $\mu\text{g/l}$ v březnu 2019, pro novou sumu a limit 2 000 $\mu\text{g/l}$ v září 2023. Je tedy patrné, že pokles u nové sumy je pozvolnější. Tyto predikce je však nutno brát s velkou rezervou, protože lineární regrese nevystihuje vývoj přesně, nýbrž se jen snaží dosáhnout optimálního výsledku pro vstupní data, a v žádném případě není schopná zahrnout nečekané vlivy a změny v chování celého procesu.

Na závěr byla také lineární regrese vizualizována v podobě křivky proložené daty. Jednotlivé záznamy byly barevně odlišeny dle zaznamenané efektivity. Je tak vidět, že až

na pár výjimek byla účinnost čištění v mraku 5 velmi vysoká. Graf byl vytvořen v RStudiu pomocí balíku ggplot2, který umožňuje výrazně lepší a atraktivnější grafické výstupy než ostatní základní balíčky, zahrnuta je také vizualizace intervalu spolehlivosti na hodnotě 95 %. Programový kód k takovému výstupu je popsán v příloze.



Obr. 9 Lineární regrese pro sumu 2 000 µg/l

6.2 Prostorové hodnocení

Všechny analýzy probíhající na PTS měly spíše statistický charakter, a nebyl kladen tak velký důraz na jejich prostorovou složku. Oproti tomu, v mraku 5 se díky vhodnému plošnému rozmístění vrtů nabízí důkladnější prostorová analýza.

K dispozici byla data z vrtů rozmístěných po celé oblasti. Na každém vrtu byly sledovány stejné informace jako ve všech předchozích částech. Pro hodnocení jsou důležité především hodnoty sumy chlorovaných uhlovodíků dle Rozhodnutí, jejichž prostorové rozložení bylo zkoumáno. Na počátku všeho stálo opět zpracování dat do vhodné podoby, kterou by bylo možno využít ke prostorovým analýzám. Prostorová složka vrtu byla zajištěna souřadnicemi X, Y v systému S-JTSK. Časově byla data sbírána od roku 2007 do roku 2014. Problém byl však s časoprostorovou kompaktností – na území bylo sice umístěno dostatečné množství vrtů, ale odběry a rozbory vody z nich neprobíhaly vždy stejně. Ve výsledku byly měření rozprostřeny do celého roku, a bylo náročné vybrat dostatek vhodně prostorově rozmístěných vrtů, které by měly záznamy ze stejného období. Cílem byl získat co nejvíce časových záznamů, aby mohl být vývoj sledován podrobně. Po rozboru všech dat bylo nakonec vytvořeno celkem 9 datasetů, přibližně s ročním krokem (pouze v roce 2009 se podařilo sestavit dva datasety, a to pro duben a říjen). Z atributové složky byla vybrána suma dle Rozhodnutí, jejíž limit 1 550 µg/l je přísnější a tudíž je vhodnější sledovat tyto nižší hodnoty.

6.2.1 Popisné charakteristiky

Na takto vybraných datech byly nejprve spočítány základní popisné statistiky, aby byl k dispozici základní přehled o datech. Brázdil popisuje dva základní druhy popisných statistik, a těmi jsou charakteristiky polohy (různé druhy průměru, modus, medián) a variability (variační rozpětí, směrodatná odchylka, rozptyl). Dále pak budou ještě použity charakteristiky asymetrie, kterými jsou koeficient šikmosti a špičatosti.

Aritmetický průměr popisuje Brázdil (1981) jako úhrn hodnot kvantitativního znaku dělený rozsahem souboru. Slabina spočívá ve velkém ovlivnění extrémními hodnotami.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Nejjednodušší charakteristikou variability je variační rozpětí, které je rozdílem nejvyšší a nejnižší hodnoty statistického souboru. Tato charakteristika je stejně jako průměr výrazně zkreslována extrémními hodnotami, jejichž původ může být v datovém souboru nahodilý či chybný (Brázdil, 1981).

$$R = x_{max} - x_{min}$$

Směrodatná odchylka je mírou proměnlivosti hodnot kolem průměru. Pokud jsou data nenulová, má vždy hodnotu > 0 .

$$s_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

Jako poslední byly počítány charakteristiky asymetrie, které charakterizují nesouměrnost v rozdělení hodnot. Souměrné rozdělení by mělo koeficient šikmosti roven nule.

$$\gamma = \frac{\sum_{i=1}^k n_i (x_i - \bar{x})^3}{ns^3}$$

kde s je směrodatná odchylka. Je-li $\gamma > 0$, jedná se o sešikmenost doleva (kladná šikmost), $\gamma < 0$ značí sešikmenost do prava (záporná šikmost).

Podle Brázdila (1981) značí koeficient špičatosti koncentraci prvků souboru kolem určité hodnoty. Poskytují představu o tom, jak je rozdělení špičaté či zploštělé. Vyšší hodnoty koeficientu značí větší koncentraci prvků kolem hodnot (průměru, střední hodnoty).

$$\varepsilon = \frac{\sum_{i=1}^k n_i (x_i - \bar{x})^4}{ns^4} - 3$$

Při kladném ε je rozložení více špičaté, při záporném zase zploštělé.

Přehled všech zjištěných charakteristik je uveden v tabulce. Z tohoto prvního náhledu je patrné, že v datech jsou velké rozdíly jak už mezi jednotlivými lety, tak v rámci rozptylu hodnot v jednotlivých letech.

Tab. 9 Základní statistiky vrtů

	2007	2008	2009	2009_2	2010	2011	2012	2013	2014
POČET	24	24	24	24	28	14	21	36	37
PRŮMĚR	7 353	8 043	10 640	10 783	9 877	29 315	3 811	4 156	7 113
MIN	12	82	59	163	49	343,5	167,7	66,5	124,8
MAX	34 287	72 391	70 376	48 428	51 267	160 608	48 364	23 385	99 374
SD	8 976	14 368	19 471	12 976	14 552	44 638	10 089	5 953	17 333
VAR ROZPĚTÍ	34 276	72 309	70 317	48 265	51 218	160 265	48 196	23 319	99 249
ŠIKMOST	1,47	4,02	2,18	1,52	1,73	2,17	4,40	1,90	4,40
ŠPIČATOST	1,76	17,55	3,88	1,98	1,86	4,61	19,80	2,91	21,99

6.2.2 Kriging

Digitální reprezentace vrtů je uložena v podobě bodové vrstvy, kde jsou body určeny diskrétně souřadnicemi X a Y . Hodnota v tomto bodě je měřená a známá. Počet vrtů je však značně omezený a jejich prostorové rozmístění není pravidelné. Pro lepší představu o prostorovém výskytu jevu je nutné znát hodnotu jevu v každém místě, a k tomu je nutné najít metodu, která převede diskrétní bodovou vrstvu na základě jejich naměřených hodnot do jednoho spojitého povrchu. Touto metodou je prostorová interpolace.

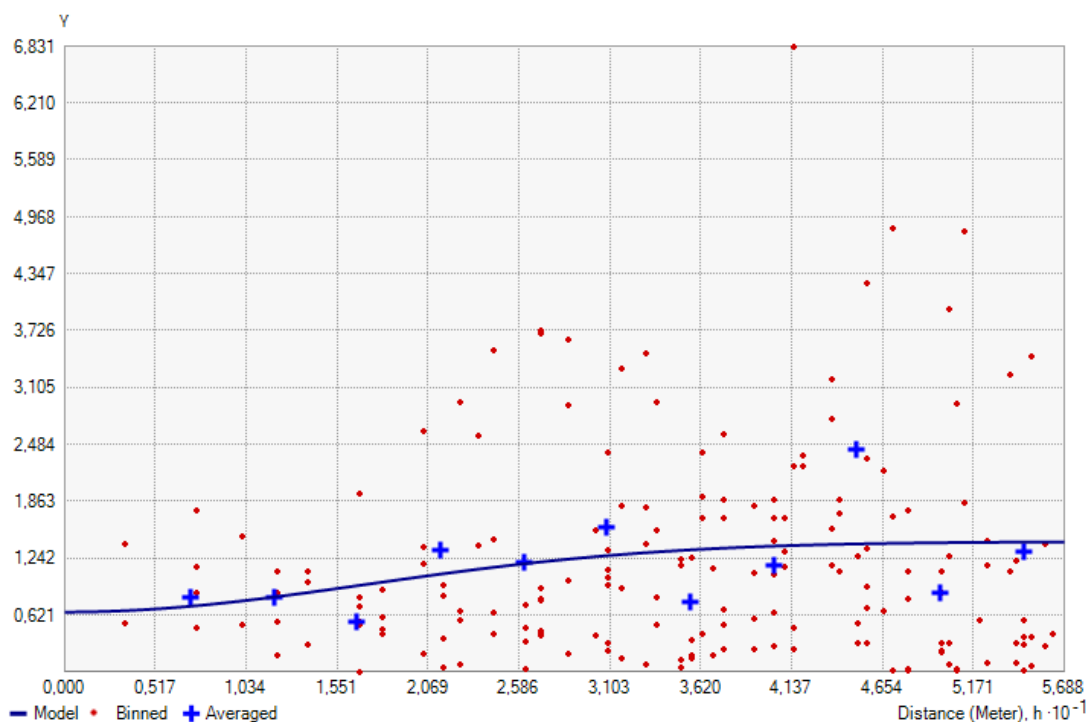
Jako interpolace je označován proces, při kterém je odvozována hodnota určitého jevu na základě okolních známých hodnot (Burian, 2008). Interpolačních metod je více, založených na vážených vzdálenostech, trendech povrchu nebo splínech. V této práci bude použita metoda kriging, která spadá do geostatických interpolačních metod.

Za otce krigingu je považován Daniel G. Krige, který se snažil odhadnout rozložení zlata na základě několika známých hodnot zjištěných z vrtů. U krigingu se předpokládá, že vzdálenost a směr mezi známými body určitým způsobem odráží jejich prostorovou závislost, která může být použita k vysvětlení proměnlivosti modelovaného jevu. Očekává se, že blízké body jsou si podobnější než body vzdálené.

Prvním krokem je vyjádření prostorové struktury a závislostí pomocí semivariogramu. Existuje určitá vzájemná závislost mezi hodnotami sledovaného jevu, závislá na vzdálenosti pozorování, tuto závislost nazýváme jako autokorelaci (Li, 2005).

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2$$

kde $\gamma(h)$ je hodnota semivariogramu, $Z(x_i)$ je měřená hodnota jevu a $N(h)$ počet párů bodů ve vzdálenosti h . Semivariogram předpokládá normální rozdělení v datech. Pokud není tato podmínka dodržena, je nutno data určitým způsobem transformovat.



Obr. 10 Ukázka semivariogramu

Z obrázku je patrné, že se rostoucí vzdáleností dvou bodů klesá jejich vzájemná autokorelace a zvyšuje se jejich nepodobnost. Semivariogram má několik parametrů – **range** indikuje maximální vzdálenost, na které je patrná autokorelace v datech, tudíž že se hodnoty mezi sebou ovlivňují. V určité vzdálenosti dosáhne variogram mezní hodnoty (**sill**), za níž se už hodnoty dále nemění. Posledním parametrem je **nugget**, který popisuje velikost posunutí semivariogramu od jeho počátku. Při vyšších hodnotách nuggetu je sledována nepravidelnost na blízkých vzdálenostech.

Dosah semivariogramu nemusí být ve všech směrech stejný. V případě, že hodnoty variogramu a jeho parametry se v různých směrech liší, hovoříme o tzv. anizotropii. Jsou-li dosahy variogramů stejného či příbuzného typu shodné, je těleso izotropní. Lze-li tyto dosahy proložit elipsou, jedná se o geometrickou anizotropii. Při rozdílných dosazích v různých směrech se jedná o zonální anizotropii (Zůvala, 2011). Anizotropie lépe zohledňuje přírodní charakter jevu, například v tomto případě analýzy znečištění může zohledňovat proudění podzemní vody (www.help.arcgis.com).

Následně je na empirický semivariogram nasazen model, který jej nejlépe vystihuje, a pomocí něj jsou predikovány neznámé hodnoty v prostoru. Jedná se o klíčový krok celé predikce. Pomocí této funkce je možné počítat neznámé hodnoty ve všech směrech a vzdálenostech. V podstatě se jedná o úlohu podobnou regresní analýze – datovým souborem musíme proložit určitou funkci, která jej nejlépe vystihuje. Modelů semivariogramu je několik, liší se podle použité funkce. Mezi hlavní typy patří:

- sférický
- exponenciální

- gaussovský
- cirkulární
- stabilní

Po výběru nejvhodnějšího modelu semivariogramu následuje již krok predikce neznámých hodnot pomocí krigingu. Stejně jako u ostatních interpolačních metod, jsou k výpočtu použity vážené hodnoty nejbližších známých bodů, avšak algoritmus výpočtu je zde dokonalejší, založený na využití semivariogramu a anizotropického chování jevu.

K dispozici je několik hlavních metod: běžný (ordinary), jednoduchý (simple), univerzální (universal) a nebo kokriging. V této práci byl použit pouze běžný kriging, který je obecně vhodný pro široké využití. Předpokládá, že prostorová proměnná je stacionární a průměr je neznámý.

Model ordinálního krigingu lze popsat vztahem (Hengl,2007):

$$Z(s) = \mu + \varepsilon'(s)$$

μ je stacionární funkce (průměr) a $\varepsilon'(s)$ zastupuje prostorově korelovanou část proměnlivosti jevu. Výpočet predikované hodnoty potom proběhne pomocí zápisu ve tvaru (www.help.arcgis.com):

$$Z(s_0) = \sum_{i=1}^n \lambda_i Z(s_i)$$

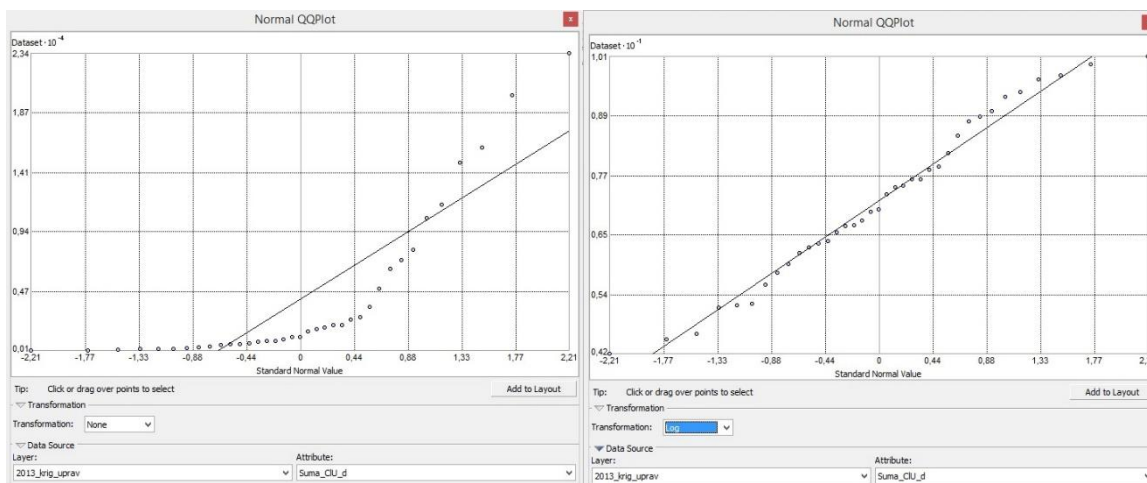
kde $Z(s_i)$ je měřená hodnota na i -lokaci, λ_i je váha a s_0 predikovaná hodnota.

Pro potřeby diplomové práce byl proces krigingu řešen v programu ArcMap, který disponuje extenzí Geostatistical analyst. Ta je zaměřená na využití geostatistických metod, v tomto případě především krigingu. Pomocí extenze lze nejprve prozkoumat data z hlediska jejich rozdělení, výskytu trendů a vizualizovat autokorelaci, následně pomocí různých typů a nastavení krigingu provést výpočet predikce neznámých hodnot v prostoru a závěrem zhodnotit kvalitu a přesnost vytvořeného modelu.

6.2.3 Prostorová analýza vrtů

Jak již bylo zmíněno výše, data z mraku M5A byla upravena do 9 datasetů, na kterých měla být provedena interpolace pomocí nástrojů z *Geostatistical analyst*. Už při prvotním náhledu na data pomocí základních popisných statistik bylo jasné, že extrémní hodnoty budou způsobovat problémy při interpolaci a zhoršovat celkovou přesnost. Tato hypotéza se potvrdila při prvních zkušebních modelech. Z tohoto důvodu byly z dat outliery odebrány – byl odstraněn nejvyšší záznam, který výrazně vynikal nad ostatními. Bylo tak učiněno ve všech letech, kromě 2007, říjen 2009 a rok 2013.

Geostatistical analyst nabízí intuitivní uživatelské rozhraní, které uživatele provede jednotlivými kroky interpolace, kde lze využívat buď výchozí nastavení a vše provést automaticky, nebo do procesu vnést větší či menší množství vlastního nastavení. Nejprve bylo provedeno testování normality dat. Ta nebyla v žádném z případů dodržena, a proto bylo nutno data transformovat. Nejlepší výsledky poskytovala logaritmická transformace, ukázka rozdílu mezi původními a transformovanými daty je na obrázku 11.



Obr. 11 Transformace dat

Po základním průzkumu dat byl pomocí možnosti *Geostatistical Wizard* spuštěn samotný průvodce interpolací. Zvolená metoda byl kriging, bylo nutno zvolit správnou vrstvu a atribut hodnot, které mají být interpolovány, tedy suma dle Rozhodnutí. V dalším kroku byla vybrána metoda Ordinary kriging a nastavení typu transformace na logaritmickou. Následující krok už přinesl okno s vizualizací semivariogramu a volby jeho nastavení. Celkový přehled použitých empirických semivariogramů a jejich hodnot je uveden v tabulce 7. V posledním okně je k dispozici přehled chybovosti v podobě střední kvadratické chyby, absolutní chyby a dalších ukazatelů.

Tab. 10 Nastavení semivariogramů pro jednotlivé roky

PARAMETRY SEMIVARIOGRAMU							
	model	anizotropie	nugett	sill	min. range	max. range	RMSE
2007	sférický	ano	0	5	24,9	59,4	12216
2008	sférický	ano	0,2	2,94	79,6	123,3	5190
2009	sférický	ano	0	4,4	38,2	66,2	7868
2009_2	sférický	ano	0	3,8	36,5	69,5	10924
2010	stable	ano	0,9	4,3	47,1	64,6	9275
2011	sférický	ano	0,8	3,5	40,8	45,9	31751
2012	sférický	ne	0,4	0,9	x	46,7	1764
2013	sférický	ano	0,3	2,6	30,5	41,6	4274
2014	stable	ano	1,5	1,8	28,8	46,3	8567

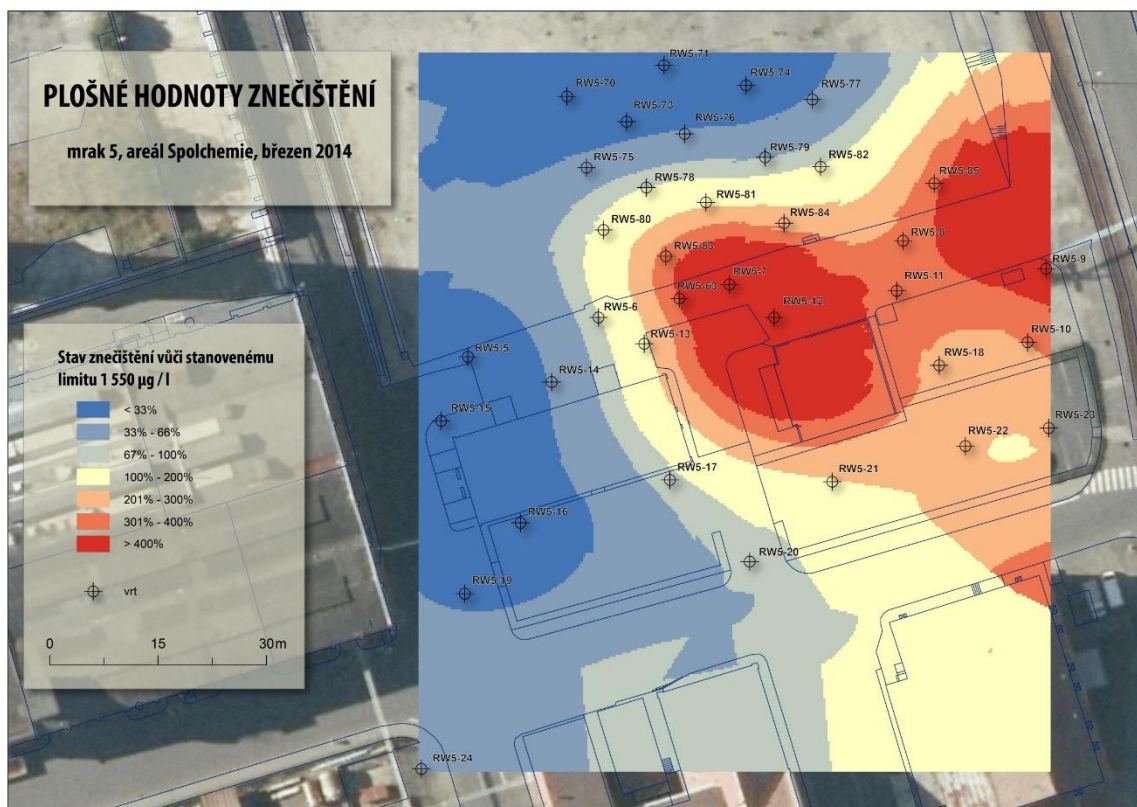
Na datech bylo testováno velké množství nastavení, především různé teoretické modely semivariogramu a využití anizotropie tak, aby bylo dosaženo co největší přesnosti. Obecně se jako nejvhodnější jevil sférický model s využitím anizotropie. Je vidět, že nastavení je proměnlivé a to z důvodu rozdílného množství vstupujících bodů do modelu a také kvůli velkým rozdílům hodnot. Za povšimnutí stojí také vysoká střední kvadratická chyba, u které vyžadujeme co nejnižší hodnoty, avšak zde byly velmi vysoké. Je to způsobeno především odchylkami u nejvyšších extrémních hodnot, kde tato chyba celkově značně narůstá. Avšak při vizuální kontrole původních dat a interpolovaného rastru bylo zjištěno,

že celkový trend byl dodržen, jen z důvodu velkého rozptylu dat je přesnost nižší. Zvoleným nastavením bylo docíleno co nejnižších chyb.

Ze všech bodových vrstev byly postupně vytvářeny souvislé plochy hodnot znečištění. Z těch byly následně vyexportovány data v rastrovém formátu .grid, s prostorovým rozlišením 0,3 m. Výsledné rastry bohužel nemají stejný rozsah pro všechny roky, protože nebyla dostupná vždy ta samá data. Tento problém je drobnou komplikací pro následné hodnocení, a tak není zájmová oblast například při rozdílových rastroch vždy stejná. Chybějící data však nelze nijak nahradit. Jako referenční rastr byl zvolen rok 2010, jehož rozsah se nejlépe shoduje s ostatními. Tímto rastrem byly zbývající ořezány za účelem dosažení stejného prostorového rozsahu pomocí nástroje geoprocessingu *Raster/Raster processing/Clip*. V případě rastrů, jejichž rozsah byl menší, byly chybějící hodnoty doplněny jako No data a to při zaškrtnutí nabídky *Maintain clipping extend*.

Takto upravené rastry bylo nutno na závěr vhodně vizualizovat a vyhodnotit. Jako první možnost přicházelo v úvahu data rozdělit do několika tříd a ty pak barevně odlišit. Vizualizovat data spojitě barevnou škálou by kvůli velkému rozsahu nebylo optimální. Nakonec byla zvolna jiná metoda, a to pomocí porovnání hodnot se stanoveným limitem 1 550 µg/l, který je planý pro sumu dle Rozhodnutí. Hodnoty byly klasifikovány do 7 kategorií za využití bipolární stupnice. Nebyly použity absolutní hodnoty, nýbrž procentuální vyjádření překročení stanoveného limitu – 200 %, 300 %, 400 % a více, hodnoty nižší než limit byly vyjádřeny jako 66 %, 33 % a méně než 33 %. Takto nastavená klasifikace potlačuje extrémně vysoké hodnoty a pro vizualizaci je vhodnější. Snadno umožňuje identifikovat oblasti, které dodržují stanovený limit, a zároveň vymezuje oblasti, které limity překračují mnohonásobně, a to i bez znalosti absolutní hodnoty, která není z důvodu výše popsaných faktorů interpolace zcela přesná.

Vizualizace proběhla pro všech 9 sledovaných období. Výsledky opět celkem kolísají, největší rozdíly jsou vidět v roce 2011, kde silně převažují plochy maximálním znečištěním, a hned v následujícím roce pozorujeme silný pokles, kde limity nejsou ani zdaleka tak výrazně překračovány. Lze však určit jedno stále jádro znečištění, a to se pohybuje přibližně ve středu zájmové oblasti, kolem vrtů RW5-63, RW5-7 a RW5-12.



Obr. 12 Vizualizace prostorové interpolace

Aby bylo možné porovnat vývoj, byly vypočítány změny mezi rastry, a to opět v procentuálním vyjádření, ze kterých bylo možno určit, o kolik procent se situace mezi dvěma lety zhoršila/zlepšila. Byla použita stejná klasifikace jako u základní vizualizace, a porovnávány byly vždy dva po sobě jdoucí datasety, a také změna každého datasetu oproti prvnímu záznamu z roku 2007. Zde přichází problém s nestejným rozsahem hodnot všech rastrů z důvodu nekompletnosti dat, a proto při výpočtu z posupně jdoucích let vzniká spousta holých míst. Ze zbývajících informací je vidět, že hodnoty opět průběžně kolísají, největší zlepšení proběhlo mezi lety 2008-2009, 2009-2010 a naposledy mezi 2013 a 2014.

Při porovnávání let s prvním měřením v roce 2007 je patrný jeden určitý prostorový vzor, který zahrnuje dvě hlavní menší centra výrazného zhoršení, a ve zbytku oblastí dochází naopak ke zlepšení. Tento vzor je pro všechny roky přibližně stejný, z čeho lze vyvodit závěr, že největší skok nastal hned po roce 2007, a od té doby hodnoty jen mírně kolísají.

6.2.4 Vizualizace v animaci

Pro rychlý přehled datových náhledů a vývoji v čase byla vytvořena animace. Jako zajímavé řešení tohoto nápadu se ukázal volně dostupný program Google Earth, ve kterém lze kombinovat časoprostorové prvky do interaktivní animace. Toolbox v ArcGIS nabízí velice snadnou konverzi rastrů do formátu kml, který je podporován v Google Earth. Takto vyexportovaný rastr se zobrazí přímo v prohlížeči programu, správně georeferencovaný.

Záměrem však bylo zobrazit všechny rastry v po sobě jdoucí sekvenci. K tomu bylo nutno upravit zdrojový kód formátu .kml, k čemuž bylo použito dokumentace na webu Google developers, kde je popsáno, jak se zdrojovým kódem formátu kml manipulovat. Úprava byla jednoduchá, spočívala v přidání dalších vrstev reprezentovaných ostatním rastry a odkazem na soubor .png, který nese obrazovou informaci a lze jej vyextrahovat z exportovaného .kml. Aby bylo možné výsledek použít jako animaci, bylo nutno přidat ještě časovou značku, a to v podobě tagu <TimeSpan>, který obsahuje dílčí tagy <begin> a <end>, které definují časový rozsah trvání daného obrazu. Datum lze zapsat ve formátu rrrr-mm-dd, v tomto případě stačilo rozlišit pouze úroveň rrrr-mm.

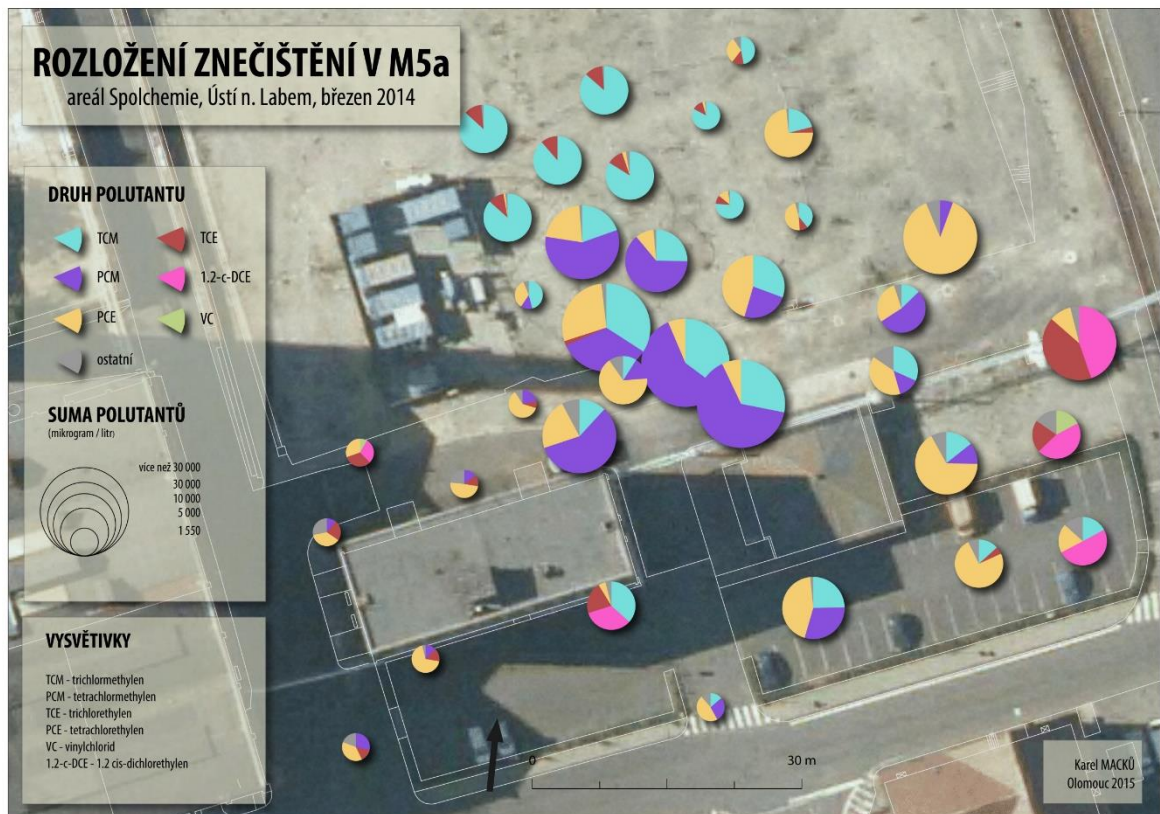
Aby bylo možno datové náhledy interpretovat, bylo nutné k animaci přidat také legendu. To je v Google Earth možné provést například vložením připraveného obrázku do mapového pole. Z ArcMap byla tedy vyexportována legenda včetně titulu, a tento obrázek byl pak vložen do kódu pomocí jednoduchého tagu <ScreenOverlay>. Obrázek byl uložen ve formátu .png, a byla mu nastavena průhlednost, aby lépe splynul s mapovým polem.

Takto modifikovaný soubor se všemi png obrázky je finálně zabalen do formátu kmz, který už lze spustit přímo v Google Earth, a pomocí ovládacího panelu časové osy pak prohlížet animaci.

6.2.5 Rozložení látek v prostoru

Poslední zkoumaný jev je chování jednotlivých polutantů v prostoru, jejich koncentrace v určitých částech mraku 5. Vycházelo se opět z dat vrtů pro jednotlivé roky, které byly v přechodích krocích použity pro prostorové interpolace. Jelikož některé polutanty výrazně převyšovaly svou koncentrací ostatní, bylo rozhodnuto, že vizualizovány budou pouze tři nejvýznamnější hodnoty, zbývající nízké koncentrace budou sloučeny do skupiny „ostatní“. Jako prvotní ukazatel sloužil součet celkových koncentrací jednotlivých látek, který ukázal nejvyšší hodnoty u trichlormethanu (975 000 $\mu\text{g/l}$), tetrachlormethanu (1 155 000 $\mu\text{g/l}$) a tetrachlorethylenu (855 000 $\mu\text{g/l}$). Je tedy vidět, že oproti PTS v mraku 5 převládají spíše chormethany.

Upravený datový soubor byl vizualizován pomocí součtových kruhových kartodiagramů, jejichž velikost byla normována pomocí celkové sumy polutantů. Takto každý datový náhled poskytuje informaci jednak o poměru nejvýznamnějších látek v daném vrtu, a také o celkovém množství těchto látek. Bylo zjištěno, že hlavně ve střední a severní části dominuje tetrachlormethan, silné jsou koncentrace také trichlormethanu a tetrachlorethylenu. Ve východní části převládá tetrachlorethylen, avšak od roku 2013 je zde silný nárůst 1.2-cis-dichlorethylenu. V jihozápadní části jsou koncentrace přibližně vyrovnané a také dost druhově proměnlivé. Z výsledků je patrné, že v rozložení látek v mraku 5 je patrný určitý opakující se vzor, který lehce kolísá v procentuálním zastoupení dominantních látek během po sobě jdoucích let. Vizualizace pro všechny roky jsou dostupné jako příloha na CD.



Obr. 13 Zastoupení látek v roce 2014

7 VÝSLEDKY

Na zájmové lokalitě proběhla datová analýza ve dvou částech – atributově zaměřená analýza podzemní těsnící stěny, a více prostorově charakteristický mrak 5A. Rozbor dat byl proveden v intervalu let 2007 – 2014. Jednotlivé dílčí výsledky jsou popsány níže, a úkolem této kapitoly je zhodnotit je, a sloučit do závěrečného tvrzení.

7.1 Výsledky analýzy PTS

Nejprve byly vypočteny základní popisné statistiky pro získání rychlého přehledu od datech. Již na těch byly patrné nestejnorození na jednotlivých branách. Pro detailnější rozbor bylo vytvořeno grafické zpracování popisující efektivitu čištění na branách. V některých měřeních byly zjištěny extrémně vymykající se hodnoty, například stonásobně vyšší koncentrace na výstupu než vstupu. Tyto hodnoty byly považovány za chybu ať už v měření či zaznamenání dat, a byly ze souboru odstraněny. Efektivita procesu je velmi dobrá, průměr na branách se pohybuje od 75 do 97 % u sumy 1 550 $\mu\text{g/l}$, a 80 – 97 % u sumy 2 000 $\mu\text{g/l}$. Výjimkou je brána 4, kde jsou průměrné hodnoty 43 %. To je však způsobeno zahrnutím hodnot z doby pilotní zkoušky, kde byla efektivita z počátku nižší. Efektivita ale nevyovídá o koncentracích vystupující z PTS – ty nesmí překročit povolené limity 1 550 a 2 000 $\mu\text{g/l}$. Tyto hodnoty nebyly nikdy překročeny, kromě brány RB4 a to opět v období pilotní zkoušky, po ní byl zaznamenány ještě 2 případy, kdy byly limity nepatrně překročeny.

Další hodnocení bylo provedeno pomocí korelačních matic a parciálních korelačních koeficientů. Účelem těchto výpočtů byla snaha o prokázání závislostí mezi některými atributy. Koeficienty byly počítány jak pro hodnoty na vstupu, tak pro změny hodnot při průchodu bránou. Zkoumány byly vztahy mezi polutanty, anorganickými látkami a také fyzikálně-chemickými parametry. Tato metoda se však ukázala pro případová data jako nevhodná: výsledné matice sice byly vhodně uspořádány do přehledné formy, ale nepodařilo se v nich objevit žádný opakující se vzor závislostí, který by platil pro více bran. Zároveň se neosvědčilo testování hypotéz mezi těmito maticemi na prokázání jejich podobnosti – testy byly nejednoznačně pravděpodobně z malého rozsahu testovaných hodnot (korelační koeficienty na intervalu -1, 1).

Vztahy mezi atributy byly hledány dále pomocí modelů beta regrese. Zde bylo na data nahlíženo ze strany efektivity čištění a zkoumáno, jaký vliv mají jednotlivé atributy na celkovou účinnost. Bohužel nešlo vytvořit zvlášť model pro každou bránu (nedostatek měření), a proto musel být použit jeden celkový pro všechny data dohromady. Tento výpočet ukázal jako nejvíce ovlivňující sumu dle rozhodnutí, vinylchlorid a sumu methylenů. Je však nutno brát v potaz, že model je zobecněn pro všechny brány dohromady, a jak bylo zjištěno už v předchozích krocích, rozptyl dat je na branách velice velký a proto může mít vliv i na tento celkový model.

Metod nelineární regrese bylo využito také k predikování vývoje efektivity bran. K tomuto účelu se ukázala logistická regrese jako velmi vhodný nástroj poskytující dobré výsledky. V některých případech byly data odivně zkreslená outliersy, a proto by bylo

vhodné mít k dispozici více měření a možnost tyto prvky vypustit. I přes to byly zjištěny trendy efektivity na branách, a výsledek říká, že převládá klesající nebo konstantní stav. Na branách 1, 8, 9 a 10 efektivita průběžně stoupala a drží se na vynikající úrovni. Avšak i u klesajících křivek není tento propad nijak výrazný či extrémní.

Poslední zkoumanou částí PTS byly změny atributů, a to anorganických látek a fyzikálně chemických parametrů. V těchto případech šlo spíše o potvrzení očekávaných průběhů – pokles Eh při průchodu bránou a nárůst hodnot pH. Tyto teorie byly úspěšně potvrzeny, v rámci Eh byl navíc zjištěn průběžný růst hodnot, který indikuje postupné slábnutí účinnosti. Za poslední období bylo však k dispozici jen málo měření, proto nelze určit, zda se nejedná pouze o dočasný výkyv.

Anorganické látky byly porovnávány graficky a také pomocí shlukové analýzy, která měla prokázat, zda změny vytváří shluky, které by odpovídaly jednotlivým branám. Rozbor většinou prokázal obecné trendy očekávané poklesy koncentrací látek, avšak ve většině případů byly rozptýly hodnot velmi velké na to, aby byla jednoznačně prokázána jasná podobnost mezi bránami či látkami. Z toho důvodu nepřinesla ani shluková analýza jasné uspokojivé výsledky, vzniklé shluky nekorespondovaly s reakčními bránami, a hodně se prolínaly mezi sebou. Nicméně bylo zajímavé vidět, že existují skupiny, které jsou si chováním velice podobné.

7.2 Výsledky analýzy M5

Na mraku 5 probíhalo méně analýz než na PTS. K dispozici byly jednak sumy koncentrací polutantů, které vstupovaly do čistícího procesu a sumy které z něj vystupovaly, a dále prostorově rozmístěné vrty. Na součtech byla provedena lineární regrese za účelem odhadu, kdy dojde k poklesu těchto hodnot pod žádaný limit. Pro sumu dle Rozhodnutí by mohlo být dosaženo limitu 1 550 µg/l v březnu 2019, pro novou sumu a limit 2 000 µg/l v září 2023. Tyto hodnoty jsou však velmi orientační.

Závěrečným krokem celé analýzy byla prostorová interpolace vrtů, ze kterých byl získán souvislý povrch znečištění pro sumy polutantů dle Rozhodnutí. Tyto interpolace byly provedeny pro celkem 9 časových údajů. Bylo zjištěno, že rozdíly v koncentracích jsou extrémně velké a to až v řádech desetitisíců. Tento fakt se také projevil na přesnosti interpolace. Koncentrace polutantů také výrazně kolísala v průběhu sledovaného období, nicméně bylo lokalizováno hlavní centrum znečištění, jehož poloha se výrazně nemění a nachází se přibližně ve středu zájmového území. Také bylo vyzorováno, že plochy koncentrací nad limit a pod limit jsou přibližně vyrovnané. Do celé časové řady nezapadá rok 2012, kdy byly pozorovány výrazně nižší koncentrace, ale v následujících letech hodnoty opět výrazně narostly.

Celkově lze zkoumané metody čištění podzemní vody označit za maximálně efektivní. S jejich pomocí byly na zájmovém území redukovány polutanty pod únosnou hranici a podzemní voda opouštějící areál hlavně oblasti PTS již není nebezpečím pro životní prostředí. Dále se ukázalo, že navržené technologie splňují očekávanou životnost, dokonce je možné počítat s delším fungováním, než je v původních plánech.

8 DISKUZE

Tato práce otestovala možnosti zpracování chemických data pomocí geostatistických analýz, a to hlediska analýzy atributové složky, tak složky prostorové. V tomto konkrétním případě nedocházelo k žádnému vlastnímu sběru dat a autor musel pracovat pouze s daty, které mu byla poskytnuta. A právě data byla nejslabší článek celého procesu.

Ve statistice se lze obecně řídit pravidlem, že čím víc dat je k dispozici, tím lepší výsledky lze dostat. S rostoucím počtem měření/záznamů se zvětšuje statistický vzorek a tím se stává více reprezentativním. V této práci byla data získávána odběrem vzorků vody v terénu a jejich následným laboratorním rozbořem. Je jasné, že takový proces je časově i finančně náročný, a proto jej nelze provádět v hustším časovém intervalu. Následně je nutné si uvědomit, že kvalita interpretace dat je závislá na kvalitě dat samotných. Hlavním problémem byla časová nesourodost dat – různé skupiny atributů byly měřeny v různý čas a nebo nebyly v daném období zaznamenávány vůbec. Následně tedy nebylo možné kombinovat mezi sebou všechny atributy tak, jak bylo původně zamýšleno. Ukázkou tohoto problému je například kombinace data fyzikálně-chemických vlastností s celkovými daty znečištění: v letech 2009 až 2014 byla pro fyzikálně-chemické parametry provedena pouze 3 měření, zatímco hodnoty znečištění byly zkoumány více než 10 krát. Takto pospojovaná data ve výsledku mají nízký počet pozorování, které činí výpočty, jako jsou například korelace, méně přesnými.

V datech byl také velmi velký rozsah hodnot. To má za následek nižší přesnost některých metod, například regrese. Ta sice na základě dostupných dat poskytne výsledek s nejmenší chybou, nicméně při velkém rozptylu hodnot může i takový výsledek být velmi zkreslený. Stejný problém nastal u změn anorganických látek, kde kvůli velkému rozptylu hodnot nebylo možné identifikovat jasné shluky a charakteristické chování látek.

Další záludností při zpracování takovýchto dat je tematická odbornost. Je jasné, že pro kvalitní analýzu je nutno znát bližší souvislosti, které však v tomto případě zcela překračují obor geoinformatiky či statistiky. Proto je nezbytné před samotnou prací provést konzultace s odborníky daného tématu, aby bylo datům a jednotlivým vztahům alespoň okrajově porozuměno. Následná interpretace výsledků je pro geoinformatika také náročná, proto je ideální řešení vhodně okomentované výsledky poskytnout tematikům, kteří již jsou schopni nalézt bližší souvislosti a hlubší význam.

Závěrem je nutné si uvědomit, že statistické hodnocení a zpracování dat není bezchybné ani všemocné, a je nutné se na zjištěné výsledky dívat kriticky, jelikož mohou být zatíženy různými chybami, např. z důvodu nedostatku dat či momentálních extrémních hodnot, které mění celkový trend. Konkrétní případ je prostorová interpolace, kde z důvodu velkého rozsahu hodnot docházelo k nepřesnostem a interpolované povrchy určitě nelze použít k tvrzení, že v souřadnici XY je hodnota jevu právě a přesně taková, ale naopak lze celkem správně tvrdit, že v určité části zájmové oblasti jsou hodnoty výrazně vyšší než v částech jiných.

9 ZÁVĚR

Úkolem této práce bylo aplikovat metody geoinformatiky a geostatistiky na tematiku znečištění podzemních vod a prověřit účinnost současných čistících technologií. Na této zajímavé problematice byly vyzkoušeny různé metody zpracování dat, a to jak na atributové tak i prostorové úrovni, a byla hodnocena vhodnost těchto metod.

Práce byla rozdělena do dvou hlavních částí – zpracování dat z tzv. podzemní těsnící stěny a dat z prostorově zajímavějšího rozložení kontrolních vrtů, pod označením mrak 5A. Tyto oblasti se lišily použitými technologiemi čištění a metodami analýzy.

Základ hodnocení byl postaven na základních popisných statistikách a grafické vizualizaci zájmových témat. Jednoduché grafické vizualizace (např. boxplot) se ukázaly jako velice efektivní nástroj pro rychlé zhodnocení situace a porovnání rozdílů, například mezi jednotlivými reakčními bránami. Při vhodné úpravě dat a volby správné metody lze celkem snadno předpovídat hodnoty do budoucna pomocí různých regresních modelů. Naopak ne příliš dobře se osvědčily výpočty korelací – datový soubor byl příliš pestrý a proměnlivý na to, aby v něm byly nalezeny závislosti opakující se v jednotlivých bránách. Vypočtené korelační matice jsou ve výsledku také kvůli velkému množství atributů a míst pozorování méně přehledné. Rozporuplné výsledky přinesly shlukové analýzy – proběhly sice bez problému a identifikovaly shluky, avšak ty se příliš neshodovaly s existujícím rozdělením. Z toho plyne další závěr, že zpracovávaný datový soubor byl velmi proměnlivý ať už z hlediska času či prostoru, a z toho důvodu v něm byl velmi těžké hledat souvislosti a jednoznačně interpretovatelné vzory. Očekávalo se podobnější chování mezi jednotlivými bránami, avšak výše zmíněné rozbory výrazné shody zamítly.

Pomocí interpolací byly určeny centra znečištění v mraku 5. Mapové vizualizace také umožňují přehledně zobrazit převažující koncentrace jednotlivých polutantů. V mraku 5 je stále identifikováno silné znečištění, ale pomocí regresního modelu je patrný celkový trend poklesu.

Závěrem této práce je nutno říct, že výsledky zjištěné geostatistickými analýzami vypovídají o vysoké efektivitě obou metod čištění použitých na lokalitě, a velké kvalitě soudobých dekontaminačních metod. Díky nim mohou být prováděny rozsáhlé sanační práce, které tak velkou mírou přispívají ke zdravějšímu životnímu prostředí a ekologické stabilitě krajiny.

POUŽITÁ LITERATURA A INFORMAČNÍ ZDROJE

Knižní zdroje

BARTOŇ, Jan. Kontaminace podzemních vod chlorovanými uhlovodíky na lokalitě DKV Břeclav. Brno, 2007. Diplomová práce. Masarykova univerzita.

BURIAN, Jaroslav. Vyznáte se v interpolacích?. *Geobusiness*. 2008, č. 11.

BRÁZDIL, Rudolf. *Statistické metody v geografii*. Brno: Univerzita J.E.Purkyně, 1981.

HENGL, T. *A Practical Guide to Geostatistical Mapping*. Luxembourg : Office for Official Publications of the European Communities, 2007. ISBN 978-92-79-06904-8.

KVAPIL, Petr. AQUATEST A.S. *Vyhodnocení výsledků pilotní brány RB-4*. 2007.

LI, B.G. Geostatistical analysis and kriging of Hexachlorocyclohexane residues in topsoil from Tianjin, China. *Environmental Pollution*. 2005, č. 142.

MELOUN, Milan.; MILÍTKÝ, Jiří. *Statistická analýza experimentálních dat*. Praha: Academia, 2004. 953 s. ISBN 80-200-1254-0.

NETOPIL, Rostislav. *Základy hydrologie povrchových a podpovrchových vod*. Praha: Státní pedagogické nakladatelství, 1970.

PLAŠIL, Miroslav a Petr VLACH. Grafická analýza vícerozměrných dat. [online]. 2007 [cit. 2015-04-05]. Dostupné z:
<https://www.vse.cz/polek/download.php?jnl=aop&pdf=43.pdf>

STEJSKAL, Vojtěch. Kombinovaná metoda – Sanace podzemních vod za využití kombinace laktátu sodného a nanoželeza. Praha, 2014. Diplomová práce. Univerzita Karlova.

STEJSKAL, Vojtěch. AQUATEST A.S. *Informativní zpráva pro 47. KD*. 2014.

ZŮVALA, Robert. *Analýza sesuvu svahu pomocí metody kriging*. Olomouc, 2011. Bakalářská práce. Univerzita Palackého.

Elektronické zdroje

Dílčí korelační koeficienty. IASTAT - interaktivní učebnice statistiky [online]. 2000 [cit. 2015-04-30]. Dostupné z: <http://iastat.vse.cz/regrese/Korelace4.htm>

FERRARI, Silvia L. P. a CRIBARI-NETO, Francisco. Beta regression for modelling rates and proportion. Instituto de matematica e estatistica [online]. [cit. 2015-04-30]. Dostupné z: <http://www.ime.usp.br/~sferrari/beta.pdf>

How kriging works. ESRI. ArcGIS Resource center [online]. 2011 [cit. 2015-05-01]. Dostupné z: <http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/009z00000076000000.htm>

Podpovrchová voda. Fyzická geografie - pedagogická fakulta Masarykovy university [online]. [cit. 2014-03-17]. Dostupné z: http://is.muni.cz/do/rect/el/estud/pedf/js13/fyz_geogr/web/pages/11-podpovrchova-voda.html

Podzemní voda. [online]. Ústav geotechniky Fakulty stavební Vysokého učení technického v Brně. [cit. 2014-03-17]. Dostupné z: <http://geotech.fce.vutbr.cz/studium/geologie/skripta/PODVODA.htm>

Podzemní voda. In: Česká geologická služba [online]. [cit. 2015-04-30]. Dostupné z: <http://www.geology.cz/rok-zeme/dokumenty/mrpz-vystava-poster-podzemni-voda.pdf>

RODRÍGUEZ, Germán. Introducing R: Generalized Linear Models. Princeton University [online]. 2015 [cit. 2015-04-30]. Dostupné z: <http://data.princeton.edu/R/glms.html>

Simple linear correlation and regression. Coastal Carolina University [online]. [cit. 2015-04-30]. Dostupné z: <http://ww2.coastal.edu/kingw/statistics/R-tutorials/simplelinear.html>

Wilcoxonův test. Statistika a výpočetní technika [online]. Brno: Fakulta veterinární hygieny a ekologie [cit. 2015-04-30]. Dostupné z: <http://cit.vfu.cz/statpotr/POTR/Teorie/Predn4/Wilcoxon.htm>

Základy hydrogeologie. Geologie - výukové multimediální texty [online]. Vysoká škola báňská - Technická univerzita Ostrava [cit. 2015-04-30]. Dostupné z: http://geologie.vsb.cz/geologie/kapitoly/10_z%20klady_hydro/10_z%20klady_hydrogeologie.htm

PŘÍLOHY

SEZNAM PŘÍLOH

Vázané přílohy:

Příloha 1 Programový kód v R

Příloha 2 Ukázkové grafické výstupy pro reakční bránu 5

Příloha 3 Změny v prostorovém znečištění na M5

Volné přílohy:

Příloha 4 CD – zpracované datové sady, kompletní grafické a tabelární vizualizace

Příloha 5 Poster

Příloha 1 - Programový kód v R:

Parciální korelace – matice

```
attach(RB3)
x <- RB3[1:12]
y <- RB3[1:12]
korelace <- cor(x,y)
partial = cor2pcor(korelace)
View(partial)
write.table (partial,file="vystupRB3.txt")
```

Shluková analýza:

1) Hierarchická metoda:

```
attach(s1)
data <- s1
data[3:8] <- scale(data[3:8]) #standartizace dat
dist<- dist(data[,3:8], method = "euclidean")
clust <- hclust(dist, method = "ward.D")
plot(clust)
cutree <- cutree(clust, k=10)
rect.hclust(clust, k=10,border="red3") #dendrogram s ramováním
data2 <- cbind(data, cutree)
sum<-describeBy(data, group=cutree,mat=FALSE,type=3)
sum
#dendrogram, jiný přístup
install.packages("ggdendro")
dhc <- as.dendrogram(clust)
ddata <- dendro_data(dhc, type = "rectangle")
p <- ggplot(segment(ddata))
  + geom_segment(aes(x = x, y = y, xend = xend, yend = yend)) +
  + coord_flip() + xlab("ID záznamu")
  + labs(title = "DENDROGRAM SHLUKOVÉ ANALÝZY")
  + theme(plot.title = element_text(size=20, face="bold",vjust=2))
```

2) K-means metoda:

```
mydata <- scale(s1[3:8])
wss <- (nrow(mydata)-1)*sum(apply(mydata,2,var)) #výpocet počtu shluků
for (i in 2:15) wss[i] <- sum(kmeans(mydata,centers=i)$withinss)
plot(1:15, wss, type="b", xlab="Počet shluků")
fit <- kmeans(mydata, 10)
aggregate(mydata,by=list(fit$cluster),FUN=mean)
mydata <- data.frame(mydata, fit$cluster) #zápis do dat
```

Beta regrese:

```
library(betareg)
model.reg <- betareg(UCIN ~ VC + X1_1_D + X1_2_C + X1_2_T + TCE + PCE +
SUM + SUM_CL + SUM_R + CL + N + S + HCO + CA +FE)
summary(model.reg)
```

Predikce s logistickou regresí:

```
rok.gml <- glm (formula = JEV ~ DAT, family=quasibinomial(logit))
```

```

roky.pred <- data.frame(DAT = 39490,39546,39625,39678,39741,39869,39925,
40107,40261,40652,40814,41012,41208,41557,41725,42005,42156,42370,42522,
42736,42887,43101,43252,43466,43617,43831) #definice predikovaných let
predict <- predict(rok.gml, roky.pred, type="response")

#příprava plotu pro vykreslení
plot(c(39500, 43000), c(0,1), type = "n", xlab = "Roky", ylab = "Efektivita (%)", font.lab="2", cex.lab=1.1, bty="1",
+ main = "PREDIKCE EFEKTIVITY rb10 \n v letech 2008 - 2017",
+ col.main=rgb(0,0,0), font.main="2",cex.main=1.5, col.sub=rgb(0,0,0),
+ font.sub="2", cex.sub=1.5)

#vykreslení predikovaných a známých hodnot
lines(roky.pred, predict, lty = "dotted", col="red", lwd="1" )
lines(roky, rok.gml$fitted, type="l", col="green", lwd="1")

```

Lineární regrese

```

lm.out <- lm(SUM ~ DAT)
lm.out
summary(lm.out)
plot(lm.out)

```

Tvorba grafu z lineární regrese:

```

ggplot(lm2.out, aes(x=DAT, y=SUM)) +
+ xlab("Datum") + xlim(39200,41800) + ylab("Suma polutantů") + coord_cartesian(ylim = c(0,50000)) + coord_cartesian(xlim = c(39150,41800))
+ scale_y_continuous(breaks=seq(0,50000,10000))
+ scale_x_continuous(breaks=seq(39200,41750,510)) +
+ labs(title = "VÝVOJ ZNEČIŠTĚNÍ V LETECH 2007 - 2014") +
+ theme(plot.title = element_text(size=20, face="bold", vjust=2)) +
+ geom_point(aes(colour = UCIN), size = 5)
+ geom_smooth(method=lm, size=1.7, col="#003366")
+ theme(plot.background = element_rect(fill = '#CCCCCC'))

```

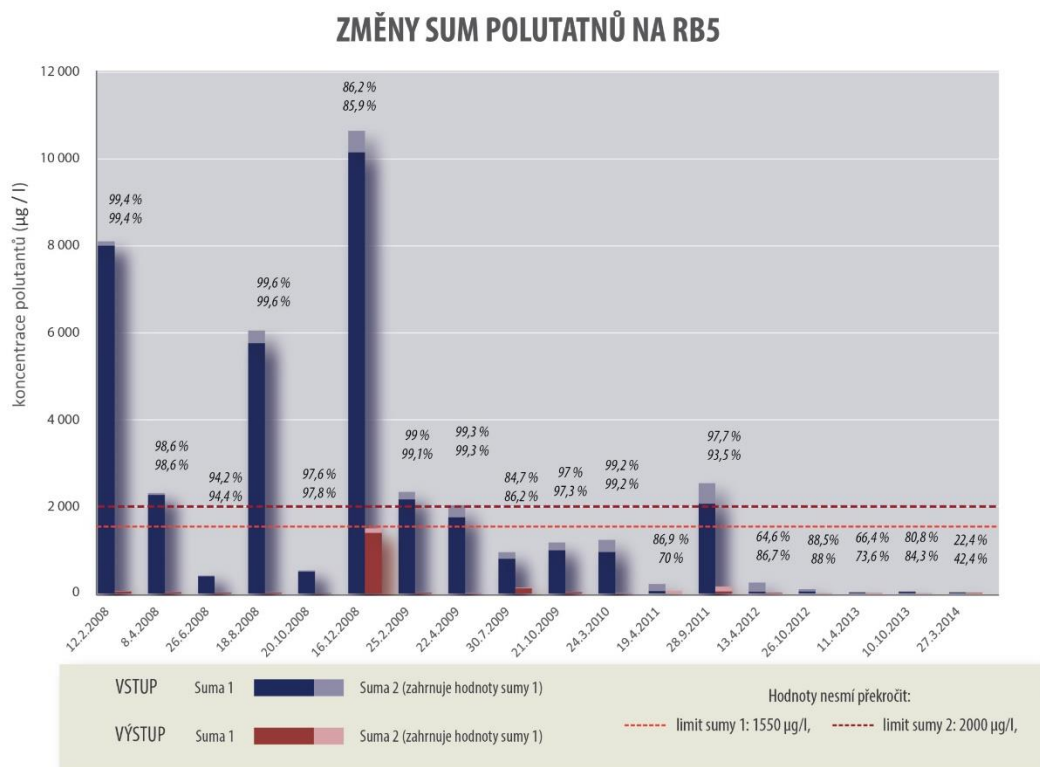
Boxploty

```

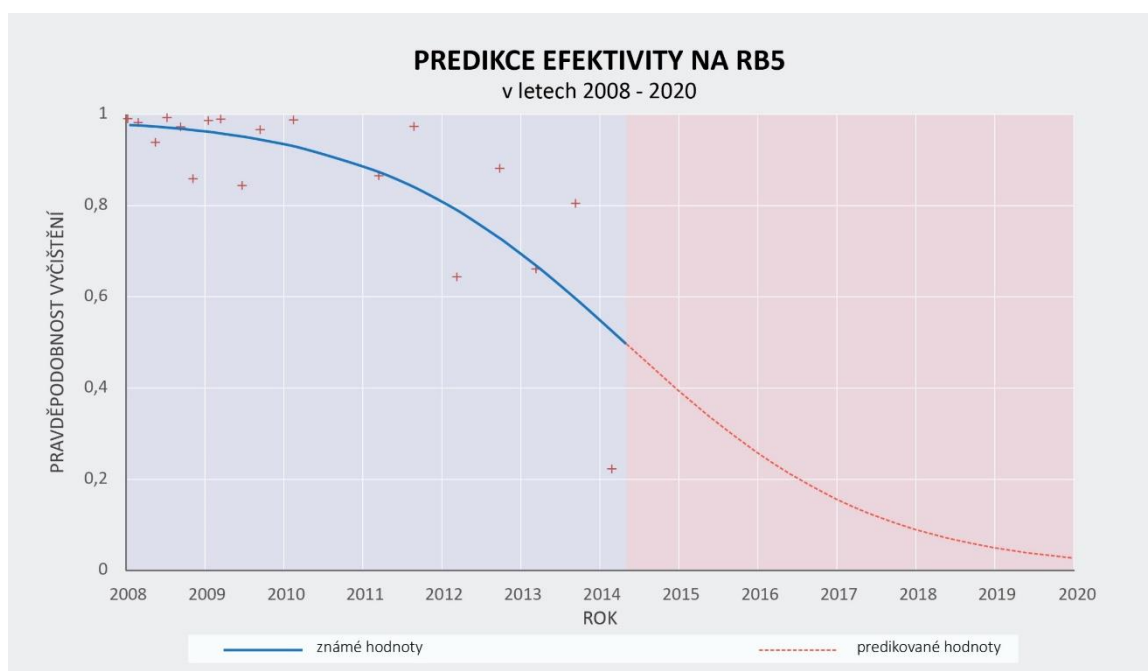
ggplot(vstup, aes(x=BRANA, y=Eh, fill = BRANA))
+ geom_boxplot()
+ scale_y_continuous(breaks=seq(-1000,500,250))
+ labs(title = "OXIDAČNĚ-REDUKČNÍ POTENCIÁL NA VSTUPU DO BRAN")
+ xlab("Brána") + ylab("Eh")
+ theme(plot.title = element_text(size=20, face="bold", vjust=2))

```

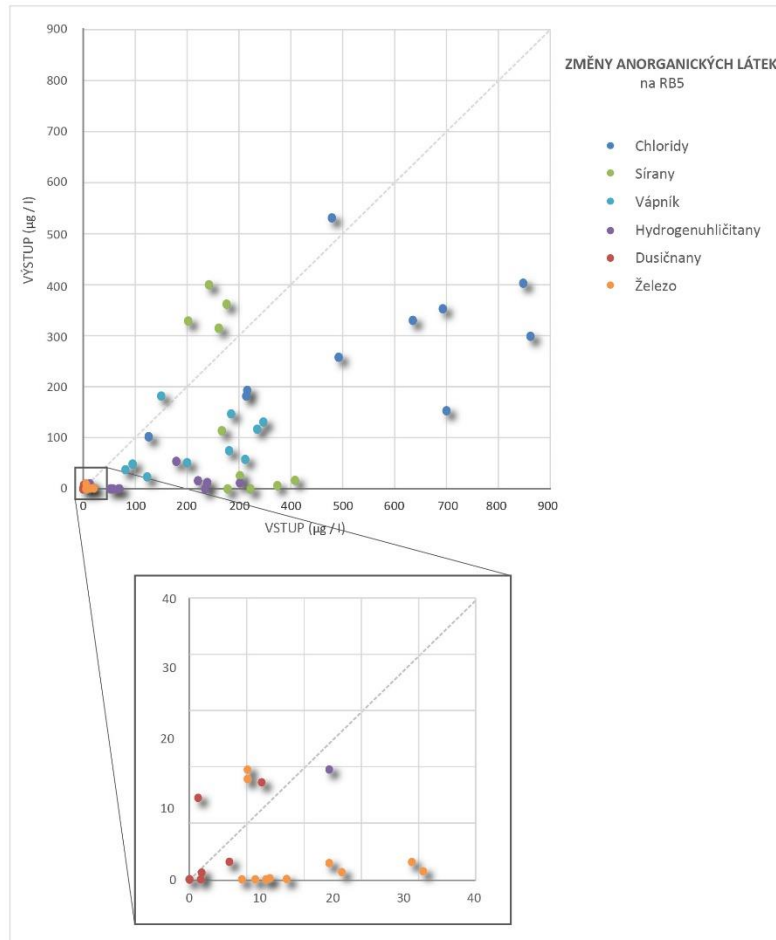
Příloha 2 – Ukázkové grafické výstupy pro bránu 5



Obr. 14 Vývoj účinnosti čištění na RB5

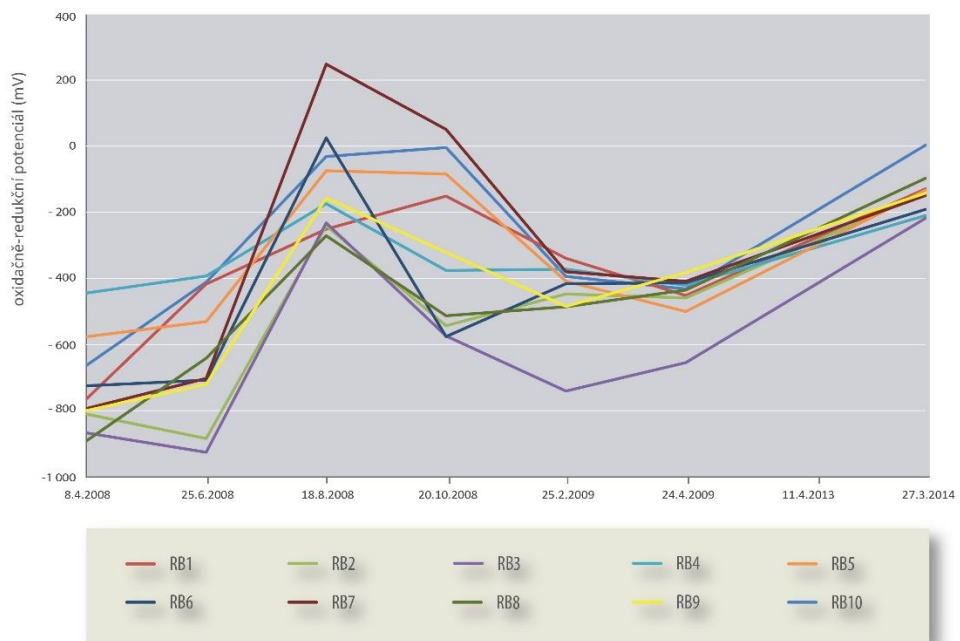


Obr. 2 Predikce efektivity logistickou regresí



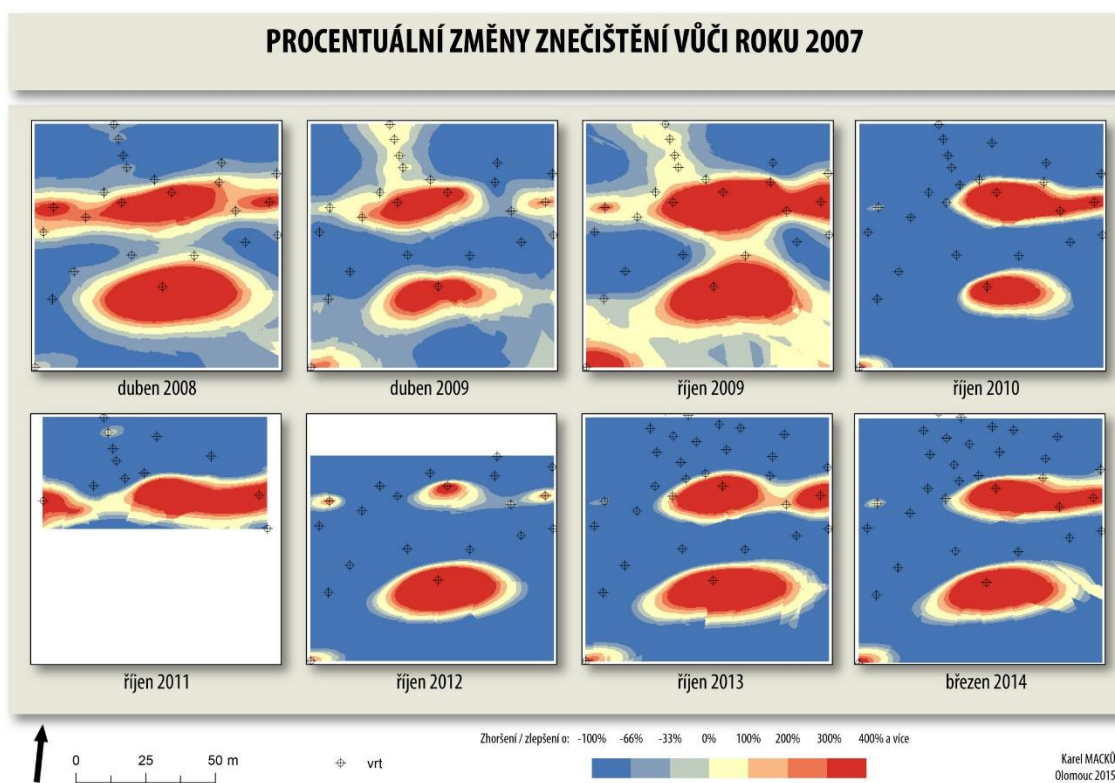
Obr. 3 Změny anorganických látek

VÝVOJ OXIDAČNĚ-REDUKČNÍHO POTENCIÁLU na výstupu z reakční brány

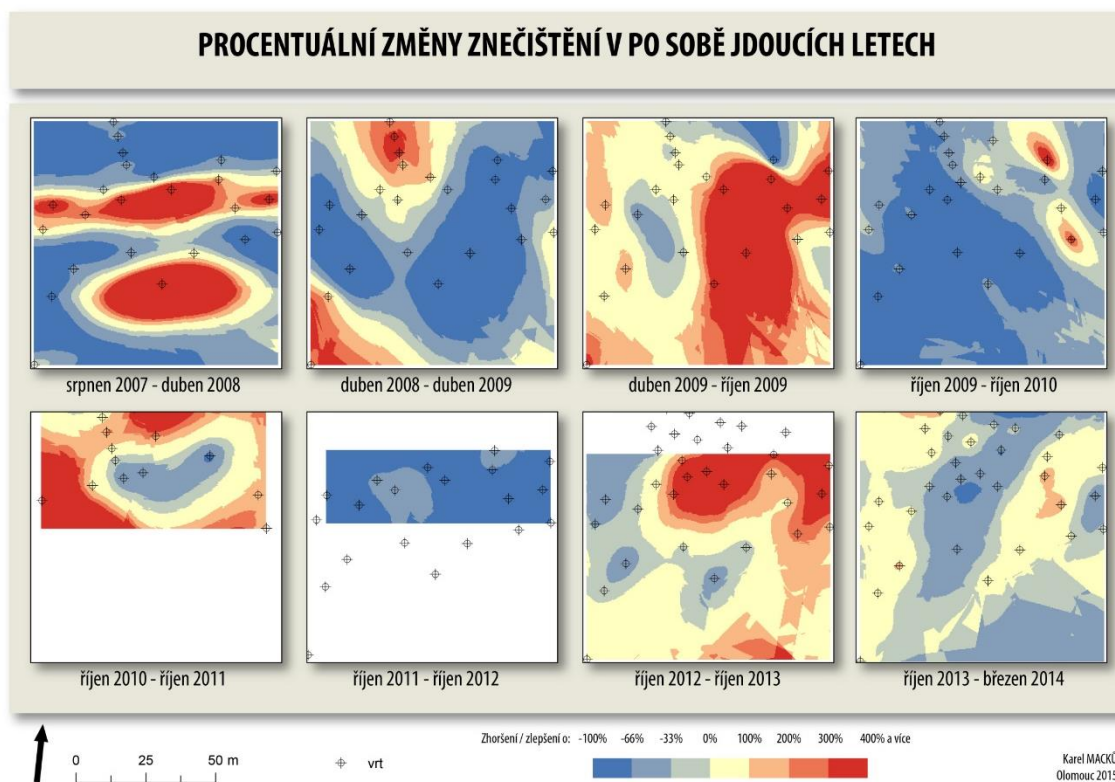


Obr. 4 Vývoj oxidačně-redukčního potenciálu na výstupu z bran

Příloha 3 - Změny v prostorovém znečištění na M5



Obr. 5 Vývoj znečištění vůči roku 2007



Obr. 6 Vývoj znečištění v po sobě jdoucích letech