

**Univerzita Karlova**

**Přírodovědecká fakulta**

Katedra aplikované geoinformatiky a kartografie

Studijní program: Geografie

Studijní obor: Geoinformatika a kartografie



**Bc. Andrea STRAMOVÁ**

**GEOINFORMAČNÍ ANALÝZA ZDRAVOTNÍCH DAT**

**SPATIAL ANALYSIS OF MEDICAL DATA**

Diplomová práce

Vedoucí diplomové práce: Mgr. Lukáš Brůha, Ph.D.

Praha 2017

**Vysoká škola:** Univerzita Karlova v Praze

**Fakulta:** Přírodovědecká

**Katedra:** Aplikované geoinformatiky a kartografie

**Školní rok:** 2016/2017

## Zadání diplomové práce

**pro** Bc. Andrea Stramová

**obor** Kartografie a geoinformatika

**Název tématu:** Geoinformační analýza zdravotních dat

### Zásady pro vypracování

Diplomová práce navrhne postup, který identifikuje rizikové faktory přírodního prostředí a objasní jejich vliv na funkci štítné žlázy.

Díličmi cíli práce jsou:

- nalezení a získání vhodných geokódovaných zdravotnických data, která pokrývají zájmové území,
- navržení postupu přípravy těchto dat ke vstupu do následných analýz,
- faktorová analýza zdravotních dat,
- nalezení a popis vztahů v prostorovém uspořádání zdravotních dat.

**Rozsah grafických prací:** dle potřeby

**Rozsah průvodní zprávy:** 70 – 100 stran

**Seznam odborné literatury:**

BIVAND, R. S., PEBESMA, E., V. GOMEZ-RUBIO. Applied Spatial Data Analysis with R. UseR! Series, Springer 2nd ed. 2013, xviii+405 pp.

R. TOMANOVÁ. Vliv geologie a radioaktivního záření na výskyt dětské tyreopatie: testovaná lokalita Příbramsko. Praha, 78 s. Rigorózní práce, Ústav geologie a paleontologie, PŘF UK, 2005. Vedoucí práce Karel Martínek.

TRYGG, J., S. WOLD. O2-PLS, a two-block (X–Y) latent variable regression (LVR) method with an integral OSC filter. Journal of Chemometrics, 2003, 17.1: 53-64.

Vedoucí diplomové práce:	Mgr. Lukáš Brůha, Ph.D.
Konzultant diplomové práce:	Ing. Martin Hill, DrSc.
Datum zadání diplomové práce:	16. 01. 2015
Termín odevzdání diplomové práce:	2. 5. 2017

*Platnost tohoto zadání je po dobu jednoho akademického roku.*

.....

Vedoucí diplomové práce

.....

Garant oboru

V Praze dne 2. 5. 2017

#### Prohlášení

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

Jsem si vědoma toho, že případné použití výsledků, získaných v této práci, mimo Univerzitu Karlovu v Praze je možné pouze po písemném souhlasu této univerzity. Svoluji k zapůjčení této práce pro studijní účely a souhlasím s tím, aby byla řádně vedena v evidenci vypůjčovatelů.

V Praze dne 2. 5. 2017

.....  
Andrea Stramová

### **Poděkování**

Ráda bych zde poděkovala svému vedoucímu práce, Mgr. Lukáši Brůhovi, Ph. D., za trpělivost a hodnotné rady; dále svým školitelům, Ing. Martinu Hillovi, DrSc. za pomoc při hledání vhodných metod a za zaškolení do zvolených metod, MUDr. Jiřímu Němečkovi za umožnění proniknout do problematiky štítné žlázy a pomoc při získávání endokrinologických dat, RNDr. Jakubovi Trubačovi, Ph.D. za konzultace a rady při práci s geologickými daty a za pomoc při jejich získání. Dále bych chtěla poděkovat jednotlivým institucím (Výzkumný ústav Silva Taroucy pro krajinu a okrasné zahradnictví, Český hydrometeorologický ústav, Ředitelství silnic a dálnic, Státní zdravotní ústav, Česká geologická služba a Endokrinologický ústav) za vstřícné jednání a poskytnutí dat. A nakonec, i když neméně, své rodině a blízkým za morální a materiální podporu v průběhu tvorby.

## **Abstrakt**

V dnešní době zatím není prokázáno, jaké všechny aspekty mohou působit na funkci štítné žlázy. U mnoha nemocí je prokázán vliv složek životního prostředí. Z tohoto důvodu se tato práce zabývá vlivem parametrů životního prostředí na funkci štítné žlázy. Cílem této práce je navrhnout metodiku, která prostorově a statisticky popíše zvolená medicínální data, odhalí pravidelnosti v jejich prostorovém uspořádání a vytvoří regresní model. Prostorové uspořádání dat je popsáno pomocí metody nejbližšího souseda a lokálním Moranovo I. Následně je aplikována metoda O2PLS, která navrhuje regresní model. V datech bylo identifikováno shlukové uspořádání v okolí měst. V práci byl navržen regresní model, který zohledňuje aspekty životního prostředí (koncentrace As, Cd, Pb a Se ve vodě). U ostatních parametrů životního prostředí nebyl prokázán vliv na funkci štítné žlázy. Tato práce je přínosem hlavně v oblasti prostorové analýzy dat a dále také v endokrinologickém oboru. Díky této práci vznikl postup, jak zpracovávat citlivá zdravotnická data a vyvozovat z nich informace o jejich souvislosti s daty prostorovými.

**Klíčová slova:** Prostorová analýza, kvantitativní metody, O2PLS, životní prostředí, štítná žláza.

## **Abstract**

Nowadays, all the aspects that can influence the function of the thyroid gland are not known. Since the influence of environment on progression of many diseases is proved, this thesis deals with the influence of environmental parameters on the thyroid function. The purpose of this thesis is to propose a methodology which will describe selected medical data spatially and statistically, reveal regularities in its spatial organization and create the regression model. The special organization of the data is described by the closest neighbour's methodology and by the Local Moran's I. Then the methodology of O2PLS, which proposes the regression model, is used. There were identified the cluster organization in the vicinity of towns. In the thesis, the regression model, which takes account of the aspects of the environment (the concentration of As, Cd, Pb and Se in water), was proposed. The influence of the other environmental parameters on the thyroid function was not proved. The thesis is a benefit mainly for the field of spatial data analysis and also for the endocrinological specialization. Owing to this thesis, the process of the way of working with sensitive medical data, that can be used for inferring important information about its connection to the spatial data, was created.

**Key words:** Spatial analysis, quantitative methods, O2PLS, the environment, thyroid gland.

## Obsah

1	Úvod.....	9
2	Rešerše literatury.....	10
2.1	Výzkum v České republice .....	11
3	Prostorová analýza dat .....	13
3.1	Metody a definice .....	13
3.2	Návrh metodiky .....	19
4	Charakteristika území a vybraných dat.....	21
4.1	Zájmové území Příbramsko.....	21
4.1.1	Topografická data .....	22
4.1.2	Výzkumný ústav Silva Taroucy pro krajinu a okrasné zahradnictví (VÚKOZ)..	23
4.1.3	Český hydrometeorologický ústav (ČHMÚ) .....	23
4.1.4	Ředitelství silnic a dálnic (ŘSD) .....	24
4.1.5	Státní zdravotní ústav (SZU).....	24
4.2	Geologie Příbramska.....	24
4.2.1	Geologická data .....	26
4.3	Endokrinologie .....	27
4.3.1	Endokrinologická data .....	28
5	Příprava dat.....	30
5.1	Environmentální data .....	30
5.1.1	Data z biomonitoring zátěže přírody .....	30
5.1.2	Data od ČHMÚ .....	32
5.1.3	Data z analýzy pitné vody .....	33
5.1.4	Intenzita dopravy.....	33
5.2	Geologická data .....	34
5.3	Medicinální data .....	36
5.3.1	Školy.....	37
5.4	Normalita dat a mocinná transformace .....	38
6	Mapování a popis výskytu zhoršené funkce štítné žlázy .....	41
6.1	Statistický popis zhoršené funkce štítné žlázy .....	41

6.2	Výskyt překročení norem na základě polohy.....	45
7	Analýza prostorových vzorů.....	48
7.1	Explorační analýza prostorové statistiky .....	48
7.2	Konfirmační analýza prostorové statistiky.....	50
7.2.1	Metoda nejbližšího souseda .....	57
7.2.2	Ripleyho K - funkce .....	58
7.2.3	Lokální Moranovo I .....	59
8	Analýza výskytu nemoci s faktory prostředí .....	61
8.1	O2PLS .....	63
8.1.1	O2PLS při nastavení závislé proměnné všech zdravotních parametrů.....	65
8.1.2	O2PLS při nastavení závisle proměnných parametrů štítné žlázy .....	66
8.2	Znázornění v prostoru.....	76
9	Diskuse výsledků .....	84
10	Závěr .....	88
Zdroje	.....	90
Seznam příloh	.....	95



## Seznam tabulek, obrázků a grafů

Tab. 1.	Přehled vodovodů použitých v práci .....	33
Tab. 2.	Překročení hladin norem ukazatelů.....	42
Tab. 3.	Počet překročení norem celkem.....	43
Tab. 4.	Počet překročení norem u Volgut, TSH, FT3, FT4 a Jodur .....	43
Tab. 5.	Počet překročení norem ve školách .....	44
Tab. 6.	Počet překročených norem u pacientů .....	46
Tab. 7.	Metoda nejbližšího souseda .....	58
Tab. 8.	Vztah mezi TSH, FT3, Volgut a vnějšími faktory hodnocený modelem O2PLS, prediktivní komponenta 1 .....	67
Tab. 9.	Vztah mezi TSH, FT3, Volgut a vnějšími faktory hodnocený modelem O2PLS, prediktivní komponenta 2 .....	69
Tab. 10.	Vztah Volgut vs. prediktory hodnocený vícenásobnou regresí .....	71
Tab. 11.	Vztah TSH vs. prediktory hodnocený vícenásobnou regresí.....	72
Tab. 12.	Vztah FT3 vs. prediktory hodnocený vícenásobnou regresí.....	74
Tab. 13.	Korelace závisle proměnných s komponentami pro všechna měření a po křížové validaci.....	76
Obr. 1	Normální (Gaussovo) rozdělení .....	16
Obr. 2	O2PLS modelující dvě datové tabulky.....	18
Obr. 3	Vymezení okresu Příbram.....	22
Obr. 4	Geologická mapa Příbramska .....	25
Obr. 5	Štítná žláza .....	27
Obr. 6	Prostorové znázornění koncentrace SO <sub>2</sub> a FE v oblasti Příbramska .....	31
Obr. 7	Mapa vybraných environmentálních dat v okrese Příbram .....	34
Obr. 8	Mapa radonového indexu a zlomů v okrese Příbram.....	35
Obr. 9	Znázornění vyšetřených pacientů, škol a typu hornin v okrese Příbram.....	37
Obr. 10	Vybraní pacienti a koncentrace PM10 v imisích v okrese Příbram.....	47

Obr. 11	Prostorová popisná statistika .....	49
Obr. 12	Koncentrace překročených norem na 1 km <sup>2</sup> u pacientů v okrese Příbram ....	50
Obr. 13	Počet překročených norem na 1000 obyvatel v okrese Příbram .....	51
Obr. 14	Koncentrace překročených norem Volgut na 1 km <sup>2</sup> u pacientů v okrese Příbram .....	52
Obr. 15	Počet překročených norem Volgut na 1000 obyvatel v okrese Příbram .....	52
Obr. 16	Koncentrace překročených spodních a horních hranic norem TSH na 1 km <sup>2</sup> u pacientů v okrese Příbram .....	53
Obr. 17	Počet překročených spodních a horních hranic norem TSH na 1000 obyvatel v okrese Příbram .....	54
Obr. 18	Koncentrace překročených spodních a horních hranic norem fT3 na 1 km <sup>2</sup> u pacientů v okrese Příbram .....	55
Obr. 19	Počet překročených spodních a horních hranic norem fT3 na 1000 obyvatel v okrese Příbram .....	55
Obr. 20	Koncentrace překročených spodních a horních hranic norem fT4 na 1 km <sup>2</sup> u pacientů v okrese Příbram .....	56
Obr. 21	Počet překročených spodních a horních hranic norem fT4 na 1000 obyvatel v okrese Příbram .....	57
Obr. 22	Výsledek metody nejbližšího souseda .....	58
Obr. 23	Typ prostorové asociace – Volgut, TSH, fT3, fT4 .....	60
Obr. 24	Algoritmus pro analýzu výskytu nemoci s vnějšími faktory .....	65
Obr. 25	Koncentrace As ve vodě a překročené normy Volgut na 1000 obyvatel .....	77
Obr. 26	Koncentrace Pl ve vodě a překročené normy Volgut na 1000 obyvatel .....	79
Obr. 27	Koncentrace As ve vodě a překročené horní hranice normy TSH na 1000 obyvatel .....	80
Obr. 28	Koncentrace As ve vodě a překročené spodní hranice normy TSH na 1000 obyvatel .....	80
Obr. 29	Koncentrace As ve vodě a překročené horní hranice normy fT3 na 1000 obyvatel .....	82

Obr. 30	Koncentrace As ve vodě a překročené spodní hranice normy fT3 na 1000 obyvatel .....	82
Graf 1.	Postupná úprava hodnot pro parametr TSH.....	39
Graf 2.	Hodnoty Chromu v mechu po transformaci bez úprav .....	40
Graf 3.	Ripleyho K- funkce .....	59
Graf 4.	Hottelingova $T^2$ statistika.....	63
Graf 5.	Komponentní váhy první prediktivní komponenty vyjádřené jako regresní koeficienty s jejich 95 % intervaly spolehlivosti.....	68
Graf 6.	Komponentní váhy druhé prediktivní komponenty vyjádřené jako regresní koeficienty s jejich 95 % intervaly spolehlivosti.....	70
Graf 7.	Výsledek vícenásobné regrese pro Volgut.....	71
Graf 8.	Výsledek vícenásobné regrese pro TSH .....	73
Graf 9.	Výsledek vícenásobné regrese pro fT3 .....	74
Graf 10.	Graf korelace predikovaných hodnot se skutečností pro Volgut .....	75
Graf 11.	Graf korelace predikovaných hodnot se skutečností pro TSH.....	75
Graf 12.	Graf korelace predikovaných hodnot se skutečností pro fT3.....	76
Graf 13.	Křížová kovariance As ve vodě a Volgut .....	78
Graf 14.	Křížová kovariance PI ve vodě a Volgut .....	79
Graf 15.	Křížová kovariance As ve vodě a TSH.....	81
Graf 16.	Křížová kovariance As ve vodě a fT3.....	83

# 1 Úvod

Již před naším letopočtem se lidé zajímali o výskyt nemocí, jejichž vznik souvisel s prostředím, v němž žili. Právě prostředí bylo jedním z faktorů ovlivňujících jejich zdravotní stav, kterému lidé přisuzovali velký význam. S přelomem přišel John Snow, který odhalil souvislost mezi výskytem cholery a složením (zdrojem) vody v pumpách. V této době byly k dispozici pouze tištěné mapy a vše muselo být zpracováno ručně. V současnosti se k analýzám zdravotních dat používají metody kvantitativní a kvalitativní statistiky, které jsou ve velké míře podporovány geoinformačními systémy. Tyto systémy se již mnohokrát při analýze zdravotních dat ukázaly jako přínosné (Ashizawa a kol. 1997, Grundloch 2009, Elliott a kol. 2001, Wang a kol. 2015).

K provedení kvalitní analýzy zdravotních dat je potřeba znát o pacientovi co největší množství informací. Důležitou složkou prostorových analýz je bydliště pacienta a jeho okolí. Údaje o těchto komponentách by měly být co nejkonkrétnější. Při zpracování zdravotních dat je třeba mít k dispozici hodnoty o pacientech, které nejsou upravené nebo kategorizované.

Součástí prostorové analýzy dat jsou také data, která charakterizují místo bydliště a okolí jedince. Tato data by měla pokrýt co nejširší spektrum informací o životním prostředí. Důležitým aspektem pro vstup do analýz je přesnost dat a zpracování. V České republice je pořizováno značné množství dat. Pro diplomovou práci je důležité vybrat vhodná data týkající se daného území a také možné korelace se zkoumanou nemocí.

Cílem diplomové práce je vytvoření regresního modelu, který vysvětluje ovlivnění zhoršené funkce štítné žlázy rizikovými faktory přírodního prostředí. K dosažení tohoto cíle je třeba vyřešit tyto problémy:

- Získat vhodná prostorová a zdravotnická data, která pokrývají zájmové území.
- Data upravit pro vstup do následných analýz.
- Prostorová data popsat ve vztahu ke zdravotním aspektům funkce štítné žlázy. Zdravotnická data statisticky vyhodnotit a popsat jejich výskyt vzhledem k poloze v okrese.
- Popsat distribuci zdravotnických dat v prostoru a odhalit pravidelnosti v prostorovém uspořádání zdravotních dat.
- Navrhnout regresní model, který dává do souvislosti aspekty životního prostředí a charakteristiky funkce štítné žlázy.
- Potvrdit nebo vyloučit vliv jednotlivých proměnných životního prostředí na funkci štítné žlázy.

## 2 Rešerše literatury

Již před naším letopočtem se lidé domnívali, že výskyt nemocí souvisí s životním prostředím. Hippokratés byl první, kdo se tomuto tématu věnoval, a proto je považován za zakladatele epidemiologie (Vrtišková 2012). Epidemiologie se zabývá výskytem nemocí v lidské populaci a zároveň studiem faktorů a podmínek, které výskyt podmiňují a ovlivňují. Tato věda taktéž řeší prevenci a potlačení jednotlivých nemocí (Geizerová a kol. 1995). Epidemiologie má velice blízko ke statistice, resp. využívá nástroje statistiky k určení závěrů. Statistika je vědní obor, který shromažďuje, zpracovává a interpretuje data, jež charakterizují dané zákonitosti a přitom podléhají náhodným vlivům (Bencko a kol. 2003).

Jedním z nejznámějších představitelů moderní epidemiologie je John Snow. Dokázal odhalit zdroj nákazy cholery díky porovnání mapy výskytu onemocnění s mapou vodních zdrojů v Londýně v roce 1854 (LSHTM 2016). Postupem času se vědci zabývali novějšími a novějšími nemocemi a epidemiemi. Jedním z velice diskutovaných témat byla a je Černobylská havárie. 26. dubna 1986 došlo v Černobylské elektrárně ke dvěma výbuchům a postupnému uvolňování vzácných plynů a těžkých štěpných produktů (Xe, Kr, J, Te a Cs). Radioaktivní oblak přecházel přes Švédsko, Finsko, střední a západní Evropu (Drábová a kol. 2001). Na území ČSSR byly prováděny testy půdy, vody, potravin a zemědělských produktů kvůli prevenci možného vystavení lidského organismu radioaktivitě (IHE 1987). Zamezit se tomu úplně nedalo. Došlo ke zvýšenému počtu potratů plodů v 8. - 12. týdnu, kdy převažovaly potraty mužského plodu (Peterka a Peterková 2004). Spad radioaktivního jódu v jednom roce po Černobylské havárii způsobil zvýšení rakoviny štítné žlázy hlavně u dětí (Knechtsberger nedatováno). V oblastech Bryansk, Kyjev a Zhitor je dokládán výskyt rakoviny štítné žlázy u 35,6 % dětí (Ashizawa a kol. 1997). Taktéž ve Velké Británii byl veden výzkum vlivu radioaktivního jódu na výskyt rakoviny štítné žlázy (Hill a kol. 1986).

Článek o Černobylském spadu a výskytu rakoviny ve Finsku (Kurtio a kol. 2013) se snaží zpětně dokázat vzájemnou korelaci rakoviny štítné žlázy a radioaktivního jódu. Na základě dat z národních statistik Finska a měření radioaktivity dokázali autoři vyloučit vliv radioaktivity na rakovinu prsu, plic a prostaty. Statistickou významnost prokázali u mladých mužů na rakovinu štítné žlázy. Výsledkem švédské studie, zabývající se taktéž tímto tématem, je opatrné prohlášení, že Černobylská havárie může za zvýšený nárůst rakoviny na území Švédska (Tondel a kol. 2004).

Termín životní prostředí v sobě nese prostorový kontext. Proto výzkumy o vlivu životního prostředí na jedince požadují prostorové informace a jejich následnou analýzu. Velice

vhodným nástrojem jsou pro tuto práci Geografické informační systémy (GIS), které jednotlivá data zvládají ukládat, připravit, analyzovat a následně zobrazit v prostorové podobě. Užitečnost GIS při řešení epidemiologických problémů byla prokázána v mnoha studiích (Grundloch 2009, Elliott a kol. 2001, Wang a kol. 2015). Pomocí nástrojů GIS je vhodné hodnotit např. možnost zvýšeného výskytu anomálií u narozených dětí, jejichž matky žijí v blízkosti skládek (Elliott a kol. 2001). V článku Wang a kol. (2015) byl využit GIS pro určení korelace znečištění ovzduší a výskytu bronchitidy ve městě Jinan v Číně. Výzkum popsany ve zmíněném článku byl omezen na případy hospitalizovaných pacientů do 18 let. Při analýze byla využita data o jejich bydlišti a pracovišti. Tato data byla vhodná pro zjištění možného vlivu přírodního prostředí na jedince. Ve finské práci, týkající se výskytu rakoviny štítné žlázy (Kurtio a kol. 2013), využívají GIS k vytvoření časových řad a jejich porovnání. V dalším článku Shahida a kol. (2014) korelují znečištění vody a závažná onemocnění žaludku pomocí GIS nástrojů. U této práce je korelace zjišťována plošně za městské části a nevypovídá nic o jednotlivcích. V článku Henke a Petropoulos (2013) se úspěšně daří prokázat korelace vzdálenosti rekreačních oblastí a délky dožití na území Velké Británie. Ve všech zmíněných článcích byly využity nástroje GIS, pomocí nichž se jim podařilo dokázat závislost mezi zkoumanými veličinami a odvodit vliv životního prostředí na jedince nebo skupiny lidí.

## 2.1 Výzkum v České republice

V České republice se problematice štítné žlázy a radioaktivity v životním prostředí významně věnuje Endokrinologický ústav v Praze. Pod jeho záštitou vznikly práce Tomanové (2005) a Grundlocha (2009). Rigorózní práce Tomanové (2005) se věnuje vlivu geologie a radioaktivního záření na výskyt dětské tyreopatie na testované lokalitě Příbramsko. Při zpracování dat v práci byly využity konkrétní hodnoty rastru k adresnímu bodu pacienta. Jelikož pacient (dítě) má pole pohybu nejčastěji v okolí bydliště a školy, je vhodné brát v potaz oba aspekty. Informace, kterou školu dítě navštěvuje, data neobsahují, ale na základě výpočtu nejkratší vzdálenosti mezi bydlištěm a adresou školy lze s jistou pravděpodobností určit, jakou školu dítě navštěvuje. V práci Tomanové (2005) i Grundlocha (2009) chybí podrobný popis zpracování a přípravy dat pro vstup do korelačních analýz. Kvalitní zpracování dat je důležité pro další postup v práci, aby nebyly výsledky negativně ovlivněny. Tomanová v práci jednotlivá data analyzovala a získala základní znaky popisné statistiky. Poté v práci použila Spearmanův koeficient pořadové korelace, který je posuzován na hladině  $p$  hodnoty 0.05.  $P$  hodnota je nejmenší hladina testu, na které zamítáme nulovou hypotézu (Netrdová nedatováno). Autorka zkoumala celý soubor i jednotlivé skupiny, které byly děleny na základě délky pobytu v dané obci, vzdálenosti od zlomu, typu geologického podloží a dávkového

příkonu gama záření. Důležitým dělicím aspektem byla hodnota jodurie. Tomanová soubor rozdělila na dvě skupiny - nízká hodnota jodurie, vysoká hodnota jodurie. Při použití korelační analýzy je takto možné daná data rozdělit. Dojde tím ale ke ztrátě velké části informací. Zároveň hodnoty jodurie nejsou nejvhodnější parametrem, na základě kterého lze usuzovat o funkci štítné žlázy. V předkládané diplomové práci jsou využity jiné metody, které dovolují využívat přímo naměřená data, která nejsou kategorizována, aby nedošlo ke ztrátě informací. V práci chybí zhodnocení prostorové závislosti a možnosti prostorových shluků. Toto se nevyskytuje ani v práci Grundlocha (2009). Analýzy prostorových vzorů jsou důležitým aspektem pro výsledek práce a mohou poukázat na nové skutečnosti, které předtím nebyly brány v potaz. Výsledky Spearmanova koeficientu pořadové korelace se ukázaly jako statisticky nezávislé. V porovnání s korelacemi ostatních proměnných nejlépe vyšla korelace jodurie a koncentrace uranu.

Grundloch (2009) se ve své práci zabýval územím Krucemburku, Radostína nad Oslavou, Dolních Rožinek a Budišova. Metodicky rozšířil práci Tomanové a využil Geografické informační systémy (GIS). Tyto systémy využil hlavně pro zpracování a přípravu dat. Tato práce byla rozšířena o množství vstupních dat, jako jsou intenzita dopravy, parametry pitné vody a atmosférické znečištění. Podařilo se zde prokázat některé slabé korelace prostředí na funkci štítné žlázy pomocí Spearmanova koeficientu pořadové korelace. V práci byly zjištěny slabé korelace mezi environmentálními aspekty (koncentrace antimonu, železa a arsenu v pitné vodě) a parametry štítné žlázy.

Z výsledků prací Tomanové (2005) a Grundlocha (2009) doposud nemůžeme jednoznačně prohlásit, jaké složky prostředí mají významný vliv na rakovinu štítné žlázy především z toho důvodu, že v pracích nebyl brán zřetel na pohyb pacientů. Proto předkládaná diplomová práce se snaží rozšířit zdrojová data a v souladu s tím zohlednit výsledný regresní model. V obou pracích nebyly provedeny analýzy prostorových vzorů. Tyto analýzy mohou poukázat na místa, u kterých je potřeba zaměřit se na jejich specifitu. Práce Tomanové (2005) a Grundlocha (2009) se zaměřily na korelace pouze jedné charakteristiky prostředí s jednou charakteristikou funkce štítné žlázy. Jelikož je lidské tělo vystaveno mnoha faktorům, které jej ovlivňují, je adekvátní brát v úvahu více složek a aspektů životního prostředí, stejně tak jako více charakteristik funkce štítné žlázy. I přesto byly práce Tomanové (2005) a Grundlocha (2009) značným přínosem při prostorové analýze endokrinologických dat. V současné době je možné využít větší množství prostorových analýz, a proto je možné hledat další souvislosti medicínálních a prostorových dat a podrobněji tyto souvislosti popisovat.

### 3 Prostorová analýza dat

Prostorové analýzy využívají soubor technik, které vznikaly v rozdílných oborech. Jejich společným cílem byla analýza dat s důrazem na jejich prostorové vztahy. Zabývají se hlavně nalezením nových vztahů mezi uspořádáním a jinými atributy nebo geoprvky ve studované oblasti a také modelováním těchto vztahů. Cílem je dosáhnout lepšího porozumění a předpovídání daných vztahů a uspořádání (Horák a kol. 2000). Jednotlivými prvky pro tento účel rozumíme geografické objekty a jiné objekty s prostorovou lokalizací, ať již fyzické nebo abstraktní povahy, často také události a jevy. Bez ohledu na konkrétní postup je zřejmé, že metody prostorové analýzy vyžadují přístup k atributům studovaných objektů i k informaci o jejich lokalizaci. Na rozdíl od jiných forem analýz tedy potřebují prostorové analýzy atributová data i geografickou lokalizaci objektů (Horák 2002). V dnešní době zaujímají prostorové analýzy významné postavení v celé řadě vědeckých oborů, v nichž byla rozpoznána jejich důležitá role k uchopení a analýze množství prostorových dat. Příčinou většího zájmu o prostorové analýzy dat v nedávné době je nárůst dostupnosti kvalitních prostorových dat a dále rozšíření použití GIS a dalšího softwaru (SPSS, Statgraphics, Simca 12+) vhodného pro práci s tímto typem dat (Spurná 2006). Podle Fortheringham, Roberson (1994) jsou tři hlavní cíle prostorové analýzy dat:

- popis událostí nebo objektů v geografickém prostoru (včetně popisu uspořádání) - 2. část práce
- interpretace procesů, které vedly k pozorovanému stavu uspořádání objektů nebo událostí ve sledovaném prostoru (systematický průzkum) - 3. část práce
- zlepšení schopnosti předpovídat a kontrolovat objekty nebo události v geografickém prostoru - 4. část práce.

#### 3.1 Metody a definice

Tato podkapitola popisuje základní pojmy a metody prostorové analýzy dat použité v diplomové práci.

##### **Metoda nejbližšího souseda**

Metoda nejbližšího souseda (Nearest neighbour distances) studuje vzdálenosti mezi body, a to mezi nejbližšími sousedními body. Metoda je založena na porovnávání pozorované průměrné vzdálenosti mezi nejbližšími sousedy a očekávané průměrné vzdálenosti. Očekávaná průměrná vzdálenost je vzdálenost mezi nejbližšími sousedy u náhodně rozmístěných bodů (Netrdová nedatováno). Pokud je průměrná vzdálenost mezi sousedy menší než průměr



za hypoteticky náhodnou distribuci, je analyzovaná distribuce považována za shluk. V případě, že je průměrná vzdálenost větší než průměrná vzdálenost hypoteticky rozmístěné distribuce, jsou analyzované prvky považovány za rozptýlené (ArcGIS Pro 2017). Pro výpočet je užíván vzorec 1:

$$ANN = \frac{Do}{De} \quad (1)$$

Kde **Do** je pozorovaná střední vzdálenost mezi jednotlivými body a **De** je očekávaná průměrná vzdálenost pro body s náhodným rozmístěním.

Součástí metody je výpočet z - skóre. Z – skóre jsou směrodatné odchylky. Na základě z – skóre a **p** hodnoty je rozhodováno, zda přijmout nebo zamítnout nulovou hypotézu (H0). V případě, že z – skóre je menší než -2.58 nebo větší než +2.58, je spolehlivost vyšší než 99 % (ArcGIS Pro 2017). Z – skóre se vypočítá pomocí vzorce 2:

$$z = \frac{Do - De}{SE} \quad (2)$$

kde:

$$SE = \frac{0.26136}{\sqrt{\frac{n^2}{A}}} \quad (3)$$

Hodnota **n** představuje celkový počet prvků a **A** je plocha minimálního opsaného obdélníku kolem všech bodů v analýze nebo plocha nastavená uživatelem.

**P** hodnota v tomto testu představuje pravděpodobnost, že pozorovaný prostorový vzor byl vytvořený náhodným procesem (ArcGIS Pro 2017).

### Ripleyho K – funkce

K-funkce poskytuje efektivní přehled prostorové závislosti událostí pro široký rozsah měřítek. U Ripleyho K-statistiky jsou nejprve vytvořeny buffer zóny kolem všech bodů o velikosti zvoleného prostorového kroku. Následně je zjištěn počet bodů v každé zóně. Poté je provedeno porovnání získaného rozdělení s náhodným rozdělením četností (Netrdová, nedatováno). Pokud je křivka hodnot naměřených vzdáleností **Ki** větší než očekávaná křivka **Kj**, má analyzovaná vrstva shlukové uspořádání. Pokud je křivka vzdáleností analyzované vrstvy

menší než očekávaná křivka, má vrstva rozptýlené uspořádání (ArcGIS Pro, 2017). Vzorec (4) pro výpočet Ripleyho K-statistiky je:

$$L(d) = \sqrt{\frac{A \sum_{i=1}^N \sum_{j=1, j \neq i}^N k(i, j)}{\pi N(N-1)}} \quad (4)$$

kde **A** je plocha, **N** je počet bodů, **d** je vzdálenost, a **k(i, j)** je váha.

U této funkce je možnost vážení určitou hodnotou, např. počtem úkazů v daném bodě. Pokud bod není vážený, je uživatelem zvolen počet permutací, který např. 99krát rozdělí náhodně 99 bodů do obalové zóny bodu pro každou iteraci. Následně je vypočítána hodnota **K**. Pokud je uživatelem zvolena možnost vážení, jsou pouze hodnoty vážení rozděleny do buffer zóny bodu. Poté je vypočítána hodnota **K** (ArcGIS Pro 2017).

### Lokální Moranovo I

Lokální Moranovo I (Local Moran's I) identifikuje prostorové seskupení prvků s vysokými nebo nízkými hodnotami. Metoda také identifikuje prostorové odchylky. K tomuto účelu nástroj vypočítá místní hodnotu Moranova I, z- skóre, p hodnotu a kód reprezentující typ clusteru pro každý statisticky významný rys (ArcGIS Pro 2017). Moranovo I se počítá pomocí vzorce 5:

$$I_i = \frac{x_i - X}{S_i^2} \sum_{j=1; j \neq i}^n w_{i,j} (x_j - X) \quad (5)$$

kde  $x_i$  je atribut funkce  $i$ ,  $X$  je průměr odpovídajících atributů,  $w_{i,j}$  je speciální váha mezi funkcemi  $i$  a  $j$ ,  $S_i^2$  je vypočítáno podle vzorce 6:

$$S_i^2 = \frac{\sum_{j=1; j \neq i}^n (x_j - X)^2}{n - 1} \quad (6)$$

kde  $n$  je celkový počet funkcí.

Výsledná hodnota **I** udává, jestli má funkce podobné okolní hodnoty. Při vysokých hodnotách **I** mají sousední funkce podobně vysoké nebo nízké hodnoty atributu. Naopak při nízkých hodnotách **I** mají sousední funkce odlišné hodnoty atributu. U obou případů musí být dostatečně malá  $p$  hodnota, aby se jednalo o významný výsledek na hladině 95 %.

Výsledkem celé metody je vznik pole značícího typ bodu. Může se jednat o významné seskupení vysokých hodnot (HH), nízkých hodnot (LL) a dále odlehlosti, ve kterých je vysoká

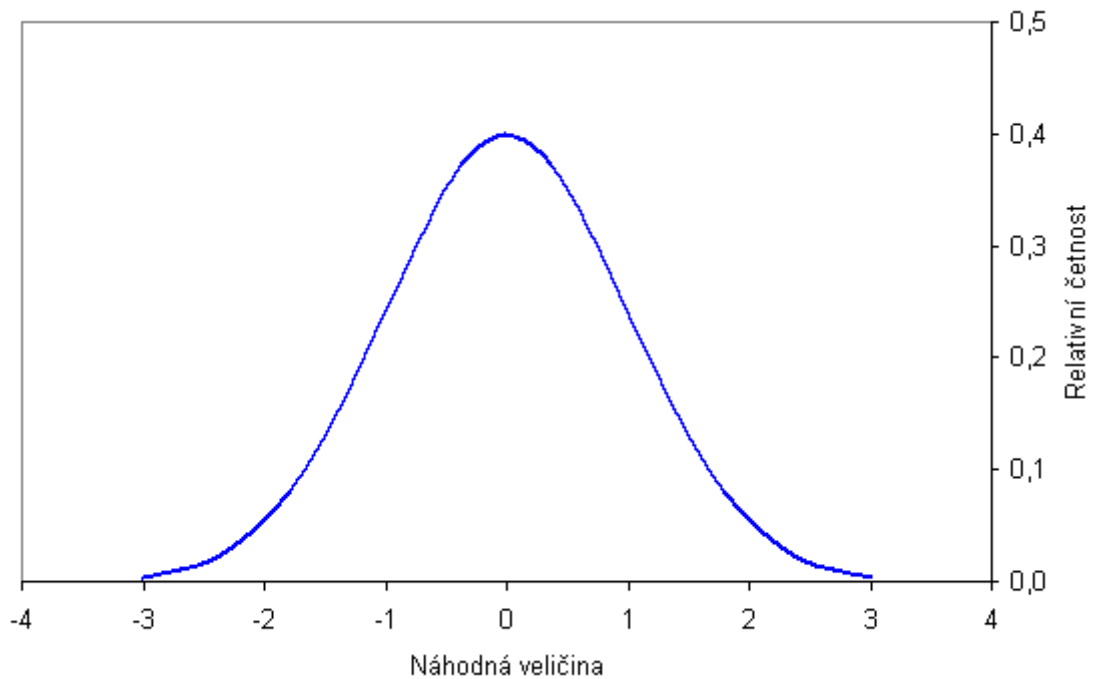
hodnota obklopena nízkými hodnotami (HL) anebo naopak, nízká hodnota obklopena vysokými hodnotami (LH).

### Normální rozdělení

Normální (Gaussovo) rozdělení je rozdělení pravděpodobnosti spojité náhodné veličiny. Má tvar křivky na Obr. 1 , která je definována Gaussovou funkcí (vzorec 7):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (7)$$

V dané funkci je důležitý průměr a směrodatná odchylka nebo rozptyl (Tomsa 2011).



**Obr. 1 Normální (Gaussovo) rozdělení**  
(Zdroj: Tomsa 2011)

### Mocninná transformace

Mocninná transformace je metoda, která se využívá pro zlepšení rozdělení dat. Vhodnou transformací dat je možné stabilizovat rozptyl, přiblížit šikmost rozdělení k nule a tvar k normálnímu rozdělení (Meloun, Hill a Militký 2000). Cílem je, na základě znalostí o výběru dat, nalézt vhodnou mocninu. Nejjednodušší je prostá mocninná transformace (vzorec 8 a 9):

$$\text{pro } P \neq 0 \quad hp(x) = \text{sign}(x) * \text{abs}(x)^P \quad (8)$$

$$\text{pro } P = 0 \quad hp(x) = \ln(x) \quad (9)$$

Kde **abs (x)** je absolutní hodnota a **sign (x)** je znaménková funkce:

$$\text{Sign}(x) = 1 \text{ pro } x > 0, \text{ sign}(x) = -1 \text{ pro } x < 0, \text{ sign}(x) = 0 \text{ pro } x = 0.$$

Tato transformace zachovává pořadí dat výběru a dochází k symetrizaci výběru (Militký a Meloun nedatováno).

### **Obousměrné ortogonální projekce do latentní struktury (O2PLS)**

Metoda obousměrných ortogonálních projekcí do latentní struktury (O2PLS) je jednou z variant metody parciálních nejmenších čtverců (PLS), či správněji projekcí do latentní struktury. O2PLS je vícerozměrná regrese s redukcí dimenzionality, která je schopna vysvětlit vztahy mezi maticí predikovaných proměnných  $Y$  a maticí prediktorů  $X$ , a to i v obráceném pořadí (Trygg a Wold 2003).

O2PLS je schopna pomocí lineárních kombinací rozložit matice  $X$  a  $Y$  tak, aby korelace mezi těmito maticemi byla maximální, a zároveň zajistit snadnou interpretovatelnost vzniklého modelu. O2PLS tvoří dvě nezávislé, ortogonální skupiny komponent, a to komponenty prediktivní sdílející rozptyl jak s prediktory v matici  $X$ , a vysvětlující strukturu vztahů mezi  $X$  a  $Y$ , tak s predikovanými proměnnými v matici  $Y$ , a na druhé straně komponenty ortogonální, vysvětlující variabilitu uvnitř matice  $X$  nezávisle na matici  $Y$  a variabilitu uvnitř matice  $Y$  nezávisle na matici  $X$  (Trygg a Wold 2003). Zatímco komponenty prediktivní jsou cílem analýzy a bývají interpretovány, komponenty ortogonální jsou z hlediska cílů O2PLS nezajímavé. Nicméně i ortogonální komponenty jsou zpravidla nezbytnou součástí celého O2PLS modelu.

Postup analýzy sestává v prvním kroku z ověření předpokladů normality rozdělení proměnných a reziduí, jejich konstantního rozptylu (homoskedasticity) a homogenity rozdělení dat a reziduí. Teprve po splnění těchto předpokladů lze data analyzovat. Data vyhovující výše uvedeným předpokladům jsou obvykle normována. Od měřené hodnoty je odečten průměr a výsledek je vztažen ke směrodatné odchylce. Teprve takto upravená data jsou podrobena analýze.

Metoda O2PLS, podobně jako jiné varianty PLS, je schopna pracovat s daty zatíženými vysokým stupněm multikolinearity uvnitř matic  $X$  a  $Y$ . Naopak od regresních metod bez redukce dimenzionality (např. vícenásobná regrese), které multikolinearita deformuje, v případě O2PLS multikolinearita zesiluje prediktivitu a při dostatečně vysokém stupni

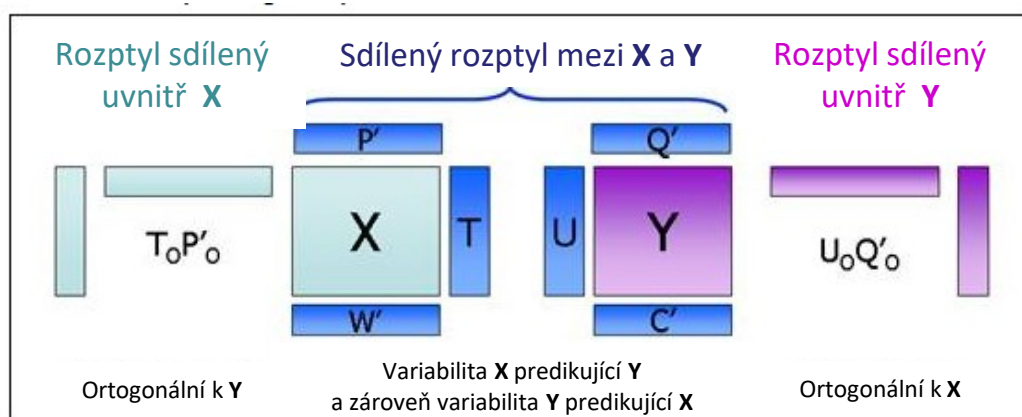
multikolinearity si lze představit i fungující modely O2PLS o počtu proměnných převyšujícím počet měření, či zahrnutí proměnných, které jsou ve vzájemném funkčním vztahu.

O2PLS model lze vyjádřit takto (vzorec 10 a 11):

$$X = T_p \cdot P'_p + T_0 \cdot P'_0 + E \quad (10)$$

$$Y = U_p \cdot Q'_p + U_0 \cdot Q'_0 + F \quad (11)$$

Kde  $X$  je matice  $s \times k$  prediktory a  $i$  subjekty,  $Y$  je matice  $s \times l$  závisle proměnnými a  $i$  subjekty.  $T_p$  je matice komponentních skóre prediktivních komponent extrahovaných z matice  $X$  a  $T_0$  je matice komponentních skóre komponent ortogonálních extrahovaných z matice  $X$ .  $P_p$  je matice komponentních vah prediktivních komponent extrahovaných z matice  $X$  a  $P_0$  je matice komponentních skóre komponent ortogonálních extrahovaných z matice  $X$ .  $E$  představuje matici chyb (reziduí) extrahovaných z matice  $X$ . Analogicky,  $U_p$  je matice komponentních skóre prediktivních komponent extrahovaných z matice  $Y$  a  $U_0$  je matice komponentních skóre komponent ortogonálních extrahovaných z matice  $Y$ .  $Q_p$  je matice komponentních vah prediktivních komponent extrahovaných z matice  $Y$  a  $Q_0$  je matice komponentních skóre komponent ortogonálních extrahovaných z matice  $Y$ .  $F$  představuje matici chyb (reziduí) extrahovaných z matice  $Y$  (Trygg, Holmes a Lundstedt 2007).



**Obr. 2 O2PLS modelující dvě datové tabulky**  
(Zdroj: Eriksson, Wold a Trygg nedatováno)

Relevance jednotlivých prediktorů je testována statistikami Variable Importance pro  $j$ -té proměnné  $k$ -té komponenty ( $VIP_{j,k}$ ). Je-li alespoň jedno z kritérií  $VIP_{j,k}$  významné, prediktor se v systému ponechává jako relevantní, čímž však není ještě zaručena významnost komponentních vah prediktivních komponent tohoto prediktoru. Proměnná může být relevantní pro výstavbu modelu, avšak zároveň významná, či nevýznamná ve vztahu ke komponentním vahám prediktivních komponent. Relevance komponent je hodnocena

metodou parciálních reziduálních součtů čtverců (Partial Residual Sum of Squares, PRESS). Model O2PLS se dokáže vyrovnat s mírným množstvím chybějících údajů, pokud jich není více jak 50 % v daném řádku či sloupci matice  $X$  či  $Y$  (Madsen, Lundstedt a Trygg 2010). A analýza O2PLS byla provedena s využitím software SIMCA-P+ verze 12.0.0.0 od firmy Umetrics AB (Umeå, Švédsko)

### **Variable importance (VIP)**

Metoda VIP je určena k odhadnutí významnosti proměnných v projekci modelu O2PLS nebo OPLS. Na základě velikosti VIP se určuje setrvání proměnné v modelu nebo jejího vyloučení (VIP nedatováno). Při hodnotě VIP větší než 1, proměnná v modelu zůstává, při hodnotě VIP menší než 1, je proměnná méně významná a může být vhodným kandidátem na odebrání z modelu. Takto striktně se nelze vždy rozhodovat. VIP hodnota se vždy mění po předělání modelu.

### **Křížová validace**

Při zpracování výsledků metody O2PLS je vypočítána křížová validace (Cross Validation). Jedná se o metodu zjišťující, jak moc budou nezávislé vzorky dat ovlivňovány výsledky statistické analýzy. Křížová validace rozdělí soubor dat do dvou skupin. Jedna skupina je určena pro trénování a druhá pro testování klasifikátoru. Trénovací soubor je využit při tvorbě klasifikátoru (Malá 2013). Testovací soubor slouží k ohodnocení klasifikátoru. Trénovací a testovací soubory by neměly mít ani jeden společný prvek. To znamená, že data, sloužící k hodnocení klasifikátoru, nejsou přítomna při tvorbě klasifikátoru. Tento postup je pak několikrát opakován. Vyhodnocení klasifikace je bráno jako průměr všech chyb predikce na rozdílných testovacích souborech (Clarke a kol. 2009).

## **3.2 Návrh metodiky**

Postup práce je rozdělen do čtyř bodů, které postupně vedou k naplnění cílů diplomové práce.

### **1. Příprava dat**

Správné využití metod prostorových analýz není možné bez dostatečného poznání té části světa, kterou popisujeme, zkoumáme, analyzujeme nebo modelujeme (Horák 2002). Nejprve bylo třeba vybrat vhodná data pro analýzu. Výběr dat proběhl ve spolupráci s MUDr. Jiřím Němečkem a Endokrinologickým ústavem v Praze. Následně proběhla příprava jednotlivých typů dat pro vstup do pozdějších analýz. K práci byla získána data bodová, liniová i plošná.

U bodových vrstev je využita metoda interpolace prostorových dat IDW. U liniových vrstev je brána v potaz vzdálenost linie od místa bydliště pacienta. U plošných dat je využita konkrétní hodnota místa bydliště a hodnota v místě školy.

## **2. Mapování a popis výskytu nemoci**

Jednotlivá data jsou pro přehlednost zobrazena pomocí map. Jedná se o:

- výskyt nemoci,
- znečištění ovzduší na základě dat od ČHMÚ a silniční sítě,
- dávkový příkon gama záření,
- jednotlivé zlomy v oblasti,
- biomonitoring mechů,
- rozbor vody ve vodovodech.

Využity jsou základní metody popisné statistiky u zdravotních dat. Následně byl vizuálně porovnán výskyt zhoršené funkce štítné žlázy s prostorovými charakteristikami životního prostředí.

## **3. Analýza prostorových vzorů**

V této části je zkoumán výskyt zhoršené funkce štítné žlázy. Nejprve jsou využity metody explorační analýzy jako geografický střed, směrodatná vzdálenost a směrodatná elipsa. Tyto metody popíší výskyt pacientů se zhoršenou funkcí štítné žlázy. Následně jsou využity metody konfirmační analýzy, a to metody nejbližšího souseda a Ripleyho K-funkce. Aplikována byla také metoda lokální Moranovo I. Jsou zde zkoumány a vizualizovány prostorové vzory ve výskytu zhoršené funkce štítné žlázy. Výsledkem této části je odpověď na otázku, zda u výskytu zhoršené funkce štítné žlázy dochází k prostorovému shlukování. Výskyt shluků upozorňuje na zaměření na jednotlivé lokality a jejich specifičnost, která může být zohledněna při tvorbě regresního modelu.

## **4. Analýzy výskytu nemoci s vnějšími faktory**

Část čtyři se zabývá identifikací možných vztahů výskytu zhoršené funkce štítné žlázy a vnějších příčin. Tato část využívá metody O2PLS, Hottelingovy  $T^2$  statistiky, Variable importance a křížové validace. Je zde vytvořen regresní model, který pomocí dvou komponent popisuje možné vlivy životního prostředí na charakteristiky funkce štítné žlázy. Následně jsou pomocí mapových výstupů znázorněny významné parametry životního prostředí a parametry funkce štítné žlázy.

## 4 Charakteristika území a vybraných dat

Aby byly naplněny cíle práce, bylo nezbytné vybrat vhodná data. Tato data musela splňovat:

- zahrnutí zájmového území,
- charakterizovat funkce štítné žlázy,
- charakterizovat prostředí okresu Příbram, ve stejné nebo blízké době jako byla získána data o funkci štítné žlázy,
- volně k dispozici z internetu nebo po dohodě s organizací pořizující data.

Data, splňující tyto požadavky, byla získána od Endokrinologického ústavu, Českého hydrometeorologického ústavu, Výzkumného ústavu Silva Taroucy pro krajinu a okrasné zahradnictví, Ředitelství silnic a dálnic, Státního zdravotního ústavu a České geologické služby.

V první podkapitole této části práce je charakterizováno zájmové území okresu Příbram. Jsou zde popsány jednotlivé společnosti, které poskytly data a jejich činnost. V druhé podkapitole je charakterizována geologie okresu Příbram. Geologická data pro tuto práci poskytla Česká geologická společnost. V závěrečné části této kapitoly je stručně definována vědní disciplína endokrinologie. Jsou zde popsány činnosti Endokrinologického ústavu, který provedl výzkum, na jehož základě vznikla data použitá v této práci.

### 4.1 Zájmové území Příbramsko

Zájmové území Příbramsko bylo vybráno z důvodu dostupnosti dat v této oblasti. Území je vymezeno statistickou jednotkou okresu Příbram. Okres Příbram, jak je vidět na Obr. 3, se nachází v jihozápadní části Středočeského kraje. Hlavním sídlem je město Příbram. Rozloha okresu je 1 563 km<sup>2</sup>, z toho 47.6 % plochy zabírá zemědělská půda. V okrese žije 114.2 tisíc obyvatel. Hustota zalidnění je 73.1 obyvatel na km<sup>2</sup>. Tím se řadí mezi tři nejhustěji osídlené okresy v kraji. Okres se dělí na tři správní obvody (Příbram, Sedlčany a Dobříš). Jeho součástí je 120 obcí (ČSÚ 2016).

Příbramsko je členitým regionem se značným zalesněním krajiny. Lesy zabírají plochu 40.4 %. Jedná se o nejvíce zalesněný okres ve Středočeském kraji. Na západě a severozápadě se rozprostírají Brdy, které tvoří přirozenou hranici regionu. Ze širokých údolí krajina pozvolně stoupá na táhlé hřebeny a plošiny. Ty přesahují výšku 800 m. Nachází se zde Tok (865 m n. m.) - nejvyšší hora středních Čech. Významný je také vrch Praha (862 m n. m.) a Třemšín (827 m n. m.). Nejnížší bod je v blízkosti soutoku říčky Křeničná s Vltavou a má nadmořskou

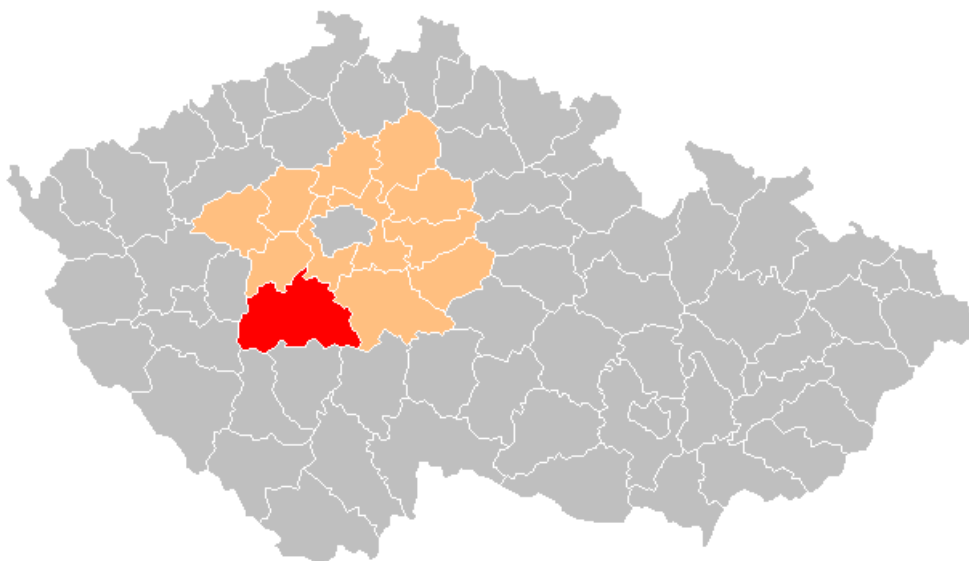


výšku 271 m. Řeka Vltava dělí okres na dvě nestejně velké části. Nachází se zde přehrada Orlík, která je co do objemu vody největší přehradou v České republice, a přehrada Kamýk.

Příbramsko je oblíbenou turistickou destinací. Je zde chráněná krajinná oblast Brdy a národní přírodní rezervace Drbákov - Albertovy skály. Jedná se o jeden z nejzachovalejších ekosystémů pro střední Vltavu, ve kterém jsou zajímavé skalní srázy a zachovalé suťové lesy s hojným výskytem tisu. Dále je na Příbramsku celá řada drobných přírodních památek. Z těchto důvodů je Příbramsko oblíbeným turistickým cílem.

Historie území je velice úzce spjata s hornickou činností. Těžily se zde hlavně uranové rudy, které do současnosti přinesly ekologické problémy. Uvnitř veselého kopce u Mokrska se nachází rozsáhlé ložisko zlata, které by ale bylo náročné vytěžit. V současnosti se na Příbramsku těží hlavně kámen a písky. Dále je zde velice rozvinutý strojírenský, potravinářský průmysl a hutnictví neželezných kovů (ČSÚ 2016).

Příbramsko je díky dálnici D4 napojeno na hlavní město Prahu. Územím dále prochází silnice I. třídy - I/4, I/18, I/19 a I/66. Vede zde také důležitá železniční trať 200 Zdice – Protivín. Další železniční tratě mají spíš lokální význam.



**Obr. 3 Vymezení okresu Příbram**  
(Zdroj: volné dílo)

#### **4.1.1 Topografická data**

Pro získání základních informací o území jsou využita data Arc ČR 500. Byla použita vrstva okres, obce, obce\_b, silnice. Vrstva okres byla využita pro definici hranic zájmového území. Vrstva obcí byla potřebná při přípravě dat, které charakterizovaly znečištění vod, a také

při vypracování druhé části metodiky. Pro stejný účel byla použita i vrstva obce\_b. Vrstva silnic byla využita při vytvoření nové vrstvy s charakteristikami o sčítání dopravy.

#### **4.1.2 Výzkumný ústav Silva Taroucy pro krajinu a okrasné zahradnictví (VÚKOZ)**

VÚKOZ je veřejnou výzkumnou institucí zřízenou v roce 2007 Ministerstvem životního prostředí ČR. Účelem tohoto ústavu je výzkum všech typů krajiny a souvisejících environmentálních rizik, výzkum biologické rozmanitosti a její ochrany, odborné podpory ochrany přírody a péče o krajinu a další (VÚKOZ, 2016).

Ústav vede velké množství projektů. Jedním z nich byl i projekt Určení zátěže krajiny ČR aktuálními úrovněmi atmosférické depozice a zajištění dat pro celoevropský biomonitorovací program 2000. Projekt se zabýval biomonitoringem mechů, u kterých se zjišťovala koncentrace 36 prvků. Z tohoto projektu jsou využita naměřená data, která se týkají oblasti Příbramsko. Data z jednotlivých projektů bývají publikována v rámci článků. Z každého projektu je závěrečná zpráva uložena v knihovně Ministerstva životního prostředí ČR.

#### **4.1.3 Český hydrometeorologický ústav (ČHMÚ)**

Aby byly informace o prostředí co nejkomplexnější, práce využívá data, která se týkají znečištění ovzduší od Českého hydrometeorologického ústavu. Jeho účelem je vykonávat funkci ústředního orgánu pro obory čistota ovzduší, hydrologie, jakost vody, klimatologie a meteorologie jako objektivní odborné služby poskytované přednostně pro státní správu České republiky. Hlavním předmětem činnosti je zřízovat a provozovat měřicí stanice a síť s využíváním telekomunikačních sítí, odborně zpracovávat výsledky pozorování, měření a monitorování, poskytovat předpovědi a výstrahy a vytvářet a spravovat databáze (Portál ČHMÚ, 2016). Právě v těchto databázích lze nalézt velké množství dat a informací.

Od Českého hydrometeorologického ústavu (ČHMÚ) jsou volně k dispozici data, týkající se znečištění ovzduší na jednotlivých měřicích stanicích za jednotlivé roky. V práci jsou využita data za období let 2000 až 2005, která bylo nutné zpracovat a připravit pro následné analýzy. Dále jsou na internetových stránkách ČHMÚ k dispozici data o znečištění vnějšího ovzduší. Úroveň znečištění je způsobena vypouštěním znečišťujících látek z různých zdrojů v důsledku lidské činnosti. Jedná se například o dopravu, průmyslovou výrobu a další. Data o znečištění vnějšího ovzduší jsou publikována v pětiletých průměrech, a to až od roku 2007. V práci jsou tedy použita data zprůměrovaná za rok 2007 až 2011.

#### **4.1.4 Ředitelství silnic a dálnic (ŘSD)**

ŘSD je příspěvková organizace zřizovaná Ministerstva dopravy ČR. Základním cílem této organizace je výkon vlastnických práv státu k nemovitým věcem tvořících dálnice, silnice I. třídy, zabezpečení správy, údržby a oprav dálnic a silnic I. třídy a zabezpečení jejich výstavby a modernizace (ŘSD ČR, 2016).

Na internetových stránkách této organizace jsou data o sčítání dopravy za jednotlivé úseky silnic a dálnic v České republice. Toto sčítání je prováděno každý rok. Data byla použita jako další z možných proměnných při hledání dopadů na funkci štítné žlázy.

#### **4.1.5 Státní zdravotní ústav (SZU)**

Další důležitou složkou jsou rozbory vody za jednotlivé vodovody na Příbramsku. Tato data jsou získána od Státního zdravotního ústavu ze sekce ochrany a podpory veřejného zdraví. Státní zdravotní ústav je příspěvkovou organizací ministerstva zdravotnictví. Ústav je zřízen za účelem přípravy podkladů pro národní zdravotní politiku, pro ochranu a podporu zdraví, k zajištění metodické a referenční činnosti na úseku ochrany veřejného zdraví, k monitorování a výzkumu vztahů životních podmínek a zdraví, k mezinárodní spolupráci, ke kontrole kvality poskytovaných služeb k ochraně veřejného zdraví, k postgraduální výchově v lékařských oborech ochrany a podpory zdraví a pro zdravotní výchovu obyvatelstva. Jeho úkolem je sledovat dlouhodobé trendy šíření infekčních nemocí, ohrožení nemocí z povolání, epidemiologii drogových závislostí a zjištěné informace dále předávat orgánům ochrany veřejného zdraví (Státní zdravotní ústav, 2016).

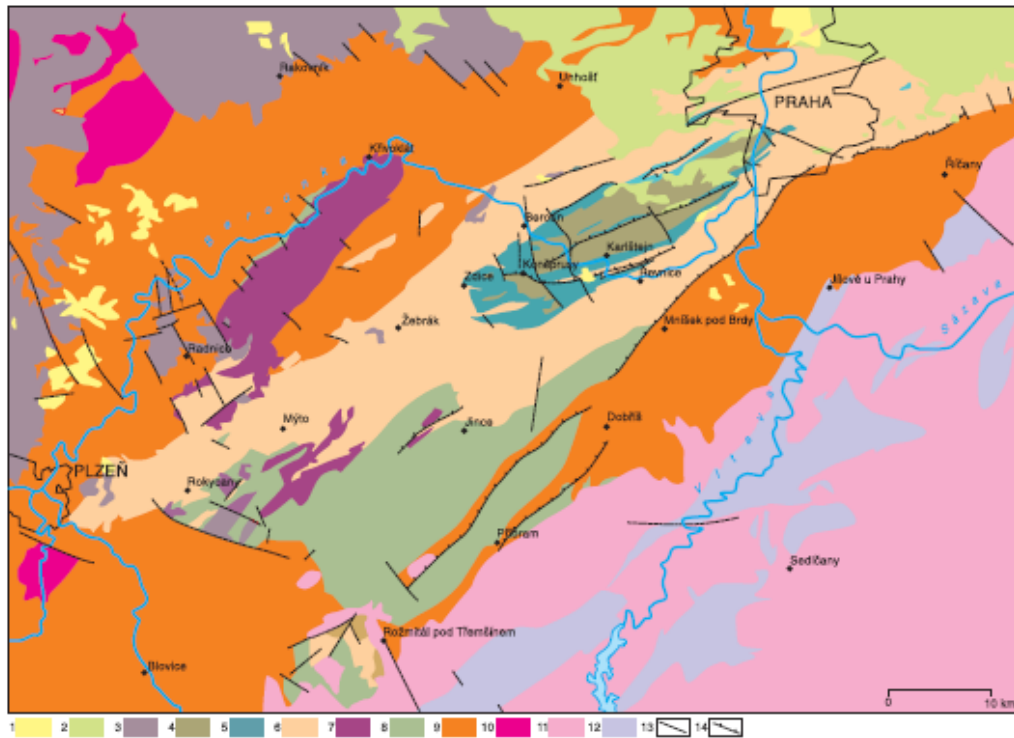
Data, publikovaná SZÚ na jeho internetových stránkách, většinou neobsahují bližší prostorové určení. Pouze data o nemocech z povolání a výskyt a šíření HIV/AIDS jsou reprezentována na úrovni krajů. Ústav zpracovává také data sbíraná a předkládaná jednotlivými hygienickými stanicemi a následně z nich vyvozuje závěry.

Pro účely diplomové práce byla využita data z analýz pitné vody z deseti vodovodů v okrese Příbram. Pro hodnocení jakosti pitné vody je závazným podkladem vyhláška Ministerstva zdravotnictví České republiky č. 252/2004 Sb., která je plně v souladu s Evropskou směrnicí o jakosti pitné vody určené pro lidskou spotřebu.

## **4.2 Geologie Příbramska**

Na území Příbramska vystupují dva základní typy horninového prostředí, a to sedimentární horniny Barrandienu v západní části zájmové oblasti a horniny středočeského plutonického

komplexu východně od Příbrami. Na Obr. 4 je názorně vidět, jaké horniny se na území vyskytují.



**1** neogénní kontinentální sedimenty; **2** svrchnokřídové kontinentální a mořské sedimenty; **3** svrchnokarbonské a permské kontinentální sedimenty; **4** devonské marinní sedimenty; **5** sílurské marinní sedimenty a vulkany; **6** ordovické marinní sedimenty a vulkany; **7** svrchnokambričské vulkanické horniny; **8** kambričské kontinentální a mořské sedimenty; **9** proterozoické marinní sedimenty a vulkany; **10** prevartiské plutony; **11** variské plutony; **12** proterozoické a paleozoické metasedimenty a metavulkany; **13** zóny středočeského plutonu; **14** významné směrné přesmyčky.

**Obr. 4 Geologická mapa Příbramska**  
(Zdroj: Kachlík 2003)

Západní a severní část Příbramska je tvořena sedimenty proterozoického (starohorního) a kambričského (nejstarší období prvohor) stáří (Chlupáč 2011, Kachlík 2003). Sedimentární horniny jsou převážně tvořeny tmavými až černými jílovými břidlicemi, prachovci a droby. Jsou zde nalezeny droby s čůčkami konglomerátů tzv. dobříšské slepence (Chlupáč 2011). Hlavním sedimentačním prostorem je příbramsko-jinecká pánev, která je vyplněna kontinentálními sedimenty spodního kambria (pískovce, slepence) a mořskými sedimenty středního kambria (prachovce, břidlice) (Geologický park 2016). V prostoru dnešních Brd a Příbrami se v době spodního kambria rozkládalo rozsáhlé sladkovodní jezero. Na dně se usazovaly písky a další hrubozrnné usazeniny, které daly vzniknout dnešním drobům, pískovcům a slepencům.

Tato oblast je ovlivněna značnou intenzitou poruchových pásem, zvláště dvou hlubinných zlomů - klatovského a jáchymovského. Sedimenty svrchního proterozoika a spodního kambria západně od Příbrami jsou od sebe odděleny nejdůležitější poruchou v oblasti tzv. jílovou

rozsedlinou. Jádro tohoto zlomu je tvořeno šedým jílem o mocnosti až 20 cm. Tento zlom je z části sledován i říčkou Litavkou.

Jižní, jihovýchodní a východní část Příbramska tvoří hlavně magmatické horniny středočeského plutonu, který se vyskytuje na hranici Tepelsko-Barrandienské a Moldanubické jednotky. Vyskytují se zde středně biotitické až amfibol - biotitické granodiority, hrubozrnný biotitický granit a křemenný diorit (Chlupáč 2011). Středočeský pluton souvisí se vznikem významných rudných ložisek na Příbramsku. Dále se projevuje výrazně zvýšenou regionální radioaktivitou a velkou variabilitou jednotlivých horninových typů (Chlupáč a kol 2002). Hrubozrnný biotitický granit vykazuje hodnoty dávkového příkonu gama záření 75 - 125 nGy/h, křemenný diorit 35-50 nGy/h a středně biotitické až amfibol - biotitické granodiority 70 – 100 nGy/h. Sedimenty brdského kambria dosahují hodnot 15 - 35 nGy/h (Matolín 2000). Nejvyšší objemové aktivity radonu jsou v granitech a granodioritech. V roce 1986 nebylo Příbramsko významně zasaženo radioaktivním spadem po havárii jaderné elektrárny Černobylu (Matolín 2000).

Geologickou zvláštností příbramské oblasti je výskyt rudných žil s rozsáhlými akumulacemi rud Ag, Pb, Zn, Sb, U a Fe, které podnítily rozsáhlou těžbu. Rudné a uranové doly v této oblasti patří mezi mineralogicky světově proslulé lokality. Zdejší rudné žíly poskytly více než 300 různých druhů minerálů často v kvalitních vzorcích světové úrovně.

#### **4.2.1 Geologická data**

Geologická data byla získána od České geologické služby (ČGS). ČGS je organizace zřízená Ministerstvem životního prostředí ČR. Jejím úkolem je sběr, zpracování a poskytování údajů o geologickém složení státního území, o ochraně a využití přírodních nerostných zdrojů a zdrojů podzemních vod a o geologických rizicích. Tyto údaje předává správním orgánům pro politická, hospodářská a ekologická rozhodování (Česká geologická služba, 2016). ČGS poskytuje data pro online mapy, jako je Geologická mapa 1 : 50 000, Geomapa 1 : 25 000, Radonové mapy, Registr svahových nestabilit a další. Pro získání dat pro soukromé účely je nejdříve nutné o konkrétní data požádat pomocí kontaktního formuláře.

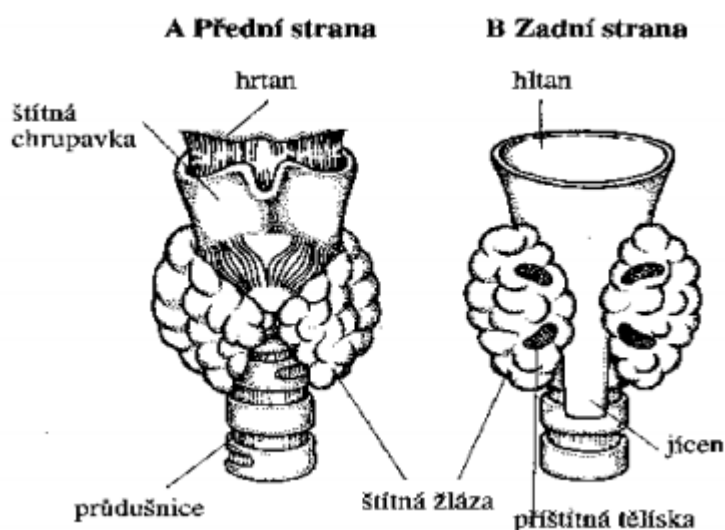
V práci byla využita digitální geologická mapa v měřítku 1 : 50 000 (GeoČR50). Mapa pokrývá celé území okresu Příbram. Byly využity informace o zlomech, dávkovém příkonu gama záření, rozložení hornin a jejich radonovém indexu.

### 4.3 Endokrinologie

Nedílnou součástí této práce je vědní obor endokrinologie. Endokrinologie je část medicíny, která se zabývá diagnózou a léčbou hormonálních poruch. Na poruchu tvorby hormonů orgánem štítné žlázy se zaměřuje tato práce.

Endokrinní systém je kontrolním systémem endokrinních žláz, které vylučují hormony. Ty cirkulují v těle pomocí krevního oběhu a ovlivňují vzdálené orgány. Účinky vylučovaných hormonů je nezbytné kontrolovat. Pokud by byl účinek moc dlouhý nebo příliš intenzivní, mohl by způsobit narušení metabolické rovnováhy. Řídící mechanismus pracuje na principu zpětné vazby, kdy je sledován výsledný efekt procesu. Na základě zjištěných údajů se upravují faktory, které akci iniciovaly (Lidské tělo, 1996).

Štítná žláza je součástí endokrinního systému. Její tvar a popis částí je znázorněn na Obr. 5. Jedná se o důležitý lidský orgán, který produkuje hormony ovlivňující rychlost látkové přeměny. Hormony zvyšují spotřebu kyslíku a mají vliv na růst a vývoj jedince. Štítná žláza udržuje optimální úroveň metabolismu v tkáních lidského těla pro jejich normální funkci (Baranovičová 2015).



**Obr. 5 Štítná žláza**  
(Zdroj: Baranovičová 2015)

Štítná žláza je umístěna v přední straně krku pod štítnou chrupavkou. Po obou stranách je ohraničena průdušnicemi. Na zadní straně štítné žlázy jsou uložena příštítná tělíska. Štítná žláza je tvořena dvěma laloky vejčitého tvaru, které jsou spojeny můstkem. Horní konec štítné žlázy se nachází pod úrovní hmatatelného výstupku hrtanové chrupavky. U zdravého člověka je průměrná hmotnost štítné žlázy 15 – 20 g.

Funkce štítné žlázy je řízena hormonem thyreotropin (TSH). Ten se tvoří v předním laloku hypofýzy. Slovo hormon je odvozené od řeckého slova „hormao“, což znamená „řídím“. Hormony jsou chemické struktury, které se vyznačují vysokou účinností i v malých koncentracích. Základní jednotkou je folikul. Zde dochází k vychytání jódu, uskladnění a tvorbě dvou hormonů - thyroxinu (T4) a trijódthyroninu (T3). Jejich uvolňování stimuluje Thyreotropní hormon (TSH). Bez těchto dvou hormonů nemůžeme dlouhodobě přežít. Štítná žláza produkuje ještě třetí hormon kalcitonin, který udržuje hodnotu vápníku v těle. Aby štítná žláza produkovala hormony, potřebuje dostatečné množství jódu, který se vyskytuje v potravě. Doporučená denní dávka jódu je přibližně 150 - 200 µg/l (WHO 2001). V České republice trpělo v roce 1990 velké množství lidí jeho nedostatkem. Proto se začal přidávat do soli a minerálních vod (Čeřovská 2007).

Hormony štítné žlázy se účastní regulace metabolismu lipidů a sacharidů, stimulují spotřebu kyslíku. Jsou nezbytné pro normální růst a zrání. Štítná žláza a její funkce není nezbytná pro život, ale její nepřítomnost se projeví celkovým zpomalením všech tělesných funkcí a způsobuje mentální a fyzické poruchy, u dětí retardaci a omezení růstu. Nadbytečná produkce hormonů štítné žlázy vede k hyperaktivitě. Člověk nemůže spát, třesou se mu prsty apod. (Endokrinní systém 2016).

### **4.3.1 Endokrinologická data**

Endokrinologická data byla dána k dispozici Endokrinologickým ústavem (EÚ). Základním úkolem EÚ je vysoce specializovaná preventivní, diagnostická, léčebná a vědecko-výzkumná činnost hlavně v oblasti endokrinologie a v dalších oborech, které s ní souvisí. Jedná se například o nukleární medicínu, klinickou biochemii, klinickou imunologii a onkologii hormonálně dependentních tkání (Agendový portál, 2016).

V roce 1990 trpělo téměř 160 milionů osob nedostatečným zásobením jódem. Nedostatek jódu je považován za příčinu fyzické a mentální retardace. Z tohoto důvodu se v České republice začal přidávat jód do soli, zavedlo se užívání jódové soli v potravinářských výrobcích, obohacovaly se krmné směsi hovězího dobytka a drůbeže, dále se zvýšila konzumace minerálních vod a stolních vod s jódem. Toto zvyšování jódu realizovala Mezirezortní komise pro šetření jódového deficitu. Jelikož se ale jednalo o téměř neregulovatelné zvýšení příjmu jódu, nastalo riziko vzniku thyreopatií z nadměrného přísunu jódu (Čeřovská 2007).

Z tohoto důvodu byla v letech 1991 - 2002 a v letech 2004 - 2006 prováděna výzkumná šetření, zabývající se monitorováním zdravotního stavu obyvatelstva se zaměřením na thyreopatii. V letech 1991 - 2002 probíhalo šetření po celé České republice. V letech

2000 - 2006 probíhalo šetření pouze ve třech částech republiky. Jednou z částí bylo Příbramsko, kde probíhalo šetření v roce 2000 a 2005.

Osoby, u kterých šetření proběhlo, byly vybrány náhodným výběrem z registru obyvatel jednotlivých regionů. Kritériem pro výběr bylo bydliště v regionu a věková kategorie. Dětská věková kategorie byla 6 - 17 let. Bylo vybráno 270 chlapců a 270 dívek. K vyšetření se dostavilo v roce 2000 291 osob a v roce 2005 274 osob. Účastníci byli zváni k vyšetření zvacím dopisem, ke kterému byl přidružen dotazník zjišťující socioekonomické charakteristiky, výživové návyky a další. U pacientů bylo provedeno vyšetření jodurie metodou založenou na Sandell - Kolthoffově reakci a vyšetření selenu použitím instrumentální neutronové aktivační analýzy (Čeřovská 2007).

Obě skupiny dat byly získány od pracovníků Endokrinologického ústavu v Praze ve formátu tabulek \*.xls. Získaná data obsahovala polohovou informaci pouze o dané části města, kde jedinec bydlí. Aby byla následná prostorová analýza dat provedena kvalitně, bylo nutné získat konkrétní adresy jedinců. Tyto adresy byly získány opisem na základě osobního čísla jedince a později byly na základě adresy určeny geografické souřadnice. Jelikož se jedná o citlivá zdravotnická data s konkrétním určením adresy pacienta, je nutné zvláštní zacházení s daty.



## 5 Příprava dat

V této části práce je podrobně popsán postup přípravy různých typů dat. Nejprve je vysvětleno zpracování dat environmentálních, následně geologických dat a poté dat medicínálních. V poslední části kapitoly je popsán postup pro získání dat s normálním rozdělením, aby data mohla vstupovat do pozdější analýzy. Zpracování bylo provedeno v programu Microsoft Excel 2010, ArcGIS Desktop a IBM SPSS Statistics 23. V programu ArcGIS Desktop byly vytvořeny mapy pro znázornění jednotlivých faktorů.

### 5.1 Environmentální data

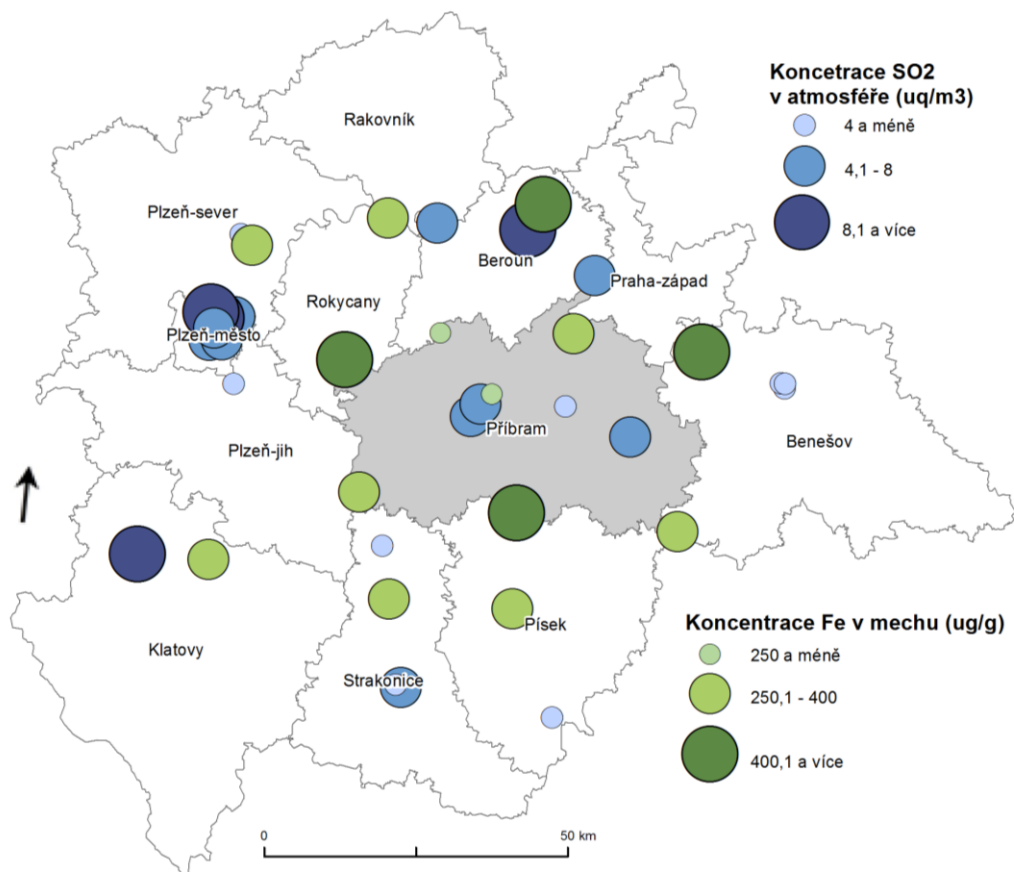
Tato podkapitola popisuje zpracování environmentálních dat. Konkrétně se jedná o naměřená data z jednotlivých stanic a také o data o znečištění ovzduší, dále data naměřená při biomonitoring zátěže přírody, následně o data analyzující pitnou vodu a v poslední řadě data zaznamenávající intenzitu dopravy. Tato data splňují nastavené podmínky pro využití v této práci.

#### 5.1.1 Data z biomonitoring zátěže přírody

Data pokrývají území ve velmi řídkém, ale pravidelném rozmístění. Jedná se o 14 bodů. Znázornění jejich polohy je na Obr. 6. Analýza vzorků byla ve výzkumu zpracována následujícím způsobem. Na každém místě bylo celkově odebráno 9 litrů vzorku mechu. Vzorky byly následně sušeny na vzduchu, poté namlety v mlýnku s titanovým rotorem a přesáty na sítech, s velikostí ok 0,2 mm. Dále byly vzorky rozloženy pomocí mikrovlnné trouby pracující se zvýšeným tlakem (MARS 5) nejprve v kyselině dusičné ( $\text{HNO}_3$ ) a následně pak v peroxidu vodíku ( $\text{H}_2\text{O}_2$ ). Analyzování vzorků probíhalo na přístroji ICP-MS (PE, Elan 6000). Ve vzorcích mechu bylo zjištěno těchto 35 prvků: Ag, Al, As, Ba, Be, Bi, Cd, Ce, Co, Cr, Cs, Cu, Fe, Ga, In, La, Li, Mn, Mo, Ni, Pb, Pr, Rb, S, Sb, Sc, Se, Sr, Th, Tl, U, V, Y a Zn. Pro analýzy v diplomové práci byly vybrány ukazatele, které překračovaly normy na území Příbramska nebo ukazovaly abnormality. Jedná se o: Ag, Bi, Cd, Cr, Cu, Fe, In, Mo, Pb, Sb, Sn a Zn. Naměřená množství jednotlivých prvků byla uvedena v jednotkách  $\mu\text{g/g}$  (ppm). Získaná hodnota sice neodpovídá přesné koncentraci prvku v atmosféře, ale jednotlivé hodnoty jsou porovnatelné. Jedná se o tzv. proxy data. V místech, kde je vyšší koncentrace prvku v mechu, je vysoce pravděpodobné, že je i vyšší koncentrace v atmosféře a naopak.

Data byla ručně přepsána ze závěrečné zprávy do tabulky v programu Microsoft Excel 2010. Tato tabulka byla uložena jako textový dokument s nastavením, jak se mají zobrazovat stupně, minuty, vteřiny a čím oddělit jednotlivé buňky. Souřadnice byly

ve formátu WGS. Proto bylo nutné převést je do S - JTSK. Předpřipravený textový soubor byl otevřen v programu Qgis Desktop 2.12. Bylo nastaveno zobrazení stupňů, minut, vteřin a způsob oddělení jednotlivých buněk. Tímto krokem vznikla vrstva mechy.shp. S touto vrstvou se dále pracovalo v programu ArcGIS Desktop, kde byla provedena interpolace jednotlivých prvků do plochy. Závěrečná zpráva z výzkumu obsahovala mapy s interpolací jednotlivých prvků. Ty ale nebyly k dispozici v elektronické podobě. Ve zprávě nebyla ale uvedena zvolená metoda interpolace. Proto bylo nutné celý proces projít znovu a dostat se k interpolaci. Interpolace byla prováděna metodou IDW, ale její nastavení okolních bodů a síly probíhalo na základě vizuálního srovnání s původními mapami. Takto vzniklo 12 vrstev, které byly pojmenovány jako Mechy\_název interpolovaného prvku (např. Mechy\_Pb) s velikostí buňky 50 m. Následně byl využit nástroj Extract Values to Point, který přiřadí hodnotu vrstvy v daném bodě k jiné vrstvě. Tím byla přidána hodnota jednotlivých rastrů k bodům medicínálních dat a k vrstvě škol.



**Obr. 6** Prostorové znázornění koncentrace  $SO_2$  a  $Fe$  v oblasti Píbramska  
(Zdroj dat: ČHMÚ, VÚKOZ, ArcČR 500)

### 5.1.2 Data od ČHMÚ

Data od ČHMÚ jsou volně k dispozici na jejich internetových stránkách. V diplomové práci byly využity údaje z měřících stanic a údaje o znečištění vnějšího ovzduší, které byly již zpracovány ČHMÚ. Data z měřících stanic byla nejprve centralizována do tabulky v programu Microsoft Excel 2010. Data byla zpracována z 23 stanic, poloha jednotlivých stanic je zaznamenána na Obr. 6 . Bohužel jednotlivé stanice neměří vždy všechny parametry, proto byly vybrány ty parametry, které měří nejvíce stanic. Jedná se o oxid siřičitý ( $\text{SO}_2$ ), prašný aerosol (SPM), oxidy dusíku ( $\text{NO}_x$ ), mangan (Mn), nikl (Ni), olovo (Pb), arsen (As) a částice velikostní frakce pod  $10 \mu\text{m}$  ( $\text{PM}_{10}$ ). Koncentrace  $\text{SO}_2$  na jednotlivých stanicích je zaznamenána na Obr. 6 . Ze získaných dat za období 2000 až 2005 byl vypočítán aritmetický průměr pro každý ukazatel. U informací o jednotlivých stanicích byla zaznamenána i poloha. Pomocí této informace byla data zobrazena v programu ArcGIS Desktop za použití funkce Display XY Data. Zde bylo nastaveno pole X a pole Y, ze kterého má vzniknout nová vrstva. Byl zvolen Souřadnicový systém jednotné trigonometrické sítě katastrální (S-JTSK).

Dále bylo nutné jednotlivá data interpolovat do plochy. K tomu byla využita extenze Geostatistical Analyst a funkce Geostatistical wizard. Daná data byla interpolována s použitím metody IDW, kdy bylo využito minimálně 5 bodů a maximálně 10 bodů z okolí. Získané vrstvy byly uloženy jako CHMU\_název interpolovaného prvku (např. CHMU\_Fe) s velikostí buňky 50 m. Následně byl využit nástroj Extract Values to Point. Tímto byla přidána hodnota jednotlivých rastrů k bodům medicínálních dat.

Data o znečištění vnějšího ovzduší za pětileté období jsou už zpracována do vrstvy .shp, která pokrývá celé zájmové území v podobě čtverců s průměrnou naměřenou hodnotou, a proto nebylo nutné ji dále upravovat. Tato vrstva byla zpracována z plošných map za jednotlivé roky. Mapy za jednotlivé roky jsou vypracovány na základě gaussovského disperzního modelu SYMOS 97. Více informací o zpracování daných dat je k dispozici na stránkách ČHMÚ. Vrstva využitá v diplomové práci je znázorněna na Obr. 7 . Data byla připojena nástrojem Join na základě polohy k medicínálním datům a k vrstvě škol.

Jelikož zpracovaná data od ČHMÚ byla až za období 2007 až 2011, byla v práci využita i data z měřících stanic za období 2000 až 2005 přesto, že některé atributy jsou shodné. Pro zpracování dat z měřících stanic byla použita metoda IDW, která není přímo určena pro tento typ dat, jež nejsou v pravidelné síti a mají méně záznamů než by bylo vhodné. Tato metoda byla zvolena proto, že při zpracování dat z biomonitoringu mechů dávala nejpřijatelnější výsledky při porovnání s mapami v závěrečné zprávě. Zároveň se v předchozí práci Grundlocha (2009) ukázala jako přínosná. U dat z měřících stanic by díky zvolené metodě

interpolace bylo nutné výsledky hodnotit důkladněji a více se zaměřit na jejich interpretaci. Stejně tak u dat o znečištění vnějšího ovzduší, která jsou již zpracována ČHMÚ, je také nutné věnovat pozornost různosti období při hodnocení výsledků.

### 5.1.3 Data z analýzy pitné vody

Data od Státního zdravotního ústavu byla získána v excelové tabulce. Záznamy jsou za období 2004 až 2005 a obsahují informace z 10 vodovodů (zobrazeno v Tab. 1). Ke každému vodovodu byl poskytnut záznam o 9 měřených atributech (arsen, bromičnany, kadmium, měď, olovo, rtuť, selen, železo a polutanty celkem). Jednotlivé vodovody byly vybrány podle počtu vyšetřených dětí v dané obci. Sledované atributy byly vyčleněny na základě možného ovlivnění lidského organismu. Tuto tabulku bylo nutné transformovat do vhodného formátu. Bylo nezbytné zohlednit některé vodovody obsluhující více obcí (např. Vodovod Příbram - svazek obcí). Následně byla tabulka napojena k vrstvě obce.shp z ArcČR 500 na základě Názvu. Získaná vrstva je znázorněna na Obr. 7 . Dále pomocí nástroje Join byla data o měřených atributech připojena k vrstvě medicínálních dat a k vrstvě škol.

Kód	Název vodovodu	Počet zásobovaných obyvatel
1	vodovod Příbram-město	35500
2	vodovod Příbram-Svazek obcí	8120
3	vodovod Dobříš-Stará Huť	8000
4	vodovod Sedlčany	7163
5	vodovod Březnice	3355
6	vodovod Rožmitál pod Třemšínem	3335
7	vodovod Jince-ÚV	1800
8	vodovody VHS (příbramská skupina)	1410
9	vodovod Petrovice	740
10	vodovod Obecnice-Panské lesy	810

**Tab. 1. Přehled vodovodů použitých v práci**

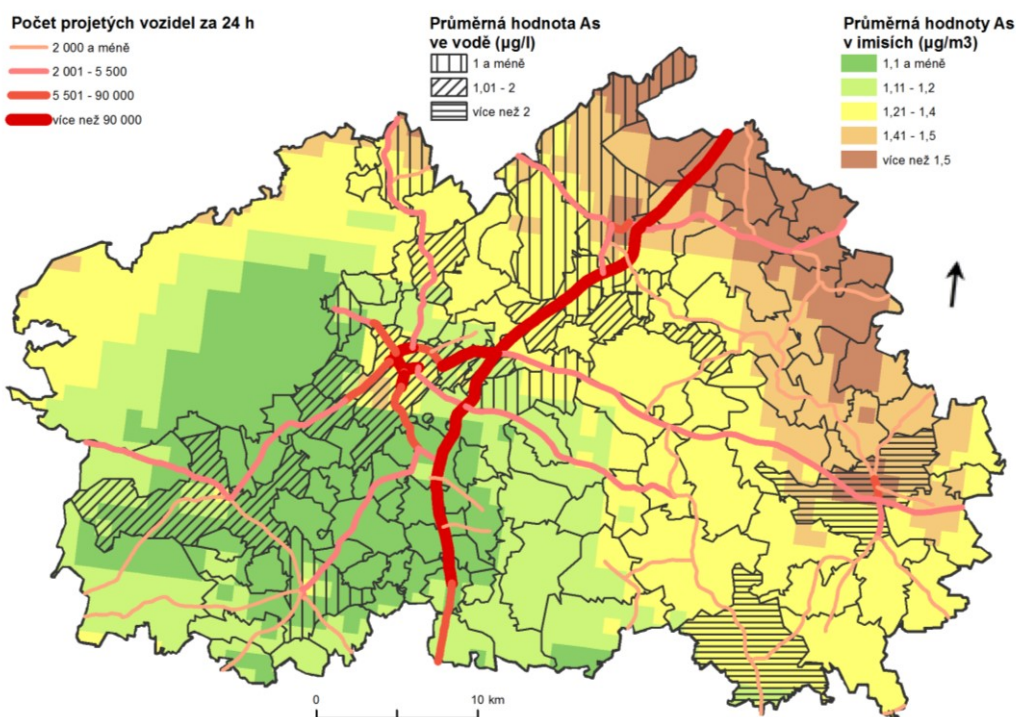
(Zdroj dat: SZÚ)

### 5.1.4 Intenzita dopravy

Data využitá pro vytvoření této mapové vrstvy byla získána od Ředitelství silnic a dálnic. Jednalo se o data z průzkumu v roce 2000 v okrese Příbram. Získaná data obsahovala velké množství informací o projetých vozidlech. Pro účely diplomové práce byl důležitý atribut číslo silnice, číslo úseku a součet všech motorových vozidel a přívěsů. Tato data bylo nutno navázat na vrstvu silnic, která byla použita z ArcČR 500. Připojení muselo proběhnout pro každý úsek zvlášť, jelikož délka jednotlivých měřených úseků neodpovídala délce úseku silnic v ArcČR 500. Zároveň se také lišilo označení těchto úseků. Vrstva silnic musela být za pomoci editoru

rozsekána na menší úseky, které odpovídaly úsekům definovaných Ředitelstvím silnic a dálnic. Následně byla k této vrstvě přidána informace o počtu všech projetých motorových vozidel a přívěsů za jeden den. Poté proběhla kontrola s online mapou na stránkách Ředitelství silnic a dálnic tak, aby nedošlo k chybě při určování úseků. Výsledná vrstva je znázorněna na Obr. 7

Takto vzniklá vrstva byla připojena k vrstvě medicínálních dat a k vrstvě škol pomocí funkce Join. Tímto byla získána vzdálenost bodu od silnic a zároveň počet projetých vozidel. Do vrstvy medicínálních dat byl poté přidán nový sloupec. Do tohoto sloupce byl vypočítán dopravní koeficient, který představuje počet automobilů projetých nejbližším úsekem vydělený vzdáleností od tohoto úseku. Toto vážení vzdáleností vzniklo proto, že osoba, žijící ve větší vzdálenosti od velmi frekventované silnice, může být pod stejným vlivem jako osoba, žijící v blízkosti méně frekventované silnice. Tento postup byl aplikován i na vrstvu škol.



**Obr. 7 Mapa vybraných environmentálních dat v okrese Příbram**  
(Zdroj dat: SZÚ, ŘSD, ČHMÚ, ArcČR 500)

## 5.2 Geologická data

Data byla dodána ve formátu .shp a obsahovala vrstvy zlomy\_50, predkvarter, kvarter, radiometrie a linie\_50. Veškerá příprava daných vrstev probíhala v programu ArcGIS Desktop.

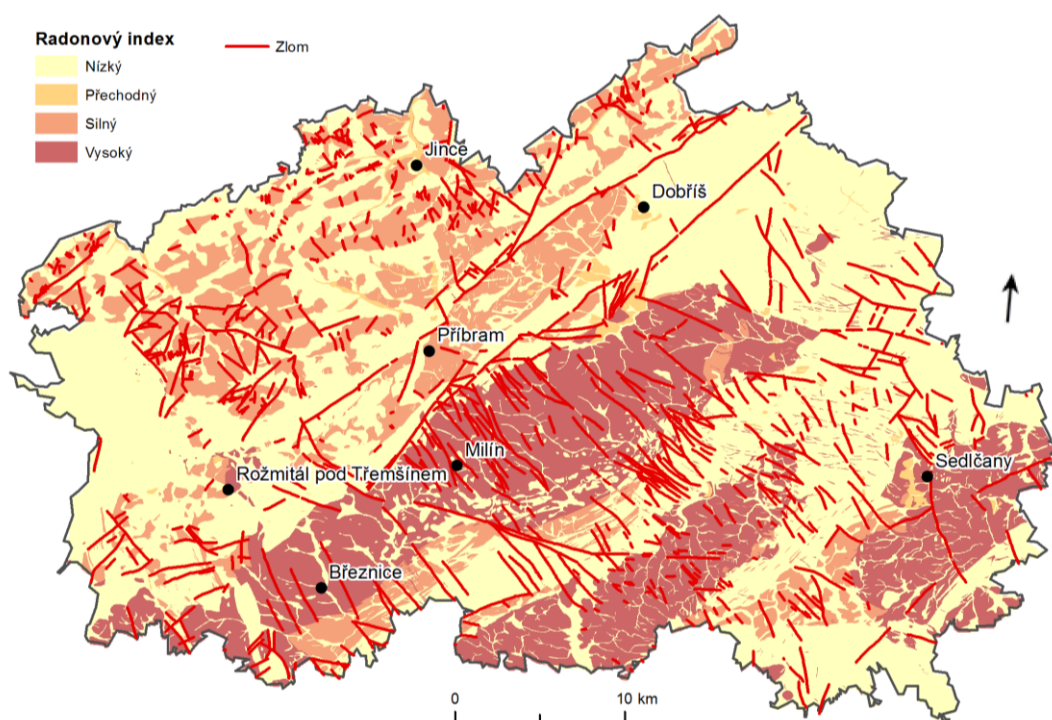
Vrstva linie\_50 obsahovala znázornění hranic jednotlivých typů hornin a hranic mapových listů, tedy data, která nejsou potřebná pro účel práce, a proto nebyla použita.

Vrstva zlomy\_50 znázorňuje zjištěné zlomy na území okresu Příbram. Tato vrstva byla připojena k vrstvě medicinálních dat a k vrstvě škol pomocí funkce Join. Tímto byla získána vzdálenost bodu od zlomu. Tato vzdálenost byla dále dělena do devíti kategorií pomocí přidání nového sloupce a funkce Field calculator. Jednalo se o kategorie do 10 m, 25 m, 50 m, 100 m, 200 m, 300 m, 500 m, 1 km a kategorie a více. K rozdělení do kategorií lze také využít základní funkce z Analysis Tools.

Vrstvy predkvarter a kvarter společně pokrývají celé území okresu Příbram. Obsahují velké množství informací, jako je typ horniny, název, zrnitost a hlavně radonový index jednotlivých hornin. Pomocí funkce Join byla připojena tabulka na základě polohy k medicinálním datům a k vrstvě škol. Pro přehlednost byla v nově vzniklé vrstvě ponechána pouze informace o radonovém indexu k příslušnému medicinálnímu bodu.

Poslední vrstvou zde byla radiometrie. Vrstva nese informaci o celkovém dávkovém příkonu gama záření hornin. Tato informace byla opět funkcí Join přiřazena k medicinálním datům na základě polohy a také k vrstvě škol.

Ze získaných geologických dat byla k medicinálním datům a k vrstvě škol přiřazena informace o vzdálenosti od zlomu. Tato informace byla i kategorizována. Dále byla připojena informace o radonovém indexu jednotlivých hornin a celkový dávkový příkon gama záření hornin. Poloha zlomů a intenzita radonového indexu je znázorněna na Obr. 8



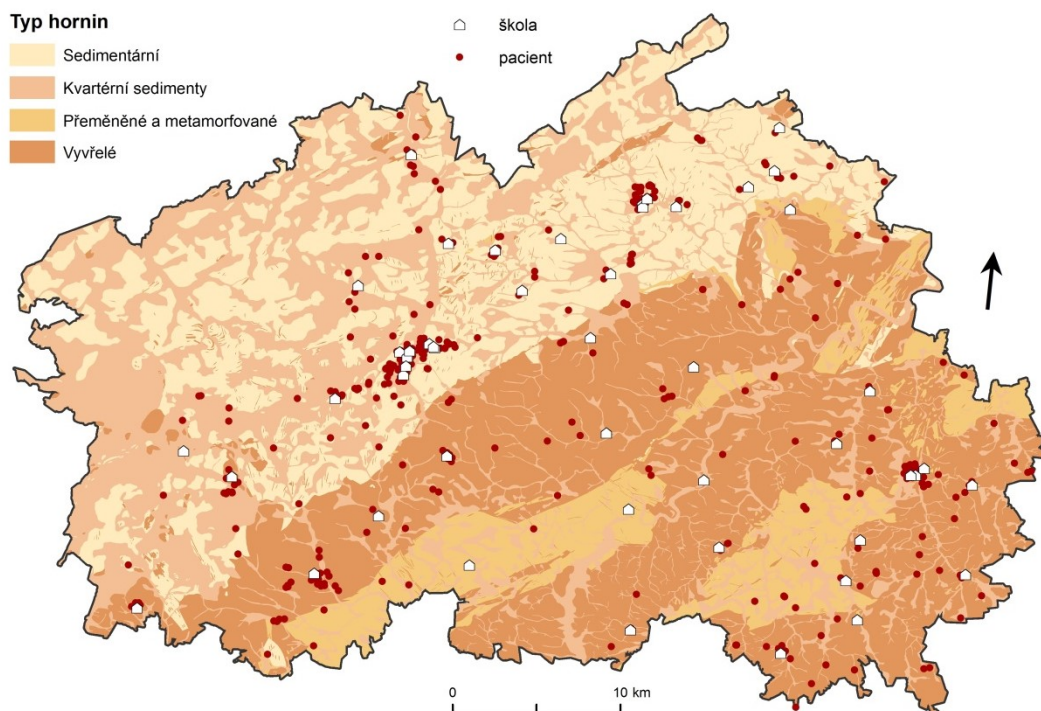
**Obr. 8** Mapa radonového indexu a zlomů v okrese Příbram  
(Zdroj dat: ČGS, ArcČR 500)

### 5.3 Medicinální data

Základní soubor dat byl získán z Pražského Endokrinologického ústavu a obsahoval 1090 záznamů. Z tohoto souboru byli vybráni pacienti ve věku 6 - 17 let (celkem 565 záznamů). Získaný soubor neobsahoval přesné adresy pacientů. Proto bylo nutné je dopsat z karet pacientů na základě jejich identifikačního čísla. K jednotlivým adresám bylo nezbytné získat S- JTSk souřadnice. Tento krok byl proveden online pomocí Registru územní identifikace adres a nemovitostí (RÚIAN). Registr umožňuje určit souřadnice na základě zadání adres. Souřadnice byly zaznamenány do nově vzniklých sloupců ke každému pacientovi. U 8 pacientů byla dodána adresa, kterou již nyní nelze nalézt nebo která byla v kartách chybně zapsána. Dále byly odstraněny duplicity na základě identifikačního čísla. Z karet pacientů byl také opsán údaj o délce pobytu pacientů v místě bydliště. Pokud byl pobyt kratší než 5 let, byli pacienti vyřazeni. Po této úpravě zůstalo v souboru 490 pacientů.

Původně bylo v souboru 392 parametrů. Protože některé z nich nebyly potřebné pro účely diplomové práce, byly vymazány. Některé z nich musely být dále dekódovány (jeden parametr v sobě nesl více informací). K atributům (TSH, Volgut, FT3, FT4, Jodur, TPO, TGL, CRP, LDL, Glukoza) následně byly vypočítány normy hladin hormonů. Seznam konečných parametrů s hladinami norem a charakteristikou kategorií je v Příloze 2.

Po těchto úpravách byl soubor nahrán do programu ArcGIS Desktop. Pro zobrazení přesného umístění pacientů byla použita funkce Display XY Data. Zde bylo nastaveno pole X a pole Y, ze kterých vznikla nová vrstva .shp a byl určen Souřadnicový systém jednotné trigonometrické sítě katastrální (S - JTSK). Znárodnění těchto dat je na Obr. 9 . K takto vzniklé vrstvě byla postupně přiřazována další data způsobem, jaký byl popsán v předchozích částech kapitoly.



**Obr. 9** Znáznornění vyšetřených pacientů, škol a typu hornin v okrese Příbram  
(Zdroj dat: Endokrinologický ústav v Praze, lags.cz, ČGS)

### 5.3.1 Školy

K údajům o pacientech byla přiřazena informace o poloze jejich školy. Nejdříve byl vytvořen seznam základních škol v okrese Příbram. Do práce byly zahrnuty pouze základní školy i přesto, že věkové rozmezí pacientů je 6 až 17 let. Je to z toho důvodu, že základní školy navštěvovali pacienti po dobu 9 let a nebere se v úvahu možnost přestupu na jinou školu. Informace byly získány od odboru školství, kultury a sportu v Příbrami a z webových stránek lags.cz. Na základě adres byly z RÚIAN získány S - JTSK souřadnice a zapsány do nových sloupců v Microsoft Excel. Pomocí těchto souřadnic byly jednotlivé školy znázorněny v prostoru v programu ArcGIS Desktop pomocí funkce Display XY Data (zobrazeno na Obr. 9 ). K takto vytvořené vrstvě byla funkcí Join na základě polohy přiřazena další data. Tento postup je popsán v předchozích částech této kapitoly.

V okrese Příbram neexistují spádová území pro jednotlivé školy, ale rodiče si v současné době mohou vybrat, do které školy bude dítě chodit. Musí pouze splňovat podmínku, že bude docházet v území obce s rozšířenou působností (ORP) svého bydliště. Při zpracování dat se předpokládalo, že dítě navštěvuje školu, která je nejbližší jeho adrese. Z tohoto důvodu byla vytvořená vrstva škol společně s informacemi o geologických a environmentálních aspektech, připojena nástrojem Join na základě vzdálenosti k pacientovi. U pacientů tedy byly informace



geologické a environmentální za polohu bydliště a polohu školy, kterou pacient navštěvuje. Následně byly přidány v atributové tabulce vrstvy nového sloupce, kde byly vypočítány aritmetické průměry geologických a environmentálních dat za polohu školy a polohu bydliště pacienta. Název nového sloupce se řídil klíčem pr\_název sloupce pouze za polohu adresy (např. pr\_PB\_mech).

## 5.4 Normalita dat a mocinná transformace

Výsledkem předchozích částí byl soubor medicínálních, geologických a environmentálních dat za polohu bydliště pacienta a zprůměrované hodnoty polohy bydliště a školy pacienta. Aby na tento soubor mohly být aplikovány metody analýzy prostorových dat, bylo nutné soubor prozkoumat a upravit ho pro vstup do jednotlivých metod.

V přípravě souboru medicínálních dat byla data rozdělena do následujících skupin: identifikační data, binární data, kategoriální nebinární data, ordinální data a metrická data. Metrická data se týkají naměřených zdravotnických hodnot a také naměřených hodnot vnějšího prostředí pacienta. Pro následné analýzy je nutné, aby metrická data splňovala předpoklad normality a konstantního rozptylu (homoscedasticita) tedy aby měla v ideálním případě normální (Gaussovo) rozdělení nebo alespoň rozdělení symetrické.

V programu IBM SPSS Statistics byla otestována normalita těchto dat. Pro testování byl využit Kolmogorov-Smirnovův test a Shapiro-Wilkův test s hladinou významnosti 95 %. Výsledky těchto testů jsou uvedeny v Příloze 3. Nejprve byla určena nulová hypotéza ( $H_0$ ), tedy že data mají normální rozdělení. Alternativní hypotéza ( $H_1$ ) popírá platnost  $H_0$  a tvrdí, že data nemají normální rozdělení. O přijetí nebo zamítnutí jednotlivých hypotéz se rozhodujeme na základě signifikance neboli významnosti testu. V tabulce je vidět, že signifikance u naprosté většiny dat nepřekračuje hodnotu 0.05. V tom případě zamítáme  $H_0$  a přijímáme  $H_1$  tedy, že data nemají normální rozdělení. U Kolmogorov - Smirnova testu několikrát vyjde signifikance  $p > 0.05$ , ale u Shapiro - Wilkova testu vyšla signifikance  $p < 0.05$ , a proto je potvrzena  $H_1$ . Je zřejmé, že ani tato data nemají normální rozdělení.

Jelikož data nespĺňují předpoklad normálního rozdělení, je nutno užít k jejich úpravě mocinné transformace, kterou bude primárně dosaženo hodnot šikmosti blízké nule a sekundárně i hodnoty špičatosti blízké 3 (méně důležité).

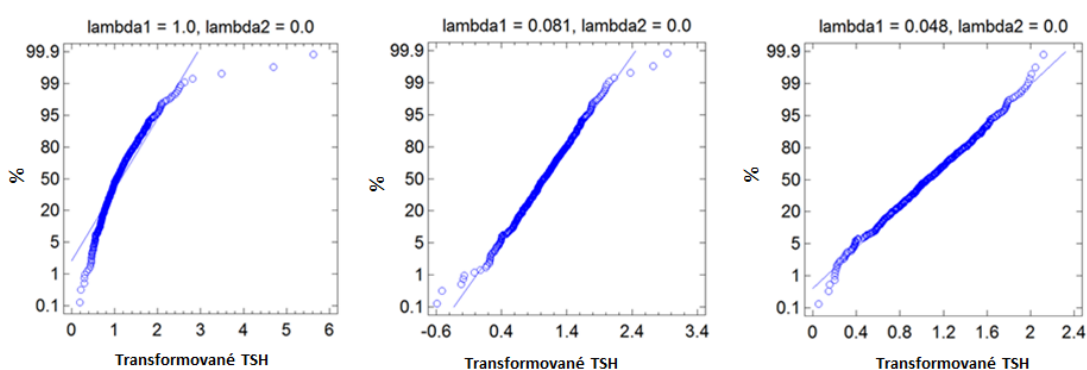
Nejprve byla data posunuta přičtením konstanty ke kladným hodnotám blízkým nule. Poté byl výsledek podělen mediánem (díky tomuto pak transformace úspěšně funguje i při vysokých mocninách). Úpravy byly prováděny v Microsoft Excel 2010. Původní soubor dat byl uložen do záložky pod názvem data, dále byla vytvořena nová záložka, která pomocí vzorce

KDYŽ upravovala vybraná data (jednalo se o data metrická) a tato data byla ukládána do záložky datac. Vzorec KDYŽ byl nadefinován takto:

$$\text{KDYŽ}(\text{data!BI2}="";"";\text{KDYŽ}(\text{datap!BI\$7}=1;(\text{data!BI2}+1.1*\text{datap!BI\$6}*(\text{ABS}(\text{datap!BI\$11})+0.05*\text{ABS}(\text{datap!BI\$13}))))/(\text{datap!BI\$13}+1.1*\text{datap!BI\$6}*\text{ABS}(\text{datap!BI\$11}));\text{data!BI2}+1.1*\text{datap!BI\$6}*(\text{ABS}(\text{datap!BI\$11})+0.05*\text{ABS}(\text{datap!BI\$13}))))).$$

Vzorec čerpá hodnoty nejprve z původního souboru dat a poté ze záložky s názvem datap. Tato záložka se všemi hodnotami je přiložena v Příloze 4. Do této záložky se průběžně počítalo minimum, maximum, medián, šikmost, špičatost a byl zaznamenán i počet hodnot. Většina hodnot byla vypočítána bez problémů. U hodnoty SONO se výpočet podařil až po záměně mediánu (jeho hodnota byla 0) ve vzorci za aritmetický průměr, jehož hodnota byla nenulová.

Formulář datac byl nahrán do programu Statgraphics. Zde byla pomocí funkce Describe - Numeric data - Power transformation získána hodnota lambda  $\lambda$  neboli mocnina. Aby bylo možné hodnotu lambda použít, bylo nutné zkontrolovat, zda jsou koeficient šikmosti a z - score pro šikmost blízké nule. Zároveň byl vizuálně zkontrolován kvantil-kvantilový graf pro normální rozdělení a zjištěné odlehlé hodnoty byly z odhadu pro  $\lambda$  vyjmuty. Postup úpravy grafu je znázorněn na Graf 1. Na prvním grafu zleva je patrné, jak vypadají data bez transformace. Na druhém grafu zleva jsou zobrazena data po automatické transformaci. Zde je nutné odstranit odlehlé hodnoty. Na posledním grafu zleva jsou znázorněny hodnoty po automatické transformaci, kdy nejsou započítány odlehlé body. Po překontrolování vhodnosti všech parametrů bylo možné vzít hodnotu  $\lambda$  a zkopírovat ji do původního formuláře datap v Microsoft Excel.

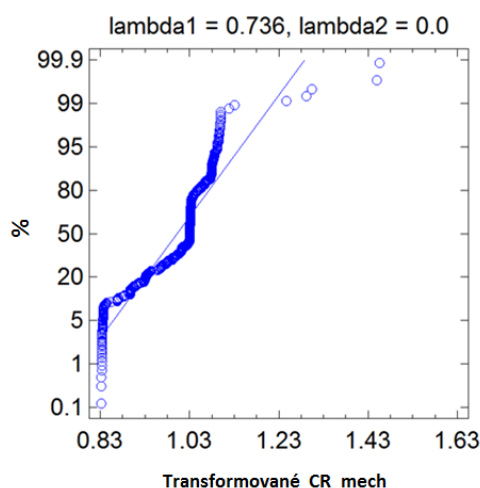


**Graf 1. Postupná úprava hodnot pro parametr TSH**

U několika případů medicínálních dat nastala situace, kdy program nedokázal určit nejhodnější hodnotu  $\lambda$ . Proto byla v těchto případech odhadnuta  $\lambda$  metodou půlení intervalu. Kritériem v těchto případech byla blízkost hodnoty šikmosti a jejího z - score nule.

U parametrů vod bylo nutné rozhodnout, zda je transformace nezbytná, neboť hodnoty se pohybovaly jen v několika málo úrovních (jsou naměřeny dvě nebo tři hodnoty, které se opakují). U parametrů mechu bylo nutno ve většině případů použít transformaci. Na Graf 2 jsou znázorněny hodnoty chromu obsaženého v mechu. Je zde vidět kolísání hodnot a hlavně výrazně odlehle hodnoty shora. Tyto hodnoty byly z odhadu pro  $\lambda$  odstraněny.

Díky interpolaci jsou data rozmístěna pravidelně. Na grafu bylo patrné, že jsou data v území různorodá. Tuto různorodost a výkyvy bylo třeba zachovat. U dat z ČHMÚ byly prováděny stejné kroky jako u mechu. Na grafech dat imisního znečištění je vidět největší soudružnost a nebylo nutné je upravovat, pouze se použila hodnota  $\lambda$  vypočítaná programem. Toto nastalo i u dávkového příkonu Gamma záření hornin. U radonového indexu bylo lepší variantou jej netransformovat, jelikož nabývá malého množství hodnot. Hodnoty znázorňující průměr hodnot naměřených v místě bydliště a u škol nebylo třeba upravovat a transformace proběhly bez problémů.



**Graf 2. Hodnoty Chromu v mechu po transformaci bez úprav**

Získané hodnoty  $\lambda$  byly zadány do sešitu Microsoft Excel pro transformace, záložky datap a pole s názvem síla. V záložce datat pak byly vypočítány transformované hodnoty za pomoci funkce KDYŽ definovanou takto:

$KDYŽ(data!BI2="" ; "" ; KDYŽ(datap!BI\$2=0 ; LOG(datac!BI2) ; SIGN(datap!BI\$2) * (datac!BI2^{datap!BI\$2}))$ .

Získané hodnoty byly retransformovány rekurentním vzorcem pro kontrolu. Transformovaná data pak bylo možno dále použít pro vstup do analýz vyžadujících Gaussovské rozdělení.

## 6 Mapování a popis výskytu zhoršené funkce štítné žlázy

Po přípravě dat, která proběhla v kapitole 5, se přistoupilo k mapování a popisu výskytu zhoršené funkce štítné žlázy. V první části této kapitoly byl analyzován základní soubor medicínálních dat. Bylo zde zaznamenáno, kolik je celkově k dispozici záznamů, kolik se nachází pacientů v jednotlivých věkových kategoriích a jakého jsou pacienti pohlaví. Soubor byl důkladně analyzován z pohledu počtu překročených norem u jednotlivých ukazatelů funkce štítné žlázy. V závěru byli pacienti přiřazeni k jednotlivým školám na základě polohy. Tyto školy pak mezi sebou byly porovnány z pohledu počtu pacientů s potenciálně zhoršenou funkcí štítné žlázy.

V druhé části této kapitoly byli analyzováni pacienti s překročenou normou parametrů štítné žlázy na základě jejich polohy. Byla zde porovnávána poloha pacientů s překročenými normami jednotlivých ukazatelů funkce štítné žlázy s daty o přírodním prostředí (charakteristiky vod, znečištění imisí, naměřená data ze stanic ČHMÚ, intenzita dopravy, geologické podloží).

### 6.1 Statistický popis zhoršené funkce štítné žlázy

Po úpravě databáze pacientů zůstaly záznamy u 490 z původních 565 pacientů, z čehož 234 pacientů jsou ženy a 241 pacientů muži. Z roku 2000 je k dispozici 249 záznamů a z roku 2005 je 241 záznamů. Ve věku do 6 let je v databázi 88 pacientů, od 7 do 10 let je 126 pacientů, od 11 do 13 let je 106 pacientů a od 14 do 17 let je 170 pacientů. U 240 pacientů se buď někdo z příbuzných léčí nebo se v minulosti léčil se štítnou žlázou.

V Tab. 2 je zaznamenáno 8 ukazatelů s uvedením, u kolika pacientů byly naměřené hodnoty snižené, zvýšené nebo v normě. Tabulka znázorňuje u každého parametru počet pacientů, kteří mají hodnotu mimo normu, a to buď nižší anebo vyšší. U parametrů TPO a TGL se sleduje překročení spodní hranice normy, zvýšená hodnota u těchto parametrů není na škodu. Naopak je tomu u parametru Volgut. Hodnoty jednotlivých hladin jsou zaznamenány v Příloze 2. Zde jsou také vysvětleny zkratky parametrů. Parametry, u kterých je 10 %, značí rozšířené hladiny norem o 10%.

Název	Hladina			Chybějící hodnoty
	Snížená	Norma	Zvýšená	
TPO	11	479		0
TPO 10 %	13	477		0
TGL	11	479		0
TGL 10%	11	479		0
Volgut		383	103	4
Volgut 10 %		346	140	4
CRP	13	223		254
CRP 10 %	14	222		254
LDL	25	211		254
LDL 10 %	46	190		254
Glukoza	4	229	3	254
Glukoza 10 %	30	168	38	254
TSH	2	411	59	19
TSH 10 %	2	409	61	18
fT3	5	382	83	20
fT3 10 %	6	347	117	20
fT4	10	460	2	18
fT4 10 %	36	433	3	18
Jodur	69	288	121	12
Jodur 10 %	117	236	125	12
Chol	79	149	8	254
Chol 10 %	136	65	35	254
UREA	3	228	5	254
UREA 10 %	8	215	13	254
Cpepti	36	180	21	253
Cpeptid 10 %	53	156	28	253
Inzulin	25	209	5	251
Inzulin 10 %	32	198	9	251
Fluor	13	203	14	260
Fluor 10 %	19	192	19	260

**Tab. 2. Překročení hladin norem ukazatelů**

Pacienti na Příbramsku mají problém s objemem štítné žlázy (Volgut), kdy převažuje její zvětšení oproti normě. Objem štítné žlázy se zvětšuje při nedostatku jódu. To je názorné v Tab. 2, 69 pacientů má nízkou hladinu jódu v moči (Jodur). Naopak 121 pacientů má vysokou hladinu jódu v moči (Jodur). To má za následek například růst strumy. Problematický se u pacientů jeví i cholesterol. Hodnoty nad normu a pod ni přispívají k vyjádření nadfunkce a podfunkce štítné žlázy. K určení onemocnění štítné žlázy se také využívají hladiny hormonů TSH, FT4 a FT3. Hladiny TSH a FT3 mají pacienti v 59 a 83 případech vysoké. Tyto ukazatele se sledují převážně proto, že jsou pro určení poruch štítné žlázy důležité. Z tohoto důvodu je k dispozici nejvíce dat na rozdíl od jiných ukazatelů, kde chybí až přes 250 záznamů.

Na Příbramsku má celkově 242 pacientů překročenou alespoň jednu z norem. Pokud by byly dostupné všechny záznamy, lze předpokládat, že by tento počet vzrostl. Překročenou normu 2 a vícekrát má 63 pacientů. Četnosti překročení jsou zaznamenány v Tab. 3. Při upravených hladinách norem o 10 % došlo k překročení u 408 pacientů. V porovnání s Tab. 4 je patrné, že při dostupnosti kompletních dat by byla naměřena překročení ještě

četnější. U Volgut, TSH, FT3, FT4 a Jodur byla překročena norma celkem 216krát a po úpravě normy až 376krát. U 44 pacientů byla překročena norma 2 a vícekrát (Tab. 4). Z toho lze vyvodit, že u pacientů s několikanásobným překročením normy je vysoká pravděpodobnost výskytu poruchy štítné žlázy.

Hladiny	Překročení normy									Součet
	1	2	3	4	5	6	7	8	9	
Normální	179	48	14	1						242
Upravené o 10%	127	94	65	54	37	23	3	4	1	408

**Tab. 3. Počet překročení norem celkem**

Hladiny	Překročení normy				Součet
	1	2	3	4	
Normální	172	39	5	0	216
Upravené o 10%	202	122	47	5	376

**Tab. 4. Počet překročení norem u Volgut, TSH, FT3, FT4 a Jodur**

V 0 je zaznamenán počet žáků, u kterých došlo k překročení normy naměřených hodnot ukazatelů na základních školách v okrese Příbram. U třech škol (Hluboš, Tochovice, Solenice) byl počet překročených norem nižší než 50 % vzhledem k počtu pacientů. U 15 škol došlo k překročení normy u všech pacientů, kteří školu navštěvují, a to i vícekrát. Jedná se o ZŠ Nechvalice, ZŠ Petrovice u Sedlčan, ZŠ Nová Ves pod Pleší, ZŠ Rosovice, ZŠ Sedlčany, ZŠ Krásná Hora nad Vltavou, ZŠ Dublovice, ZŠ Obecnice, ZŠ Bohutín, ZŠ Malá Hraštice, ZŠ Nový Knín, ZŠ Příbram - Březové Hory, 2. ZŠ Sedlčany, ZŠ Něčín a ZŠ Jiráskovy sady. Tyto školy jsou v severní a severovýchodní části okresu Příbram. Tato informace ukázala první zaměření na konkrétní území. Při upravené normě o 10 % se ukazuje, že počet škol, kde mají pacienti překročenou normu alespoň jednou, je ještě více.

ID školy						Celkem						celkem 10%	Počet pacientů
	Volgut	TSH	FT3	FT4	Jodur		Volgut 10%	TSH 10%	FT3 10%	FT4 10%	Jodur 10%		
1. ZŠ Dobříš	3	2	2	0	5	12	3	2	3	2	6	16	17
2. ZŠ Dobříš	0	0	0	0	1	1	0	0	0	0	1	1	1
ZŠ Něčín	2	0	4	1	8	15	3	0	4	1	8	16	14
ZŠ Malá Hraštice	2	1	1	0	4	8	2	1	2	0	4	9	7
ZŠ Mokrovraty	2	0	2	0	1	5	2	0	2	1	3	8	5
ZŠ Nová Ves pod Pleší	2	0	0	0	1	3	2	0	0	0	1	3	2
ZŠ Obořiště	4	0	2	0	3	9	5	0	2	0	4	11	9
ZŠ Rosovice	0	0	2	0	1	3	0	0	2	1	1	4	2
ZŠ Stará Huť	1	0	0	0	1	2	2	0	0	0	1	3	3
ZŠ Nový Knín	2	2	2	0	2	8	3	2	3	0	2	10	7
ZŠ Dobříš, Lidická	3	3	3	0	9	18	4	3	5	1	9	22	21
ZŠ Sedlčany	1	0	3	0	4	8	1	0	3	0	5	9	6
ZŠ Chlum	1	0	0	0	3	4	1	0	0	0	3	4	6
ZŠ Jesenice u Sedlčan	2	1	2	0	2	7	2	1	2	1	3	9	9
ZŠ Krásná Hora nad Vltavou	0	2	1	0	1	4	0	2	1	0	1	4	3
ZŠ Počepice	0	1	1	0	2	4	1	1	2	0	3	7	6
1. ZŠ Sedlčany	2	3	3	0	1	9	3	3	5	0	2	13	10
2. ZŠ Sedlčany	7	7	4	1	7	26	8	7	6	2	11	34	24
ZŠ Dublovice	2	2	0	1	3	8	2	2	0	1	4	9	6
ZŠ Nechvalice	1	3	2	0	2	8	2	3	2	0	3	10	5
ZŠ Kosova Hora	2	1	3	0	4	10	3	1	3	1	4	12	10
ZŠ Klučenice	1	1	0	0	2	4	1	1	0	1	2	5	4
ZŠ Petrovice u Sedlčan	5	4	6	0	10	25	6	4	6	0	13	29	16
ZŠ Vysoký Chlumeč	2	0	0	0	1	3	2	0	1	0	1	4	6
ZŠ Bohutín	4	1	6	0	8	19	5	1	7	0	9	22	16
ZŠ Březnice	11	7	5	0	12	35	14	7	9	6	14	50	39
ZŠ Dolní Hbity	0	0	0	0	2	2	0	0	0	0	3	3	3
ZŠ Hvozdčany	1	0	2	0	2	5	2	0	2	0	2	6	7
ZŠ Chraštice	0	1	0	0	1	2	0	1	0	0	2	3	2
ZŠ Kamýk nad Vltavou	0	0	1	0	0	1	0	0	1	0	0	1	2
ZŠ Milín	3	1	5	1	7	17	5	2	6	2	7	22	17
ZŠ Obecnice	3	1	3	1	0	8	4	1	3	1	2	11	6
ZŠ pod Svatou Horou	3	1	3	0	7	14	4	1	8	0	9	22	14
ZŠ Příbram - Březové Hory	4	0	2	0	2	8	4	0	2	1	3	10	7
ZŠ Jiráskovy sady	4	5	3	1	7	20	7	5	4	2	9	27	19
ZŠ 28. října	1	0	3	1	3	8	3	0	5	2	4	14	12
ZŠ Příbram 7	7	7	4	2	20	40	13	7	6	7	31	64	47
ZŠ J.J. Ryby v Rožmitále pod Třemšínem	3	2	0	1	8	14	4	2	3	1	11	21	17
ZŠ Pičín	1	0	0	0	3	4	1	0	1	0	4	6	7
ZŠ Solenice	0	0	0	0	0	0	1	0	0	0	2	3	3
ZŠ Tochovice	0	0	0	0	1	1	1	0	1	0	1	3	4
ZŠ Višňová	2	0	0	0	2	4	2	0	1	0	3	6	6
ZŠ Hluboš	0	0	0	0	2	2	0	0	0	0	2	2	5
ZŠ Suchodol	0	0	1	0	2	3	0	0	1	0	2	3	3
ZŠ Věšín	0	0	2	0	2	4	0	0	2	0	2	4	6
ZŠ Jince	2	0	3	0	5	10	3	1	3	2	6	15	11
ZŠ Příbram 4	0	0	0	0	4	4	0	0	1	0	4	5	5
ZŠ Příbram 2	0	1	0	1	3	5	0	1	0	1	3	5	6
ZŠ Bratří Čapků	7	1	2	1	9	20	9	1	3	2	12	27	27

**Tab. 5. Počet překročení norem ve školách**

V tomto souhrnném statistickém popisu nemoci se ukázalo, že děti na Příbramsku mají často překročené normy ukazatelů funkce štítné žlázy. V porovnání s prací Grundlocha (2009) se jedná o podobná čísla. Když se tato práce zaměřila na definování škol, které prokazují vyšší množství pacientů s onemocněním štítné žlázy, bylo identifikováno 15 škol převážně v severní a severozápadní části okresu Příbram.

## 6.2 Výskyt překročení norem na základě polohy

Následně byli v mapě zobrazeni pacienti. Ti byli vždy kategorizováni podle určitého parametru. V mapě byly také znázorněny kategorizované školy podle % počtu pacientů, kteří překračují normu daného parametru. Pod vrstvu škol a pacientů byly podkládány vrstvy mechů, imisí a silniční dopravy, dat z ČHMÚ, vrstva vod a vrstvy týkající se geologie (gamma záření, radonový index a vzdálenost od zlomu).

### Objem štítné žlázy (Volgut)

Při zobrazení pacientů a rozdělení podle parametru normy objemu štítné žlázy (Volgut) bylo rozložení překračujících norem pravidelné v celém území. Při zobrazení škol podle normy Volgut měly tři školy počet pacientů vyšší než 50 %, jedná se o ZŠ Rosovice, ZŠ Nalžovice a ZŠ Nový Knín. Tyto školy leží v severní a severovýchodní části okresu. Proto byly hledány takové vrstvy, které vykazují abnormality ve stejných místech. Zajímavé je, že rizikové školy leží na podloží s nízkým radonovým indexem. Naopak školy, které mají nízké procento překročení norem, leží na podloží s vysokým radonovým indexem. Tyto školy jsou také na území, kde je v imisích větší množství arsenu a oxidu siřičitého. Znázornění imisních hodnot arsenu je na Obr. 7 V severní části okresu je také zvýšený výskyt mědi v mechách.

U vod jsou dostupné informace jen z nejpočetnějších obcí, jak je znázorněno na Obr. 7. Z tohoto důvodu nebylo možné tři nejrizikovější školy vizuálně porovnat s měřením vod. Proto byly k porovnání vybrány rizikové školy s 30% podílem pacientů překračujících normu. Jednalo se o 10 škol. Tyto školy leží na území, které má nejmenší znečištění arsenem, bromičnanem, kadmíem, olovem a selenem. V obcích, kde je znečištění arsenem vyšší jsou naopak školy, které mají velice nízký podíl pacientů překračujících normu Volgut.

### Thyreotropní hormon (TSH)

Pacienti se zvýšenou nebo sníženou hladinou TSH jsou rozmístěni spíše ve větších městech. ZŠ Dublovice, 1. a 2. ZŠ Dobříš mají 50 % pacientů s překročenou hladinou TSH. Jsou to školy v severovýchodní části okresu. Stejně jako u Volgut leží tyto školy na podloží



s nízkým radonovým indexem. V blízkosti škol jsou silnice, kde projede minimálně 5 000 aut denně. V imisích na daném území je větší množství arsenu a oxidu dusičitého.

### **Volný trijodthyronin (fT3)**

Pacienti, trpící zvýšenou nebo sníženou hladinou fT3, jsou nejvíce rozmístěni na spojnici měst Příbram, Dobříš a okolí. Tomu odpovídají i zobrazené školy. Školy s pacienty trpícími zvýšenou nebo sníženou hladinou fT3 o více než 50 % lze rovněž najít v této oblasti. Pacienti překračující normu jsou často ve vzdálenosti do 1 km od zlomu. Rizikové školy leží na horninách s nízkým dávkovým příkonem Gamma záření do 50 nGy/h a v blízkosti silnic s počtem projetých aut minimálně 8 000 denně. Stejně jako u Volgut a TSH i zde je v imisích naměřeno větší množství arsenu, oxidu dusičitého a oxidu siřičitého.

### **Volný Thyroxin (fT4)**

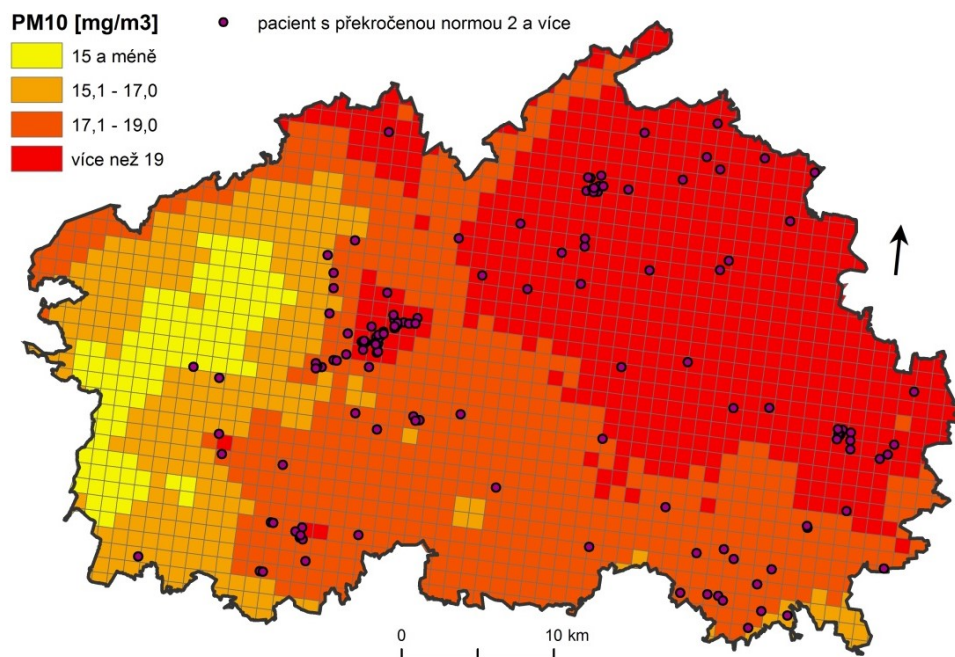
U hladin fT4 žádná ze škol nedosahuje 20% hranice pacientů překračujících normu. Celkem normu překračuje pouze 12 pacientů. Tito pacienti se vyskytují v okolí Příbrami a Sedlčan. Jedná se o území, kde je ve vodě obsažena nízká hodnota olova a mědi. V imisích je zvýšená koncentrace SO<sub>2</sub>, BaP (benzo(a)pyren) a NO<sub>2</sub>. V mechách je vyšší koncentrace Cu a nižší koncentrace Bi.

### **Celkem**

V posledním kroku této části byli v mapě zobrazeni pacienti, kteří mají součet překročených norem 2 a vícekrát (Obr. 10). U těchto pacientů je vysoká pravděpodobnost, že trpí poruchou funkce štítné žlázy. Jelikož nejsou k dispozici závěry z vyšetření pacientů, jsou brány v úvahu právě četnosti překročení norem. Počet těchto pacientů je zaznamenán v 0. Poloha pacientů odpovídá rozmístění nízkých nálezů ve vodě Se, Pb a As, a vyšších nálezů koncentrace Cu v mechách. Také je zde v imisích vyšší koncentrace Pb, As, PM25 (jemné prachové částice s poloměrem 2,5 μm) a PM10 (jemné prachové částice s poloměrem 10 μm). Znázornění pacientů s 2 a vícekrát překročenou normou společně s koncentrací PM10 je na Obr. 10.

Součet překročení norem	Počet pacientů
0	133
1	212
2	107
3	35
4	3

**Tab. 6. Počet překročených norem u pacientů**



**Obr. 10 Vybraní pacienti a koncentrace PM10 v imisích v okrese Příbram**

Z popisu výskytu nemoci a srovnáním s podklady v podobě vrstev mechů, imisí a silniční dopravy, dat z ČHMÚ, vrstev rozborů vod a vrstev týkajících se geologie (dávkový příkon gamma záření, radonový index a vzdálenost od zlomu) vyplynuly možnosti zaměření se na rozbor vod, a to konkrétně na naměřené hodnoty As, Pb a Se. U pacientů překračujících normu často korespondovaly zvýšené hodnoty naměřené v imisích, především obsah SO<sub>2</sub>, NO<sub>2</sub>, As a PM10, a také přítomnost Cu v mechu.

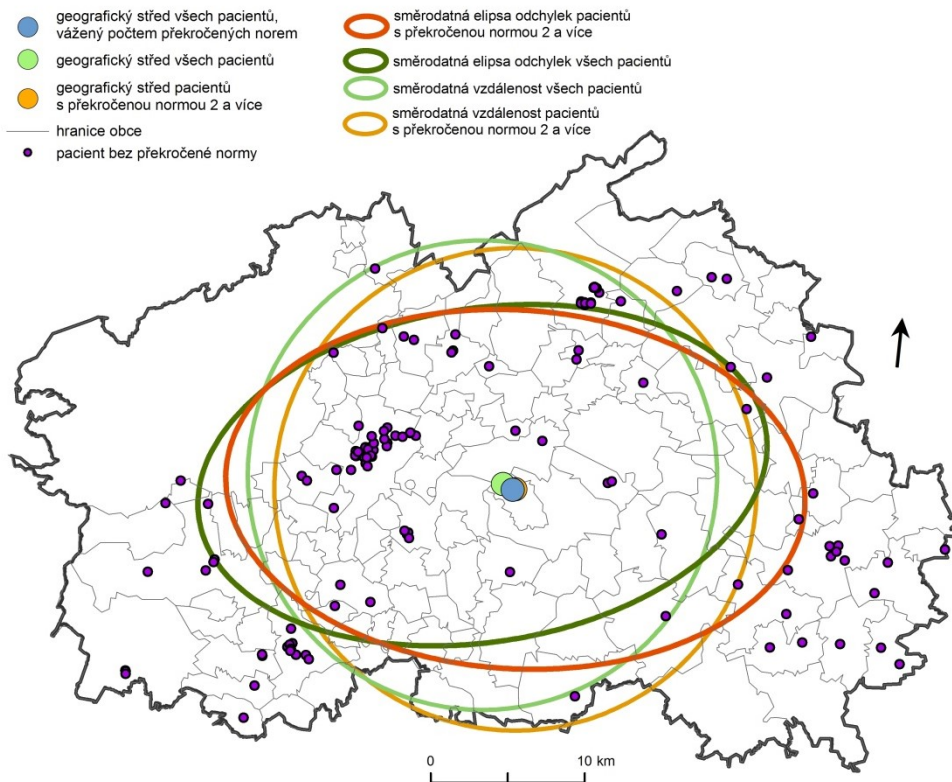
## 7 Analýza prostorových vzorů

V kapitole 6 byla medicínální data popsána na základě základních statistických metod. Následně byl vizuálně porovnán výskyt jednotlivých faktorů zhoršené funkce štítné žlázy v souvislosti s geologickými a environmentálními daty. Tato část práce rozšiřuje informace získané z předchozí části o metody prostorové statistiky. V první podkapitole jsou zachyceny výsledky popisné prostorové statistiky. V druhé podkapitole jsou popsány výsledky konfirmační prostorové statistiky. Je zde využito několik metod k popisu náhodnosti rozmístění a míry shlukování. Tato kapitola byla vypracována primárně v programu ArcGIS Desktop.

### 7.1 Explorační analýza prostorové statistiky

Na Obr. 11 jsou znázorněny výsledky metody popisné prostorové statistiky. Jedná se o popis distribuce prvků pomocí geografického středu, směrodatné vzdálenosti a směrodatné elipsy odchylek.

**Geografický střed** je průměr souřadnic. Může být vážený a nevážený. Geografický střed endokrinologických dat se nachází v severní části obce Jablonná. Jedná se o středovou obec okresu Příbram. Geografický střed vážený počtem překročených norem je posunut mírně jihovýchodně od geografického středu neváženého, ale stále zůstává v obci Jablonná. Velice blízko se nachází geografický střed pouze pro pacienty s překročenou normou 2 a vícekrát. V předchozí části na základě výsledků vyplynulo, že pacienti se zhoršenou funkcí štítné žlázy se nachází spíše v severní oblasti okresu. Na základě geografického středu je ale zřejmé, že pacienti s překročenými normami jsou rozmístěni po celém okresu s výjimkou CHKO Brdy, kde nežijí žádní obyvatelé.



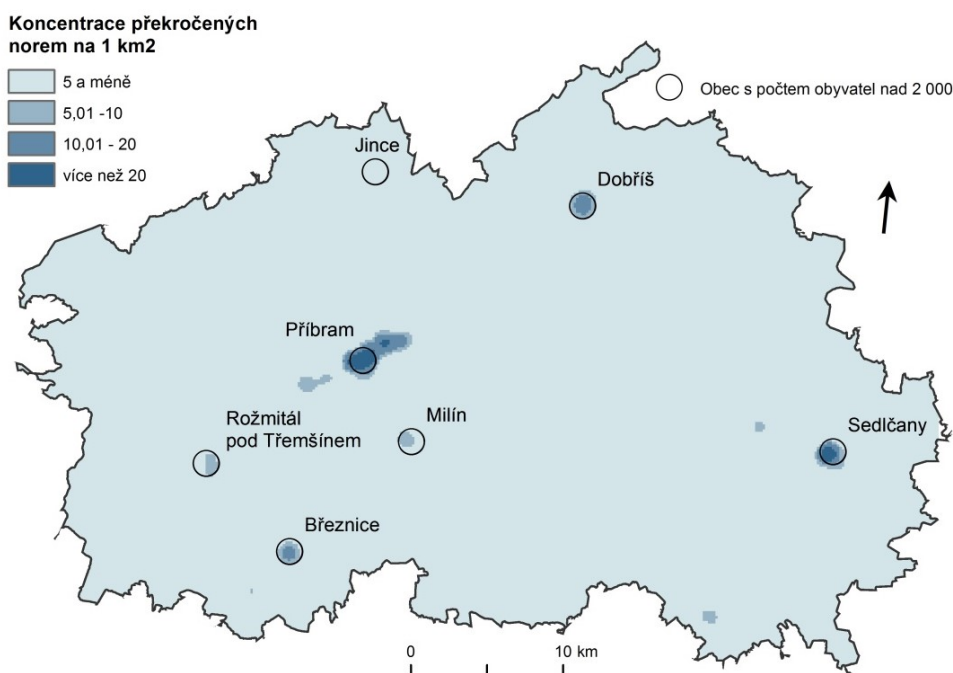
**Obr. 11** *Prostorová popisná statistika*

**Směrodatná vzdálenost** je míra rozptýlu hodnot kolem geografického středu. Jedná se o kružnici kolem geografického středu, jejímž poloměrem je hodnota směrodatné vzdálenosti (Netrdová nedatováno). U všech pacientů je směrodatná vzdálenost 15 229.5 m, u pacientů s překročenou normou 2 a vícekrát je hodnota 15 650.1 m. Rozdíl mezi směrodatnými vzdálenostmi je přibližně 400 m. U pacientů s překročenou normou 2 a více je rozptýl kolem geografického středu větší než u všech vyšetřených pacientů. Tento výsledek potvrzuje předchozí zjištění z výsledků zkoumání polohy geografického středu, tedy ten, že pacienti jsou rozmístěni po celém okrese.

**Směrodatná elipsa** odchylek vykazuje, zda prostorové rozmístění jeví vykazuje rysy směrovosti (Netrdová nedatováno). Směrodatná elipsa všech vyšetřených pacientů je užší a mírně protáhlejší než směrodatná elipsa pacientů s překročenou normou 2 a více. Nejvýraznější rozdíl je v jejich rotaci. Rotace elipsy, znázorňující pacienty s překročenou normou 2 a vícekrát, je posunuta přibližně o 10 ° protisměru hodinových ručiček. Tento výsledek naznačuje rozmístění pacientů s překročenou normou hlavně v oblasti měst Příbram a Sedlčany.

## 7.2 Konfirmační analýza prostorové statistiky

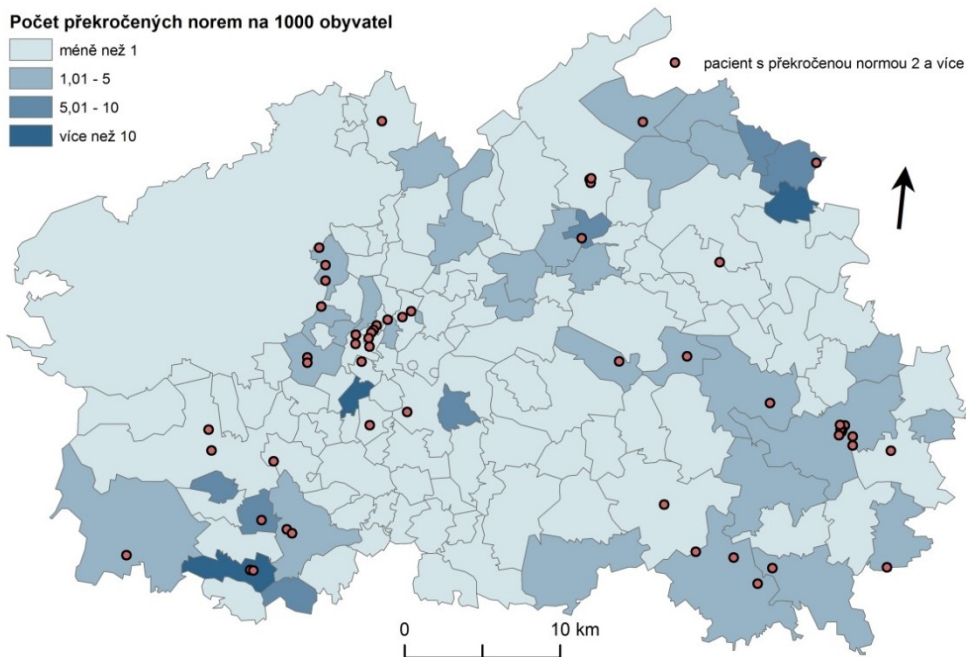
V programu ArcGIS Desktop byla znázorněna koncentrace bodů funkcí Kernel Density (Spatial Analyst Tools/ Density/ Kernel Density). Výsledek byl vážen počtem překročených norem. Na Obr. 12 je znázorněna koncentrace překročených norem na 1 km<sup>2</sup> u vyšetřených pacientů žijících v okrese Příbram. Zároveň jsou zde zobrazena města s počtem obyvatel vyšší než 2 000. Na této mapě je viditelná vysoká koncentrace ve městech Příbram, Sedlčany, Dobříš a Milín. Ve městě Příbram je nejvyšší koncentrace počtu překročených norem, druhá nejvyšší je ve městě Sedlčany. Tento výsledek podporuje tvrzení vycházející z tvaru směrodatné elipsy, že pacienti s překročenou normou se nacházejí hlavně v oblasti města Příbram a Sedlčany. Koncentrace 5.01 až 10 překročených norem na 1 km<sup>2</sup> je pozorovaná ve města Rožmitál pod Třemšínem a Milín. Stejná koncentrace, jako v těchto městech, je v obcích Bohutín, Dubovice a Petrovice. Městys Jince, který má počet obyvatel podobný jako Milín, má koncentraci překročených norem na 1 km<sup>2</sup> nižší než 5.



**Obr. 12 Koncentrace překročených norem na 1 km<sup>2</sup> u pacientů v okrese Příbram**

Následně byl vypočítán počet překročených norem na počet obyvatel v dané obci. Účelem tohoto kroku byla eliminace vlivu většího počtu obyvatel ve městech než v obcích. Město Příbram bylo pro větší přehlednost rozděleno na jednotlivé městské části a katastrální území. Na Obr. 13 je pomocí kartogramu znázorněn počet překročených norem u parametrů štítné žlázy na 1000 obyvatel a pomocí bodů jsou vyneseni pacienti s překročenou normou 2 a vícekrát. Největší překročení norem na počet obyvatel bylo zjištěno v obci Hudčice (22.4)

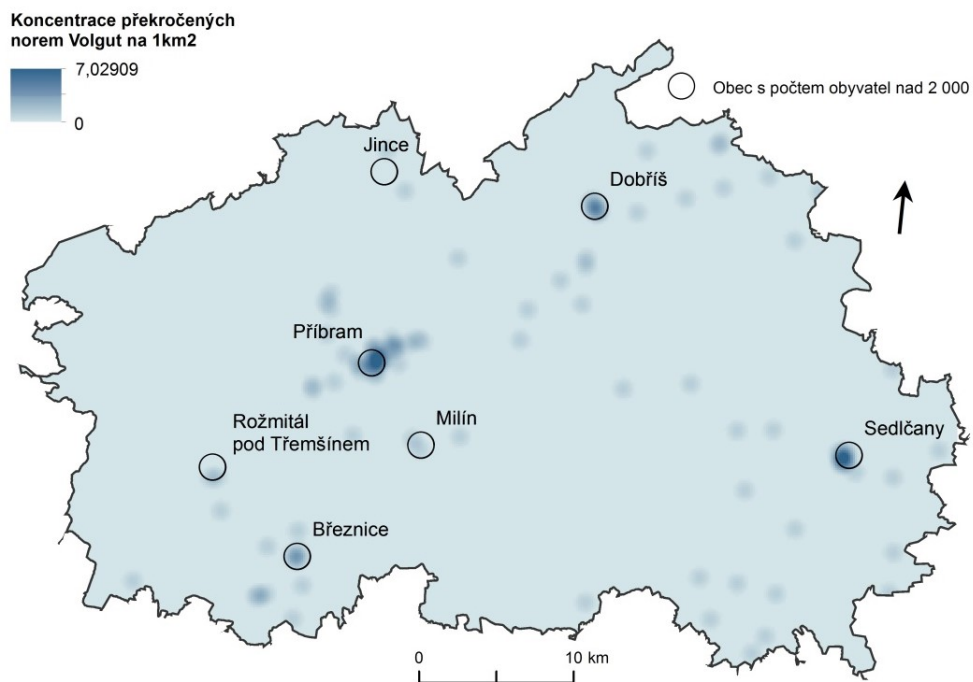
a v obci Hlubyně (22.1). Celkově bylo u 9 obcí zjištěno 10 a více překročení norem na 1000 obyvatel. Oproti mapě znázorňující koncentraci překročených norem (Obr. 12 ) na Obr. 13 nejsou tak výrazná města Příbram, Dobříš a Sedlčany. V těchto městech i v jejich okolí je počet překročených norem většinou vyšší než 1 na 1000 obyvatel. V Příloze 5 jsou zaznamenány vypočítané hodnoty překročených norem na 1000 obyvatel za jednotlivé obce. Jsou zde hodnoty celkového součtu překročených norem a hodnoty parametrů Volgut, TSH, fT3 a fT4. Obce, které nejsou v příloze jmenovány, měly všechny hodnoty rovny 0.



**Obr. 13 Počet překročených norem na 1000 obyvatel v okrese Příbram**

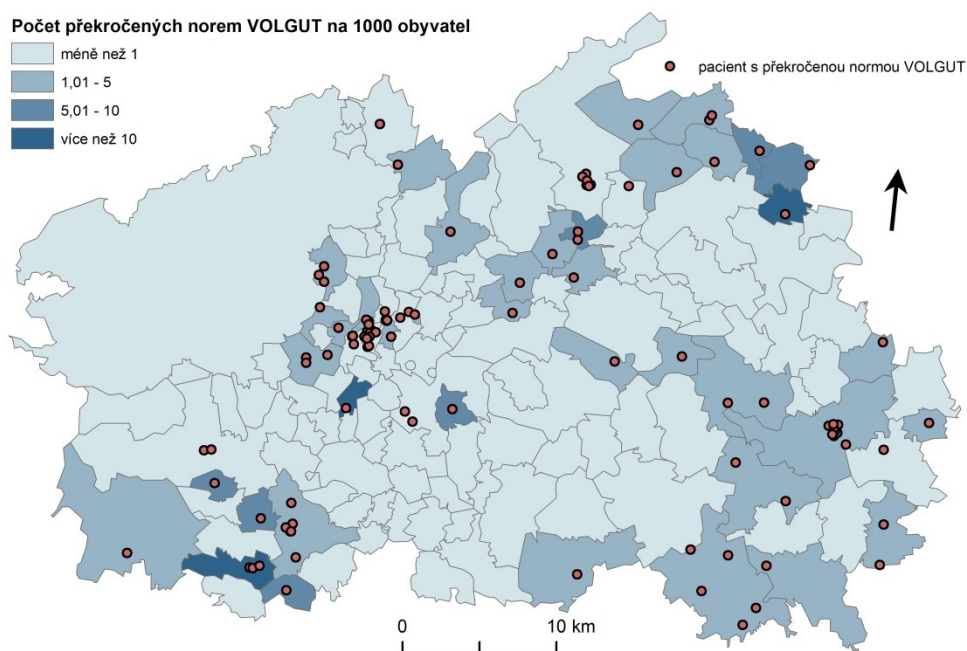
### **Objem štítné žlázy (Vulgut)**

Na Obr. 14 znázorňujícím koncentraci překročených norem Volgut na 1 km<sup>2</sup> mají nejvyšší hodnoty města Příbram, Dobříš, Sedlčany a Březnice. Zároveň jsou zde malé oblasti, které mají zvýšenou koncentraci. Opět v obci Jince je pozorována nejnižší koncentrace překročených norem objemu štítné žlázy z obcí s počtem obyvatel více než 2000.



**Obr. 14 Koncentrace překročených norem Volgut na 1 km<sup>2</sup> u pacientů v okrese Příbram**

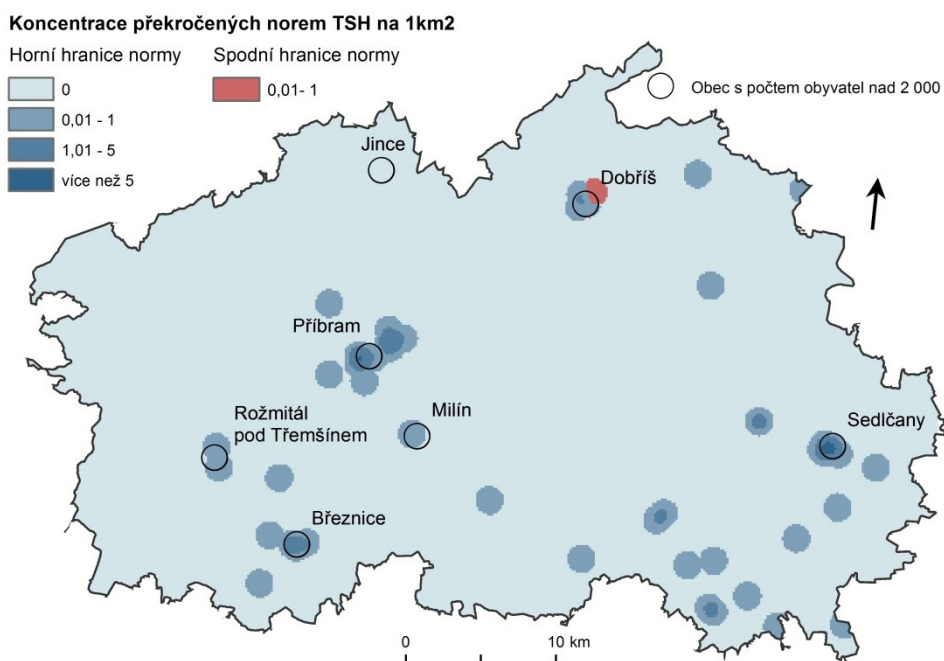
Počet překročených norem Volgut byl vážen počtem obyvatel v dané obci a následně byl vytvořen kartogram (Obr. 16 ). Na této mapě jsou znázorněni pacienti s překročenou normou Volgut. Počet překročených norem Volgut 10 a více na 1000 obyvatel je v obcích Zavržovice (19.3), Hudčice (13.4) a Korkyně (10.2). V oblasti velkých měst je pouze mírně zvýšen počet překročených norem Volgut. Veškeré výsledky jsou k dispozici v Příloze 5.



**Obr. 15 Počet překročených norem Volgut na 1000 obyvatel v okrese Příbram**

## TSH

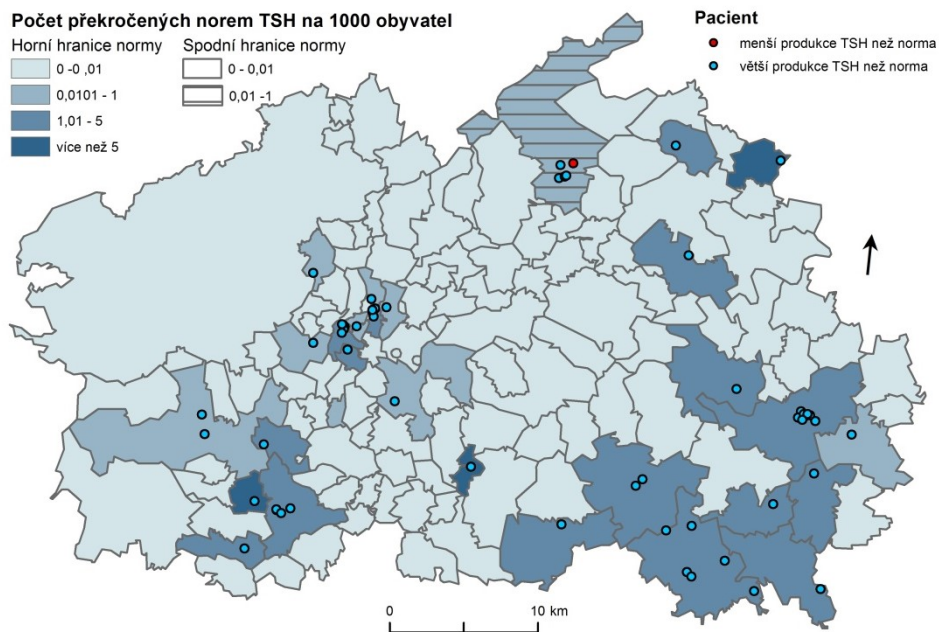
Koncentrace překročených norem TSH na 1 km<sup>2</sup> je znázorněna na Obr. 16. Hodnoty větší než 0 znázorňují oblasti se zvýšenou produkcí hormonu TSH, naopak červeně znázorněná místa identifikují oblasti se sníženou produkcí hormonu TSH. Koncentrace vyšší než 5 je zaznamenána ve městě Sedlčany a Příbram, tedy vyskytují se zde pacienti, kteří mají zvýšenou produkci TSH než je norma. Koncentrace 2 až 5 překročených norem TSH na km<sup>2</sup> je ve městě Březnice. Naopak pacienti se sníženou produkcí TSH jsou identifikováni v oblasti města Dobříš.



**Obr. 16 Koncentrace překročených spodních a horních hranic norem TSH na 1 km<sup>2</sup> u pacientů v okrese Příbram**

Počet překročených norem TSH vážený počtem obyvatel je znázorněn na Obr. 17. Stejně jako u koncentrace překročených norem na Obr. 16 je i zde identifikováno město Dobříš jako oblast se sníženou produkcí TSH než je norma. Při vážení počtem obyvatel vychází tato oblast s hodnotou 0,2, což je velice blízké 0. Je to dáno tím, že pacienti se sníženou produkcí TSH se vyskytují pouze v této oblasti a ještě v malé míře. Naopak zvýšená produkce TSH než je norma, je identifikována v obci Hlubyně (7.4), Cetyně (6.2) a Nové Dvory (5.2). Města Sedlčany a Příbram mají hodnoty mírně překračující 1 normu na 1000 obyvatel. U překročení norem celkově a u Volgut byla vždy identifikována obec Hudčice. U parametru TSH má obec Hudčice 4.5 překročených norem na 1000 obyvatel.

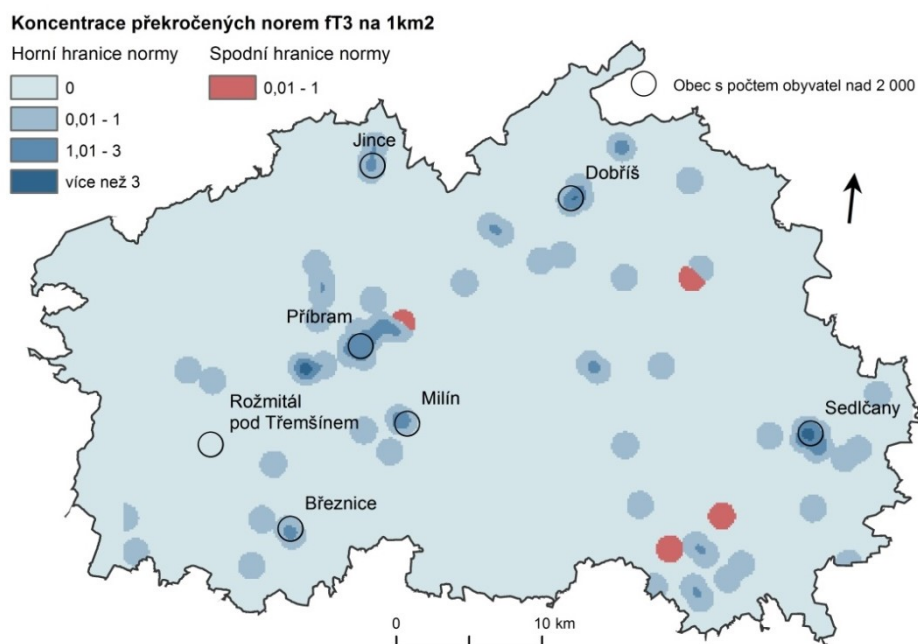




**Obr. 17 Počet překročených spodních a horních hranic norem TSH na 1000 obyvatel v okrese Příbram**

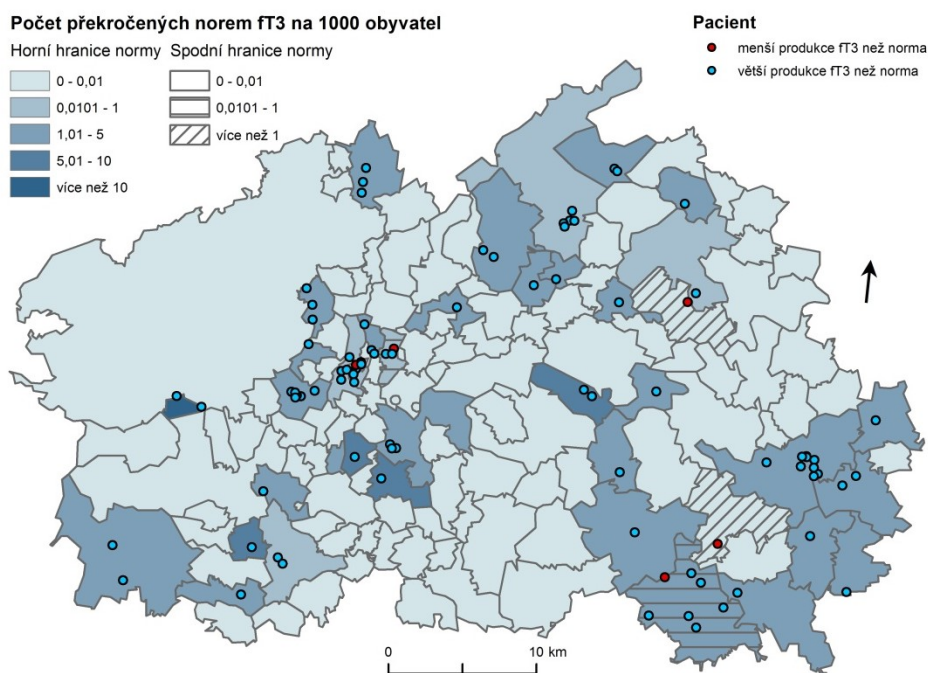
### ft3

Koncentrace překročených norem ft3 je znázorněna na Obr. 18 . Hodnota větší než 3 je zaznamenána ve městě Sedlčany a Dobříš a v obci Bohutín. Ve městě Příbram je zaznamenána hodnota 2.6 překročených norem ft3 na 1 km<sup>2</sup>. V těchto obcích se tedy nachází více pacientů, kteří mají zvýšenou produkci hormonu ft3 než je norma. Naopak pacienti se sníženou produkcí hormonu ft3 než je norma se nachází v obcích Borotice, Vysoký Chlumeč a Petrovice. Ve městě Příbram byla zaznamenána zvýšená i snížená produkce hormonu ft3 než je norma.



**Obr. 18 Koncentrace překročených spodních a horních hranic norem fT3 na 1 km<sup>2</sup> u pacientů v okrese Příbram**

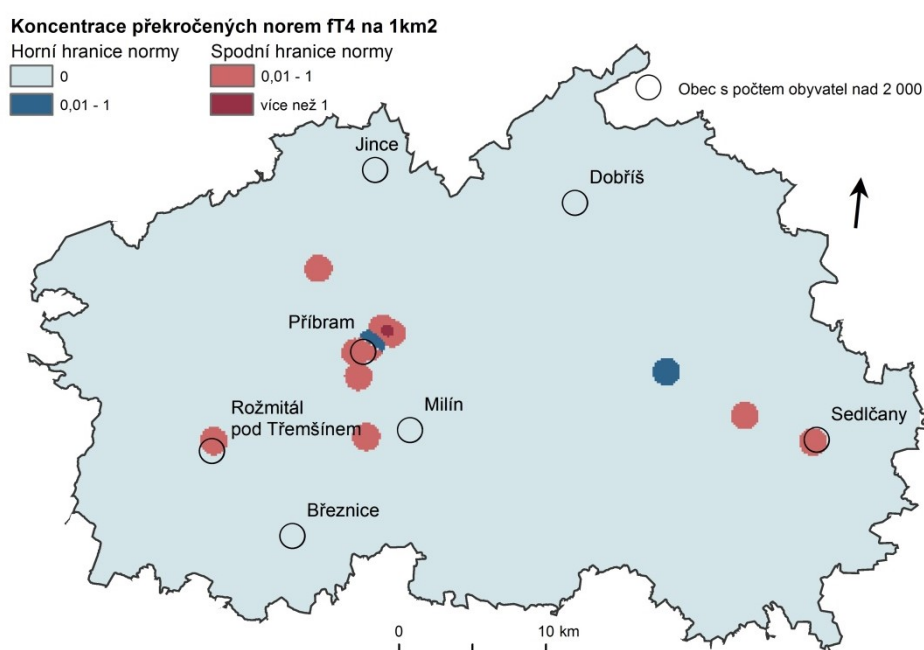
Při vytvoření kartogramu, znázorňujícího počet překročených norem fT3 na 1000 obyvatel v jednotlivých obcích okresu Příbram (Obr. 19 ), byly identifikovány obce Borotice, Vysoký Chlumeck a městská část Příbram 2, jako místa se sníženou produkcí fT3. Naopak zvýšená produkce hormonu fT3 překračující normu byla identifikována v obcích Nepomuk (10.7), Obory (8.6), Hlubyně (7.4) a Vrančice (7.3). Obec Hlubyně měla nejvícekrát překročenou normu TSH v celém okrese, je tedy již po druhé označena za rizikovou oblast.



**Obr. 19 Počet překročených spodních a horních hranic norem fT3 na 1000 obyvatel v okrese Příbram**

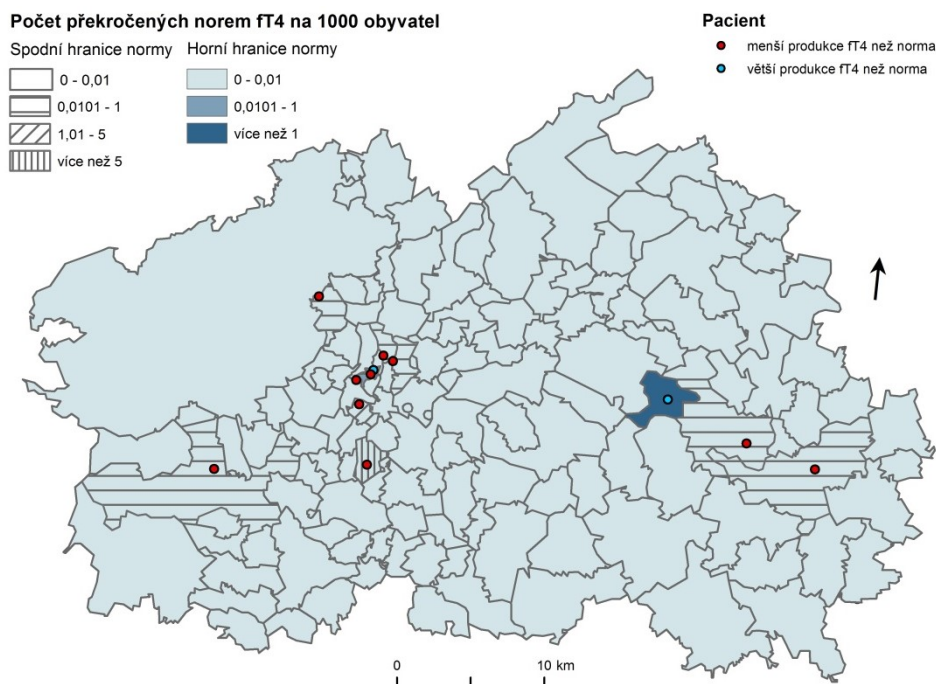
## ft4

Pro ukazatel funkce štítné žlázy ft4 byla také vytvořena mapa znázorňující koncentraci překročených norem na 1 km<sup>2</sup> (Obr. 20 ). Zde není identifikována jediná oblast s hodnotou vyšší než 1 při překročení horní hranice normy ft4. Naopak oproti mapám ft3, Volgut a TSH je identifikováno větší množství oblastí, kde je překročena spodní hranice normy v produkci ft4. Tato hranice je nejvíce překročena v severní části města Příbram (1.4). Dále je zvýšená koncentrace v počtu překročených spodních hranic normy ve městě Sedlčany a Rožmitál pod Třemšínem a v obcích Lazsko, Dublovice a Obecnice.



**Obr. 20 Koncentrace překročených spodních a horních hranic norem ft4 na 1 km<sup>2</sup> u pacientů v okrese Příbram**

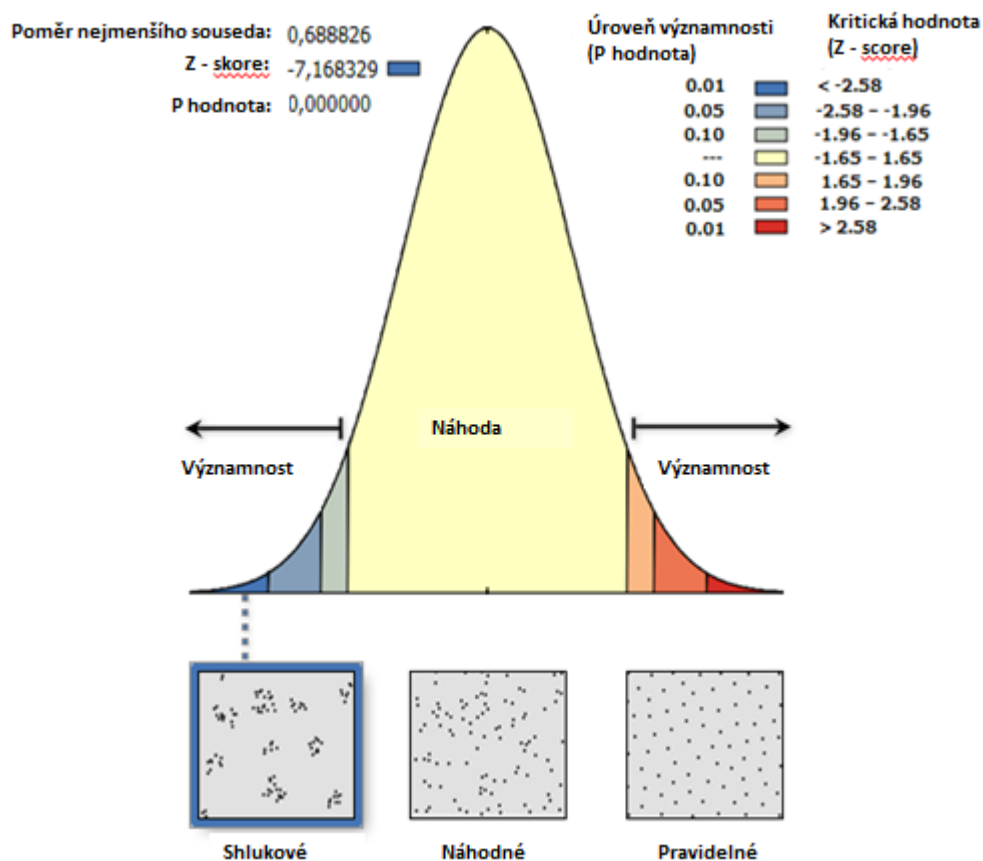
Dále byl vytvořen kartogram značící počet překročených norem ft4 na 1000 obyvatel. Obec Hříměždice má hodnotu 2.6 překročených horních hranic normy ft4 na 1000 obyvatel. V okrese Příbram se nachází pouze 2 pacienti, kteří mají zvýšenou hodnotu ft4 než je norma. První pacient žije v obci Hříměždice a druhý v městské části Příbram 7. Překročení dolní hranice normy parametru ft4 vážené počtem obyvatel má podobné rozložení jako u mapy koncentrace překročených norem na 1 km<sup>2</sup>. Nejvyšší hodnotu má obec Lazsko (6.3) a Žežice (2.1).



**Obr. 21 Počet překročených spodních a horních hranic norem fT4 na 1000 obyvatel v okrese Příbram**

### 7.2.1 Metoda nejbližšího souseda

Následně byla aplikována metoda nejbližšího souseda. Do této metody byli zahrnuti jen pacienti s překročenou normou 2 a vícekrát. U těchto pacientů je pravděpodobné, že trpí onemocněním štítné žlázy. Jelikož nejsou k dispozici lékařské závěry z vyšetření, je tento způsob nejvhodnějším řešením, jak nemocné pacienty identifikovat. Na Obr. 22 je vidět výsledný protokol metody. V okrese Příbram jsou pacienti s překročenou normou 2 a vícekrát rozmístěni ve shlucích. R je 0.688 a značí shlukové rozmístění, kdyby  $R = 0$ , jednalo by se o zcela shlukové rozmístění. P hodnota je 0.0, jedná se tedy o platný výsledek na hladině pravděpodobnosti 95 %. Z - skóry jsou -7.2 a indikují méně než 1% pravděpodobnost, že shluky jsou výsledkem náhody. Výsledky této metody se ztotožňují se znázorněním koncentrace překročených norem u pacientů v okrese Příbram.



Obr. 22 Výsledek metody nejbližšího souseda

V 0 jsou zaznamenány veškeré výsledky metody nejbližšího souseda. U parametrů Volgut, TSH a fT3 je výsledkem shlukové rozmístění překročených norem těchto ukazatelů na hladině pravděpodobnosti 95 %. U parametru fT4 se jedná o rozmístění náhodné.

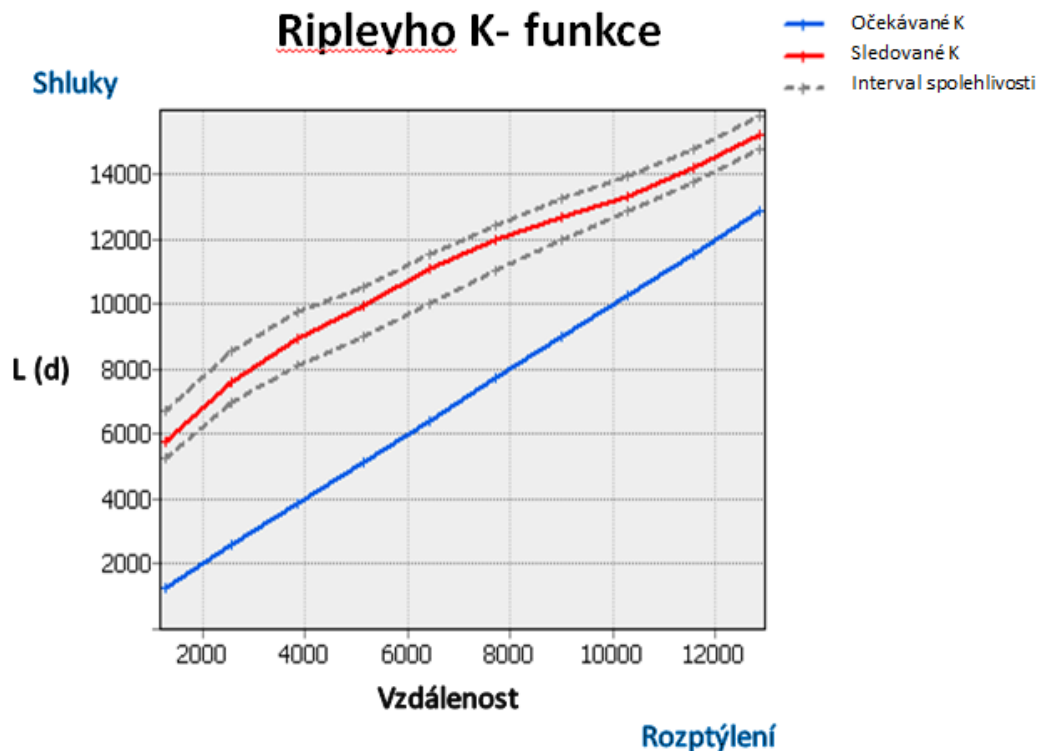
	R	z- skóre	P hodnota	Rozmístění
Celkem	0.689	-7.168	0	shlukové
Volgut	0.698	-5.855	0	shlukové
TSH	0.651	-5.21	0	shlukové
fT3	0.595	-7.285	0	shlukové
fT4	1.094	0.622	0.534	náhodné

Tab. 7. Metoda nejbližšího souseda

## 7.2.2 Ripleyho K - funkce

K - funkce odhaluje lokální odchylky v prostorové distribuci. Tato metoda byla vážena počtem překročených norem u pacientů. Výsledkem je Graf 3. Průběh sledovaných hodnot leží výrazně nad křivkou očekávaných hodnot, což značí lokální shlukové uspořádání. Kdyby byla křivka sledovaných hodnot pod křivkou očekávaných, jednalo by se o náhodné uspořádání.

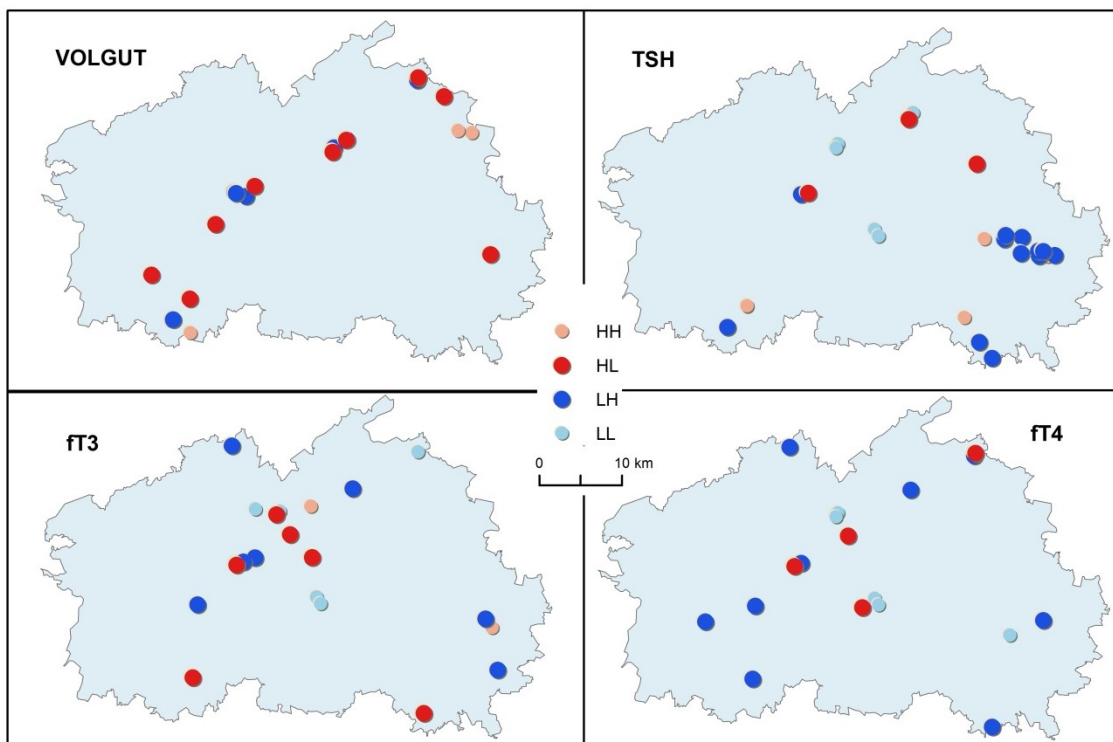
Výsledky této metody taktéž potvrzují předchozí tvrzení o výskytu pacientů s překročenou normou ve shlucích.



**Graf 3. Ripleyho K- funkce**

### 7.2.3 Lokální Moranovo I

Lokální Moranovo I identifikuje statisticky významná seskupení prvků s nízkými nebo vysokými hodnotami a prostorové odchylky. Na Obr. 23 jsou znázorněny typy prostorové asociace pro jednotlivé parametry štítné žlázy. Jednotlivé typy byly určovány na hladině významnosti 95 %. U parametru Volgut je identifikováno 7 oblastí, kde se vyskytují zvýšené hodnoty parametru. Tyto oblasti se nachází v blízkosti měst Příbram, Dobříš a Březnice. U parametru TSH bylo identifikováno 14 oblastí se zvýšenou hodnotou parametru, než je okolí. Jedná se o oblasti v jihovýchodní části okresu Příbram. Oblastí se sníženou hodnotou TSH bylo označeno 7, a to hlavně v severní a středové části okresu. U parametru fT3 byly identifikovány 2 oblasti se zvýšenou hodnotou, než je okolí a 7 oblastí se sníženou hodnotou, než je okolí. U parametru fT4 nebyly vybrány žádné oblasti se zvýšenou hodnotou, než je hodnota okolí. Bylo identifikováno 6 oblastí se sníženou hodnotou, než je okolní hodnota. U parametrů TSH, fT3 a fT4 byly identifikovány oblasti se sníženou hodnotou vždy v severní části a ve středu okresu.



**Obr. 23 Typ prostorové asociácie – Volgut, TSH, fT3, fT4**  
(Pozn.: HH-vysoká- vysoká, HL- vysoká- nízka, LH- nízka- vysoká, LL- nízka- nízka)

## 8 Analýza výskytu nemoci s faktory prostředí

V předchozí kapitole 7 byly identifikovány shluky v okolí větších měst v okrese Příbram. Následně byli pomocí map zobrazené hodnoty překročených norem jednotlivých parametrů štítné žlázy vážené počtem obyvatel v jednotlivých obcích. Při vizuálním porovnání pravděpodobně nemocných pacientů a prostorových aspektů byl nalezen možný vliv životního prostředí na nemoc. Z těchto důvodů jsou v této části práce vytvořeny analýzy, které definují dané vztahy. V následujících podkapitolách je rozebrán postup analýz výskytu nemoci s vnějšími faktory. V závěru kapitoly výsledky znázorněny v prostoru.

Pro analýzu výskytu nemoci s vnějšími faktory bylo důležité vybrat nejvhodnější metodu. V této části práce byla využita vícerozměrná regrese s redukcí dimenzionality, metoda obousměrných ortogonálních projekcí do latentní struktury (O2PLS). Často se k podobným účelům využívají také analýzy hlavních komponent (PCA) nebo faktorová analýza (FA).

Analýza hlavních komponent (Principal Component Analysis, zkratka PCA) je statistická metoda, sloužící k nalezení transformací které převádějí původní matici korelovaných proměnných (znaků)  $\mathbf{X}$  (sloupce matice představují jednotlivé znaky) na novou matici nekorelovaných proměnných  $\mathbf{T}$ , nazývaných hlavními komponentami (Meloun, Militký a Hill 2005). Tyto komponenty jsou extrahovány v pořadí postupně se snižující důležitosti (relevance) s cílem redukovat dimenzionalitu dat. Ačkoliv transformace produkuje přesně tolik hlavních komponent, kolik je proměnných v matici  $\mathbf{X}$ , prakticky se využívá jen určitý počet relevantních hlavních komponent, nejlépe 2-3, které vysvětlují optimální množství variability u proměnných matrice  $\mathbf{X}$ . K tomuto účelu lze využít tzv. vlastních čísel  $\lambda_m$  ( $m$  je pořadí hlavních komponent v matici  $\mathbf{T}$ ) (Meloun, Militký a Hill 2005). Hodnota  $\lambda_m$  vyjadřuje podíl variability vysvětlené příslušnou komponentou k celkové variabilitě v matici  $\mathbf{X}$  či  $\mathbf{T}$ . Relevance hlavních komponent se zjistí z tzv. Catellova grafu konstruovaného jako jednotlivá vlastní čísla  $\lambda_m$  či podíly celkové variability vysvětlené jednotlivými hlavními komponentami na ose y proti pořadí hlavních komponent  $m$  na ose x. Relevantní jsou ty hlavní komponenty, jejichž zahrnutím ještě podstatně klesne podíl nevysvětlené variability v matici  $\mathbf{X}$  či  $\mathbf{T}$ . To znamená, že místo zlomu v grafu označuje počet relevantních hlavních komponent. Další možností je užití tzv. Kaiserova kritéria, kdy jsou za relevantní považovány ty hlavní komponenty, jejichž vlastní čísla  $\lambda_m > 1$  (Meloun, Militký a Hill 2005).

Matice hlavních komponent  $\mathbf{T}$  jsou lineárními kombinacemi původních proměnných v matici  $\mathbf{X}$  a maticí tzv. komponentních vah což znamená, že matice  $\mathbf{T}$  po pronásobení kompletní transponovanou maticí  $\mathbf{P}'$  dává matici  $\mathbf{X}$  dle vzorce 12:



$$X = TP' \quad (12)$$

V případě, že se využije jen část komponent, tj. relevantní komponenty a zbylé se zanedbají, přejde vzorec na tvar:

$$X = TP' + E \quad (13)$$

kde  $E$  je matice reziduí (nevysvětlený rozptyl, odchylky měřených hodnot od predikcí). PCA také může sloužit k nalezení závislostí mezi proměnnými matice  $X$ . Výše uvedené vzorce platí pro normovaná data. Od nenormovaných dat v matici  $X$  se odečtou aritmetické průměry příslušných proměnných a podělí se příslušnými směrodatnými odchylkami. Komponentní váhy pro jednotlivé proměnné pak nabývají hodnot od -1 (absolutní negativní korelace příslušné hlavní komponenty s příslušnou původní proměnnou) do 1 (absolutní pozitivní korelace příslušné hlavní komponenty s příslušnou původní proměnnou). Komponentní váha  $P_{m,j}$  tedy odráží vliv  $j$ -té původní proměnné (znaku) na  $m$ -tou hlavní komponentu (Meloun, Militký a Hill 2005). FA provede stejný postup, ale poté dále pracuje, než se dobere k výsledku. Faktorová analýza je o něco matematicky dokonalejší než analýza hlavních komponent, ale plní stejný účel a v zásadě je stejný i jejich výstup.

FA lze využít podobným způsobem jako PCA, i když matematická podstata je poněkud odlišná a výsledky obvykle bývají lépe interpretovatelné, než v případě PCA, zvláště po provedení tzv. rotace faktorů (podrobnosti lze nalézt v knize Meloun, Militký a Hill 2005).

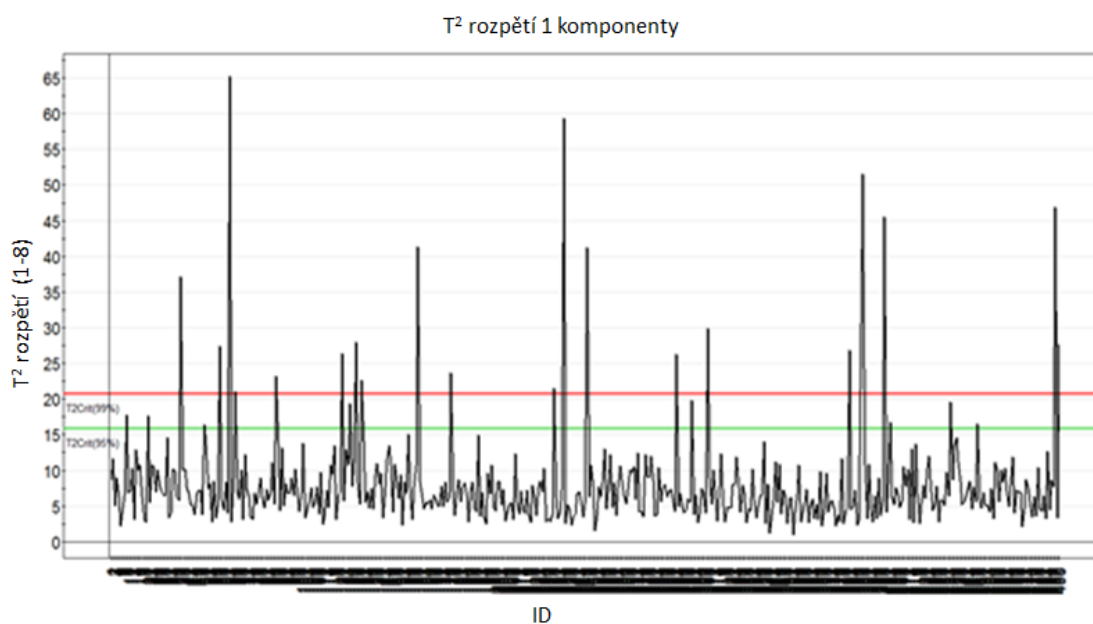
V této práci se ale ptáme přímo na vztahy mezi skupinami proměnných (medicinální data vs. charakteristiky životního prostředí). Z tohoto důvodu metody PCA a FA nejsou optimální z hlediska kladené otázky. Obě metody by sice bylo možné použít, ale výsledek by byl hůře interpretovatelný a méně adekvátní kladené otázce, než v případě O2PLS. Cílem O2PLS není interpretace vztahů uvnitř matice prediktorů  $X$  či predikovaných proměnných  $Y$ , i když tyto vztahy nezávisle na vztazích mezi maticemi  $X$  a  $Y$  jsou vysvětlovány v ortogonálních komponentách za účelem konstrukce predikčního modelu. Model O2PLS je, na rozdíl od běžných regresních modelů, schopen pracovat se silnou korelovaností uvnitř matic  $X$  a  $Y$ , tzv. multikolinearitou a model nepřestává fungovat ani v případě výskytu lineárních funkčních vztahů uvnitř těchto matic. Navíc model O2PLS nerozlišuje prediktory a predikované proměnné a je bez problémů schopen vysvětlit vztahy  $Y$  proti  $X$  i naopak.

Výsledkem uvedeného postupu je v této práci regresní model, který určuje parametry ovlivňující ukazatele funkce štítné žlázy.

Tato kapitola byla vypracována v programu Microsoft Excel 2010, IBM SPSS Statistics 23, Statgraphics Centurion XV od firmy Statpoint Inc. (Herndon, MA, USA) a SIMCA P+ 12.0 od firmy Umetrics (Umeå, Švédsko).

## 8.1 O2PLS

V programu SIMCA P+ verze 12.0 byla analyzována data transformovaná směrem k symetrii rozdělení konstantnímu rozptylu (homoskedasticitě). Nejprve byly z analýzy vyloučeny parametry, které mají více než 50 % chybějících údajů. Následně byla provedena analýza s využitím vícenásobné regrese s redukcí dimenzionality (metoda obousměrných ortogonálních projekcí do latentní struktury, O2PLS). Byly vybrány matice závisle proměnných  $Y$  a nezávisle proměnných (prediktorů)  $X$ . Poté byl spuštěn algoritmus analýzy metodou O2PLS (Obr. 24). Nejprve byla zkontrolována homogenita v matici  $X$  (prediktory, nezávisle proměnné) s využitím Hotellingovy  $T^2$  statistiky. Hotellingova  $T^2$  statistika je odvozena z Mahalanobisovy vzdálenosti a má F-rozdělení. Mahalanobisova vzdálenost, na rozdíl od vzdálenosti euklidovské, uvažuje také korelovanost mezi proměnnými. Odlehlé body detekované Hotellingovou  $T^2$  statistikou, je nutné z další analýzy vyřadit. Na Graf 4 je zaznamenána Hotellingova  $T^2$  statistika, červená příčka označuje hladinu významnosti  $p < 0.01$ . S ohledem na počet dat jsme hranici pro eliminaci nehomogenit posunuli odhadem na zhruba  $p < 0.002$ . V tomto případě bylo odstraněno kolem 15 bodů.



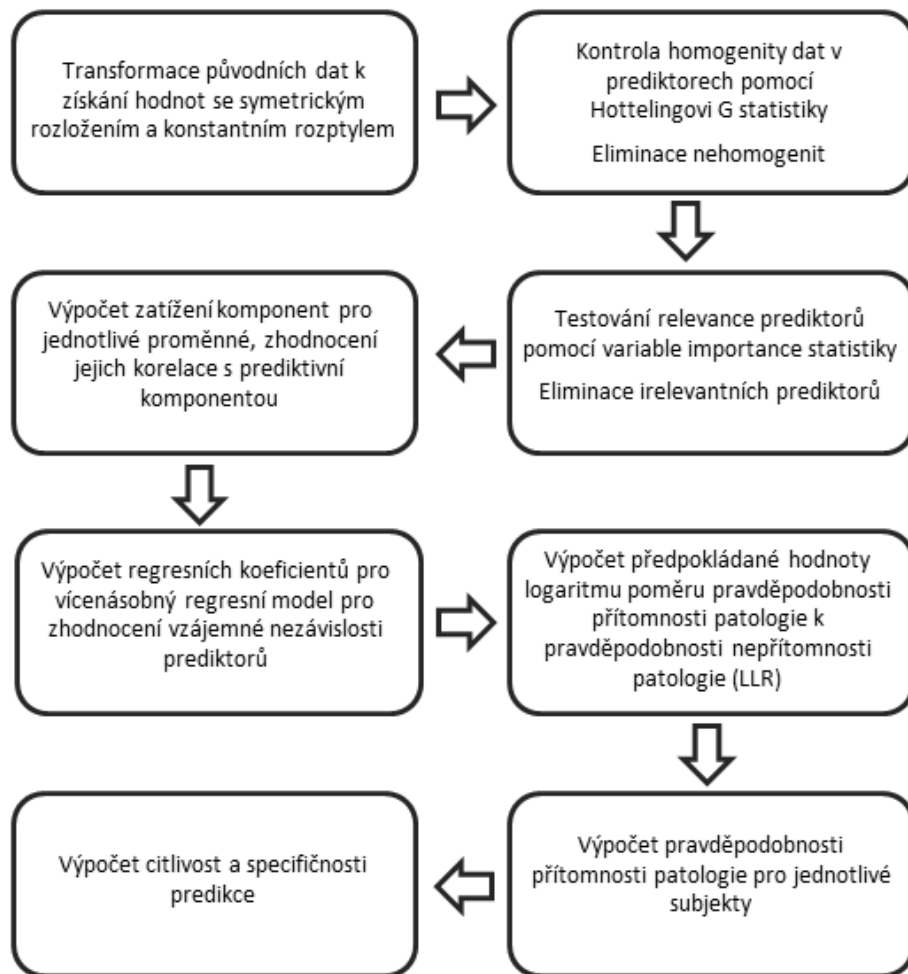
**Graf 4. Hotellingova  $T^2$  statistika**

Po vyřazení všech odlehlých bodů se opět spouští O2PLS a opět se zobrazí Hotellingovy  $T^2$  statistiky. Tento postup se opakuje, dokud nejsou všechny body s  $p < 0.002$  odstraněny. Obvykle nelze všechny nehomogenity detekovat napoprvé s důvodu tzv. maskování dalších nehomogenit výraznými nehomogenitami a teprve po odstranění extrémních nehomogenit se tyto, zatím skryté nehomogenity, projeví. Následně je aplikována statistika *VIP* (variable importance), pro ověření relevance jednotlivých prediktorů. Do další analýzy jsou pak zařazeny jen prediktory relevantní. Ostatní, které se významně ( $p < 0.05$ ) nepodílejí na konstrukci modelu, jsou z další analýzy vyloučeny. Kdyby se do analýzy zahrnuly i nerelevantní proměnné, analýza by se snažila „vysvětlit nevysvětlitelné“ a vnášel by se pouze další nevysvětlený rozptyl. V důsledku toho by se pak snižovala prediktivita modelu. *VIP* statistiky se testují ve všech prediktivních komponentách. Za nerelevantní prediktory se považují jen ty, které nenabudou významnosti u žádné z významných prediktivních komponent. Významnost a tedy i počet prediktivních komponent program SIMCA stanovuje v automatickém režimu.

V dalším kole jsou opět testovány Hotellingovy  $T^2$  statistiky v upraveném modelu. Postup je ukončen v okamžiku, kdy nejsou nalezeny k vyřazení žádné experimentální body ani proměnné v matici prediktorů  $X$ . Následně je nutné provést záměnu matice prediktorů  $X$  a matice predikovaných proměnných  $Y$ . Postup se aplikuje u matice závislé proměnných  $Y$  stejně, jako tomu bylo u matice  $X$ . V okamžiku, kdy v systému nejsou nerelevantní proměnné ani odlehlé body, se proces ukončí.

Pak se opět zamění matice  $X$  za  $Y$  a spouští se testování homogenity prediktorů a relevance proměnných. Celý postup je doveden ke konečnému výsledku v okamžiku neexistence nerelevantních proměnných a nehomogenit v maticích  $X$  i  $Y$ . Následně se zobrazí rezidua (rozdíly mezi měřenými a predikovanými hodnotami) u závisle proměnných. Je zobrazen kvantil-kvantilový graf pro rezidua jednotlivých závisle proměnných a odlehlé body příslušející těmto reziduím se z analýzy vyloučí. Jedná se o data, která nepatří do populace. Algoritmus je doveden do cíle v okamžiku, kdy neexistují nerelevantní proměnné v modelu OPLS, nehomogenity v maticích  $X$  a  $Y$  a analýza reziduí nenachází nehomogenity v jednotlivých proměnných matice závisle proměnných.

Jednotlivé části postupu jsou znázorněny na Obr. 24 .



*Obr. 24 Algoritmus pro analýzu výskytu nemoci s vnějšími faktory*

### 8.1.1 O2PLS při nastavení závislé proměnné všech zdravotních parametrů

Jako závislé proměnné  $Y$  byly vybrány lékařské parametry a jako nezávislé proměnné  $X$  byly vybrány ukazatele životního prostředí. Dříve popsany postup byl proveden 6krát tak, abymohl být ukončen a interpretován. Po úpravě zbylo pouze 10 proměnných v matici  $X$  a 20 proměnné v matici  $Y$  zbyly 2 proměnné, a to ukazatel alergik, který říká, jestli pacient trpí alergií a ukazatel LDL představující hladinu nízkodenzitního a tedy nežádoucího cholesterolu. Model ukazoval, že prostředí stejně ovlivňuje alergiky a hladinu nízkodenzitního cholesterolu, ato významně negativně i pozitivně. Přesto že tyto výsledky jsou zajímavé, nejsou účelem této diplomové práce. Z tohoto důvodu bylo nutné změnit nastavení proměnných  $X$  a  $Y$ .

### 8.1.2 O2PLS při nastavení závisle proměnných parametrů štítné žlázy

Tato diplomová práce se zabývá možností vlivu životního prostředí na funkci štítné žlázy. Z tohoto důvodu jako závisle proměnné  $Y$  byly vybrány proměnné Volgut, fT3, fT4, ras, uzle, tgl, ldl, tpo, trigl, TSH (vysvětlení názvů proměnných v Příloze 2). Za nezávisle proměnné  $X$  byly označeny parametry životního prostředí, dále proměnné vek, muz, bmi a jodur (vysvětlení názvů proměnných v Příloze 2).

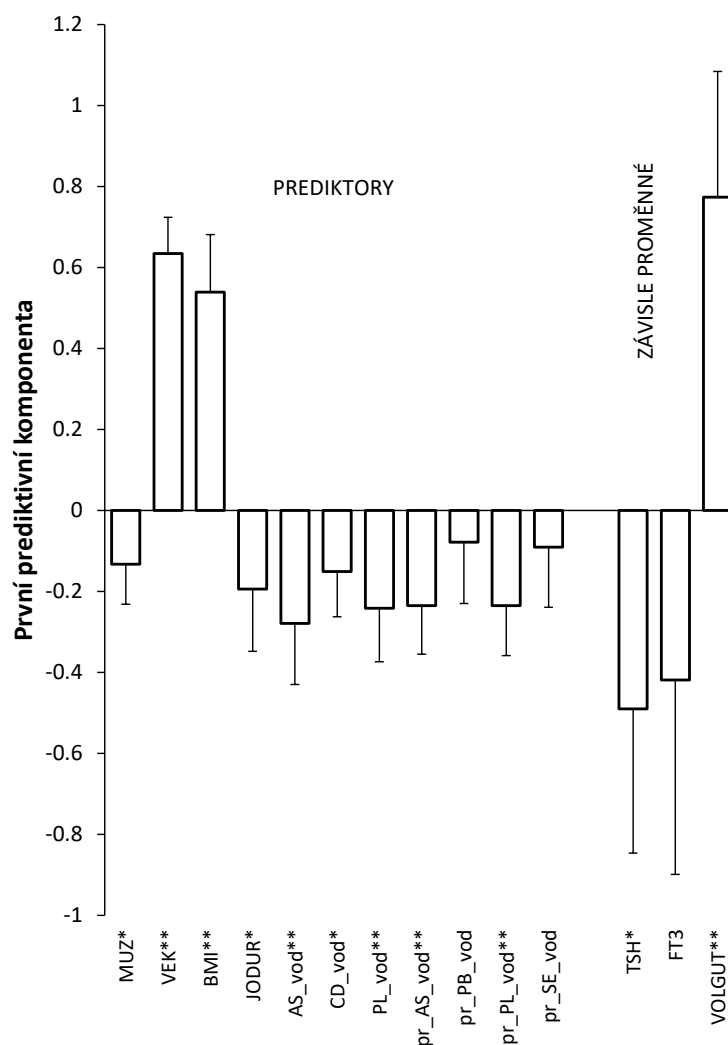
Při analýze metodou O2PLS byla nejprve zkontrolována homogenita v matici prediktorů  $X$  s využitím grafu Hottelingových  $T_i^2$  statistik pro každý bod. Poté byla zkontrolována relevance proměnných s využitím statistik Variable Importance  $VIP_{j,k}$  pro  $k$  relevantních komponent všech prediktorů  $X_j$  a v matici  $X$ , byly jako nehomogenity detekovány a poté z další analýzy odebrány pouze 2 body, avšak bylo odebráno zhruba 65 % proměnných, které nebyly relevantní. Po provedení stejného postupu v případě matice závisle proměnných  $Y$  bylo odebráno 10 bodů a 7 proměnných. Zbyly pouze proměnné Volgut, fT3 a TSH. Následně byl opět opakován tento postup pro matici  $X$ , kde byla odebrána další proměnná. Při opakování postupu pro matici  $X$  již nebyly nutné žádné změny a přistoupilo se ke kontrole homogenity reziduí s využitím kvantil-kvantilového grafu. Poté byly 4 experimentální body s nehomogenními rezidui z další analýzy odebrány. Následně byl zopakován postup pro matici  $X$ , kde byly odebrány ještě další 3 nerelevantní proměnné. Dále pak proběhla finální kontrola, která potvrdila, že není nutno provádět další změny. Na základě nevýznamných hodnot VIP statistik byly z modelu odstraněny všechny nerelevantní ukazatele vlivu prostředí.

Výsledkem je vznik regresního modelu o dvou komponentách. První komponenta, zachycená v 0, vysvětluje 26 % variability modelu (24 % variability po křížové validaci). Vysokou míru vlivu ( $R$ ) má proměnná věk a BMI. Zároveň je u těchto ukazatelů i vysoká míra statistické jistoty ( $t$ -statistika). Z proměnných vnějšího prostředí mají největší míru vlivu arsen ve vodě (samostatný za místo bydliště i zprůměrovaný s polohou školy) a polutanty ve vodě (samostatné za místo bydliště i zprůměrované s polohou školy). Míra statistické jistoty je u těchto proměnných nižší, avšak proměnné jsou stále významné. Velmi důležitě ovlivňují model také proměnné pohlaví, jodurie a přítomnost kadmia ve vodě v místě bydliště.

		Prediktivní komponenta 1			
Proměnné		Nakládací komponenta	t-statistika	$R^2$	
Relevantní prediktory (Matice X)	MUZ	-0.133	-2.54	-0.172	*
	VEK	0.634	13.40	0.868	**
	BMI	0.540	7.20	0.743	**
	JODUR	-0.194	-2.39	-0.261	*
	AS_vod	-0.279	-3.50	-0.304	**
	CD_vod	-0.150	-2.55	-0.121	*
	PL_vod	-0.242	-3.47	-0.286	**
	pr_AS_vod	-0.235	-3.70	-0.324	**
	pr_PB_vod	-0.078	-0.98	-0.125	
	pr_PL_vod	-0.235	-3.60	-0.300	**
	pr_SE_vod	-0.091	-1.16	-0.141	
(Matice Y)	TSH	-0.490	-2.60	-0.373	*
	FT3	-0.418	-1.65	-0.307	
	VOLGUT	0.774	4.72	0.744	**
<b>Vysvětlená variabilita</b>		26.2% (24% po křížové validaci)			
<sup>a</sup> R ... zatížení komponent vyjádřené jako korelační koeficienty s prediktivní složkou, * p < 0,05, ** p < 0,01					

**Tab. 8. Vztah mezi TSH, FT3, Volgut a vnějšími faktory hodnocený modelem O2PLS, prediktivní komponenta 1**

Z Graf 5 první komponenty je názorné, že objem štítné žlázy je větší a zároveň hladiny FT3 a TSH jsou nižší tehdy, má-li pacient větší BMI, je starší, má nižší jodurii, je ženského pohlaví, přijímá méně arsenu z vody, ať už jde o jeho hodnoty v místě bydliště nebo zprůměrované hodnoty s místem školy, má méně kadmia ve vodách v místě bydliště. Vlivy selenu a olova ve zprůměrovaných hodnotách již nedosahují statistické významnosti. Také u proměnné FT3 lze konstatovat, že její vliv v modelu O2PLS již nedosahuje statistické významnosti, i když se jí více blíží ve srovnání s vlivem selenu či olova.



**Graf 5. Komponentní váhy první prediktivní komponenty vyjádřené jako regresní koeficienty s jejich 95 % intervaly spolehlivosti**

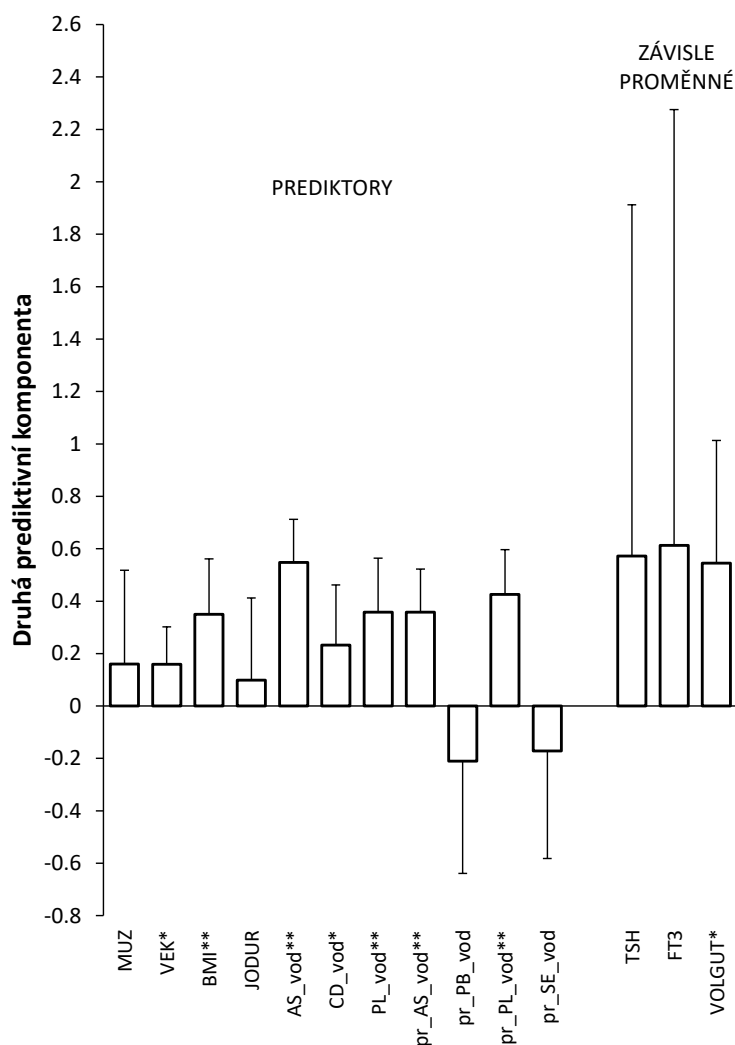
Druhá prediktivní komponenta, zachycená v Tab. 8, vysvětluje pouze 2.4 % variability modelu (1.8 % po křížové validaci). Na rozdíl od první prediktivní komponenty, kdy ke zvyšování objemu štítné žlázy dochází zároveň ke snížení hladin THS a FT3, ve druhé prediktivní komponentě dochází pouze k významnému ovlivnění objemu štítné žlázy nezávisle na hladinách tyreoidálních hormonů, a to opět pozitivně věkem a BMI bez vlivu pohlaví a negativně jodurí. Z parametrů vlivu prostředí vykazuje nejvyšší míru vlivu s objemem štítné žlázy kladně korelující arsen ve vodě (samostatně za místo bydliště i zprůměrovaný s polohou školy) a opět kladně korelující polutanty ve vodě (samostatně za místo bydliště i zprůměrovaný s polohou školy). U všech uvedených ukazatelů je významná míra statistické jistoty. U druhé komponenty je vliv vnějších faktorů na objem štítné žlázy s opačnou polaritou, než je tomu u první prediktivní komponenty. Nicméně tento opačný vliv ukazatelů na objem štítné žlázy je vcelku podstatně slabší vzhledem k zanedbatelné míře variability závisle proměnných

vysvětlení touto komponentou. Také interpretovatelnost této komponenty je v důsledku jejího slabého vlivu na závisle proměnné sporná. Grafické vyjádření druhé prediktivní komponenty je na Graf 6.

		Prediktivní komponenta 2		
Proměnná		Nakládací komponenta	t-statistika	R <sup>2</sup>
Relevantní prediktory (Matice X)	MUZ	0.160	0.85	0.211
	VEK	0.159	2.10	0.203 *
	BMI	0.350	3.14	0.448 **
	JODUR	0.098	0.59	0.131
	AS_vod	0.548	6.29	0.750 **
	CD_vod	0.232	1.91	0.308 *
	PL_vod	0.358	3.30	0.725 **
	pr_AS_vod	0.358	4.11	0.424 **
	pr_PB_vod	-0.211	-0.93	-0.316
	pr_PL_vod	0.426	4.73	0.724 **
pr_SE_vod	-0.171	-0.79	-0.265	
(Matice Y)	TSH	0.573	0.81	0.070
	FT3	0.613	0.70	0.079
	Volgut	0.545	2.20	0.250 *
<b>Vysvětlená variabilita</b>		2.4% (1.8% po křížové validaci)		
<sup>a</sup> R ... zatížení komponent vyjádřené jako korelační koeficienty s prediktivní složkou, * p < 0,05, ** p < 0,01				

**Tab. 9. Vztah mezi TSH, FT3, Volgut a vnějšími faktory hodnocený modelem O2PLS, prediktivní komponenta 2**



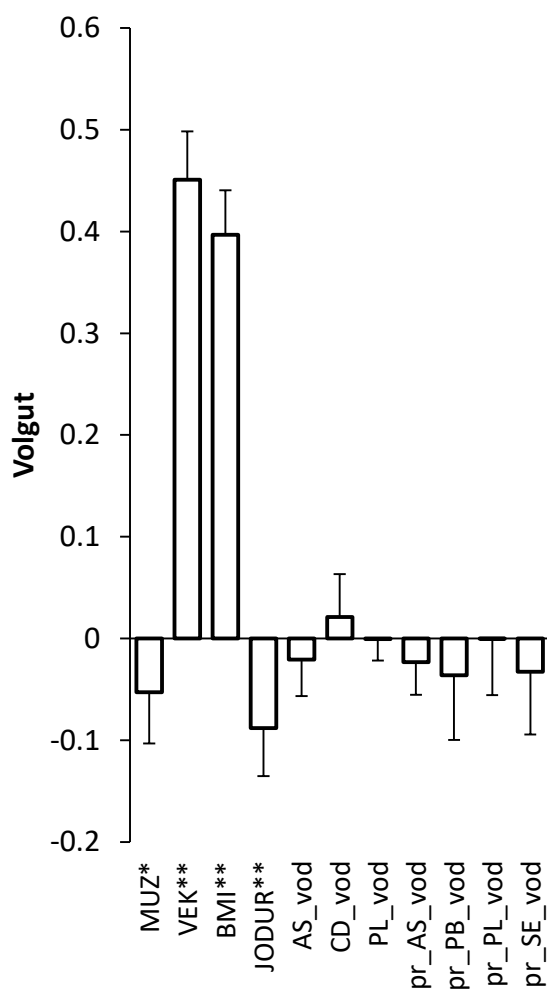


**Graf 6.** Komponentní váhy druhé prediktivní komponenty vyjádřené jako regresní koeficienty s jejich 95 % intervaly spolehlivosti

Následně byly vytvořeny tabulky a grafy vícenásobné regrese pro proměnné TSH, FT3 a Volgut přes obě komponenty. V grafu pro Volgut (Graf 7) je vidět pozitivní korelace věku a BMI, tedy čím větší věk a BMI, tím větší objem štítné žlázy. V Tab. 10 je zachycen vztah Volgut vs. nezávisle proměnné nabývá statistická jistota vyjádřená ve formě t-statistiky u věku hodnoty 18.03 a u BMI hodnoty pak 17.18. Vysoká je i míra vlivu vyjádřená regresním koeficientem. Negativní korelace byla zjištěna u jodurie, tedy čím menší hladina jodurie, tím větší objem štítné žlázy. Její t-statistika je -3.54. Z předchozích grafů a tabulek vyplývá, že také přítomnost arsenu a polutantů ve vodách koreluje s objemem štítné žlázy. Mezi těmito prediktory je však vysoká korelovanost, vedoucí k zanedbatelným regresním koeficientům v modelu vícenásobné regrese. Proměnná Volgut sdílí svou variabilitu z 60.9 % s prvními 4 proměnnými  $X$  (muž, věk, BMI, jodurie). Ostatní proměnné  $X$  vysvětlují zanedbatelnou část variability, přičemž 39.1 % je nevysvětlená variabilita, tedy residua.

Proměnná	Prediktivní komponenta	
	Regresní koeficient	T - statistika
MUZ	-0.053	-1.99 *
VEK	0.451	18.03 **
BMI	0.397	17.18 **
JODUR	-0.088	-3.54 **
AS_vod	-0.021	-1.09
CD_vod	0.021	0.94
PL_vod	0.000	-0.04
pr_AS_vod	-0.023	-1.35
pr_PB_vod	-0.036	-1.07
pr_PL_vod	0.000	0.00
pr_SE_vod	-0.033	-1.00
<b>Vysvětlená variabilita</b>	60.9% (58.9% po křížové validaci)	
* $p < 0.05$ , ** $p < 0.01$		

Tab. 10. Vztah Volgut vs. prediktory hodnocený vícenásobnou regresí

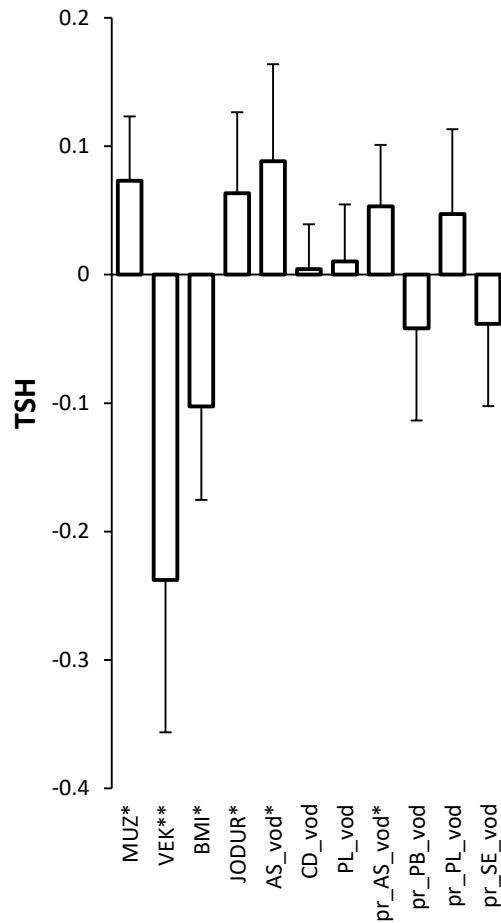


Graf 7. Výsledek vícenásobné regrese pro Volgut

V grafu znázorňujícím výsledky modelu vícenásobné regrese pro závisle proměnnou TSH (Graf 8) je zřejmá negativní korelace s věkem (t-statistika -3.79), a negativní korelace s BMI. TSH zároveň pozitivně koreluje s mužským pohlavím (muži mají vyšší TSH), a s jodurii. Veškeré výsledky jsou zaznamenány v Tab. 11. Pozitivní je také korelace TSH s přítomností arsenu ve vodě (samostatně za místo bydliště i zprůměrovaný s polohou školy). Lze tedy říci, že arsen ve vodě povzbuzuje aktivitu štítné žlázy na úrovni hypofýzy s vyšší produkcí TSH (který pak stimuluje produkci bioaktivního fT3). Prediktory v tomto modelu vysvětlují 14.5 % variability TSH (11.8 % po křížové validaci).

		Prediktivní komponenta	
		Regresní koeficient	T - statistika
Proměnná			
Relevantní prediktory (Matice X)	MUZ	0.073	2.77 *
	VEK	-0.238	-3.79 **
	BMI	-0.103	-2.68 *
	JODUR	0.063	1.91 *
	AS_vod	0.088	2.22 *
	CD_vod	0.004	0.24
	PL_vod	0.010	0.44
	pr_AS_vod	0.053	2.10 *
	pr_PB_vod	-0.042	-1.10
	pr_PL_vod	0.047	1.35
	pr_SE_vod	-0.038	-1.14
	<b>Vysvětlená variabilita</b>		14.5% (11.8% po křížové validaci)
* $p < 0.05$ , ** $p < 0.01$			

**Tab. 11. Vztah TSH vs. prediktory hodnocený vícenásobnou regresí**

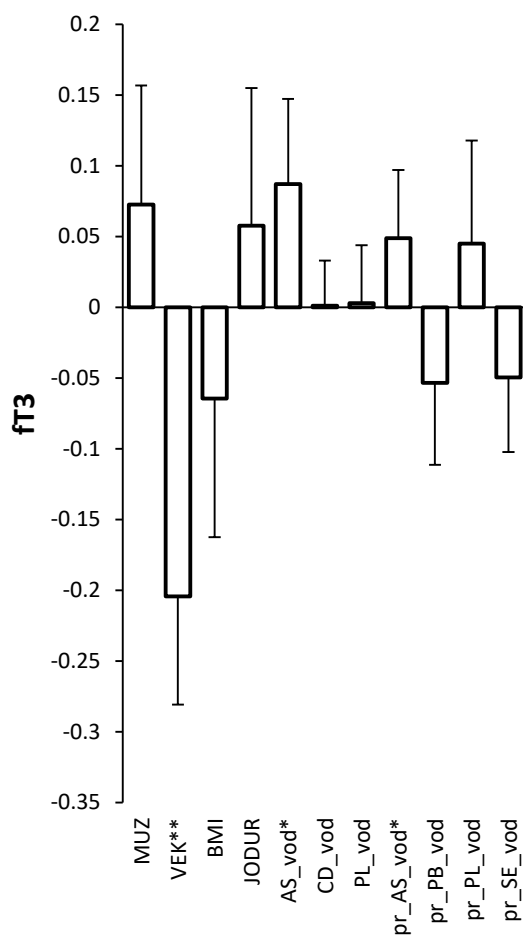


**Graf 8. Výsledek vícenásobné regrese pro TSH**

U grafu, znázorňujícího výsledky vícenásobné regrese pro proměnnou  $fT_3$  (Graf 9), je vidět negativní korelace s věkem. Čím vyšší věk, tím menší hodnota  $fT_3$ . V Tab. 12 pro  $fT_3$  je hodnota t-statistiky pro věk -5.07 a hodnota regresního koeficientu je -0.2. Z tabulky je viditelná míra vlivu arsenu ve vodě u obou zkoumaných hodnot, která je vyšší u hodnoty příslušné místu bydliště pacienta. Jedná se o pozitivní korelaci, tedy čím více arsenu ve vodě, tím vyšší produkce  $fT_3$ . Z tabulky je zřejmé, že hodnoty  $fT_3$  ovlivňují i přítomnost olova a selenu ve vodě, avšak méně výrazně, i když stále významně. Variabilita  $fT_3$  vysvětlená prediktory je 10.1 % (zejména věkem a hodnotami arsenu ve vodě). Po křížové validaci je vysvětlená variabilita 6.6 %.

		Prediktivní komponenta	
		Regresní koeficient	T - statistika
Proměnná			
Relevantní prediktory (Matice X)	MUZ	0.073	1.64
	VEK	-0.204	-5.07 **
	BMI	-0.065	-1.25
	JODUR	0.058	1.12
	AS_vod	0.087	2.75 *
	CD_vod	0.001	0.06
	PL_vod	0.003	0.14
	pr_AS_vod	0.049	1.92 *
	pr_PB_vod	-0.053	-1.75
	pr_PL_vod	0.045	1.17
	pr_SE_vod	-0.050	-1.79
	<b>Vysvětlená variabilita</b>	10.1% (6.6% po křížové validaci)	
	<i>*p&lt;0.05, **p&lt;0.01</i>		

Tab. 12. Vztah ft3 vs. prediktory hodnocený vícenásobnou regresí



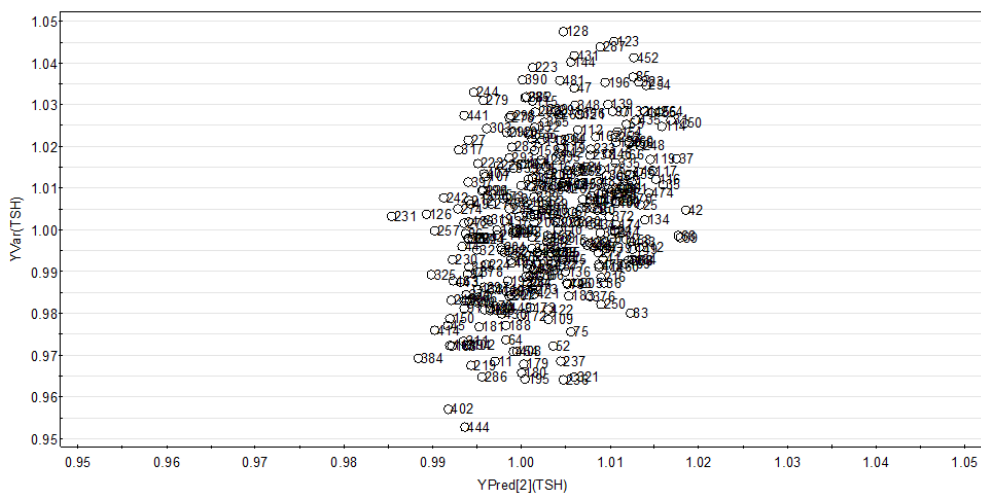
Graf 9. Výsledek vícenásobné regrese pro ft3

Proměnná FT4 (prekurzor) na rozdíl od FT3 (produkt) nebyla v modelu O2PLS relevantní. To je překvapivé, neboť FT4 je konvertováno na FT3 dejodázami. Tuto nesrovnalost by bylo možno vysvětlit metodickými důvody, například nižší přesností měření a tudíž větším podílem nevysvětlené variability.

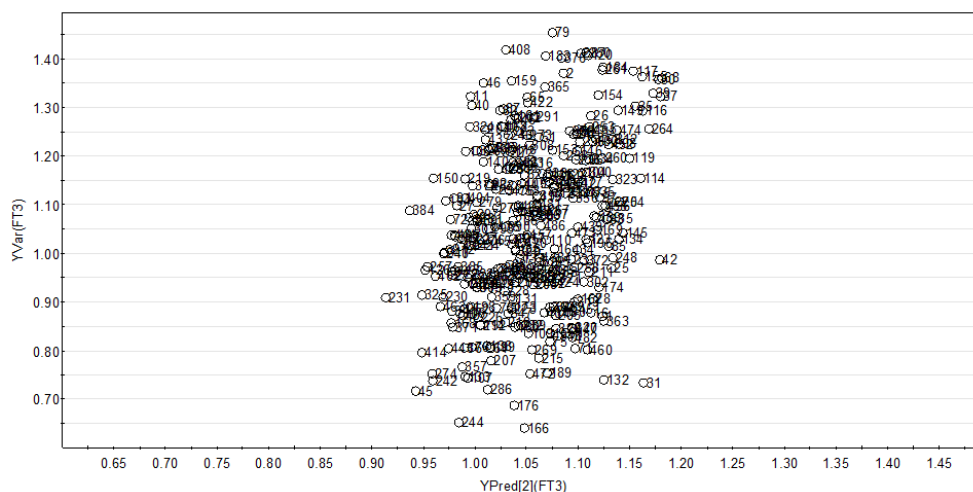
Pro kompletní představu o výsledku byl pro každou závisle proměnnou vytvořen graf shody predikcí a měřených hodnot u dat po mocninné transformaci a normování transformovaných hodnot  $((\text{měřená hodnota} - \text{průměr})/\text{směrodatná odchylka})$  (Graf 10, Graf 11, Graf 12). Nejsilnější korelace je patrná u grafu pro objem štítné žlázy (Graf 10).



**Graf 10.** Graf korelace predikovaných hodnot se skutečností pro Volgut



**Graf 11.** Graf korelace predikovaných hodnot se skutečností pro TSH



**Graf 12. Graf korelace predikovaných hodnot se skutečností pro FT3**

V Tab. 13 jsou uvedeny míry rozptylu predikovaných parametrů vysvětlené modelem O2PLS, a to zvláště pro první a pro druhou prediktivní komponentu. Nejvíce koreluje objem štítné žlázy, a to s oběma komponentami.

Komponenta	Vysvětlený rozptyl TSH (%)	Vysvětlený rozptyl po křížové validaci TSH (%)
1	11.8	11.6
2	14.5	11.8
Komponenta	Vysvětlený rozptyl FT3 (%)	Vysvětlený rozptyl po křížové validaci FT3 (%)
1	7.5	6.6
2	10.1	6.6
Komponenta	Vysvětlený rozptyl Volgut (%)	Vysvětlený rozptyl po křížové validaci Volgut (%)
1	59.2	53.8
2	60.9	58.9

**Tab. 13. Korelace závisle proměnných s komponentami pro všechna měření a po křížové validaci**

## 8.2 Znázornění v prostoru

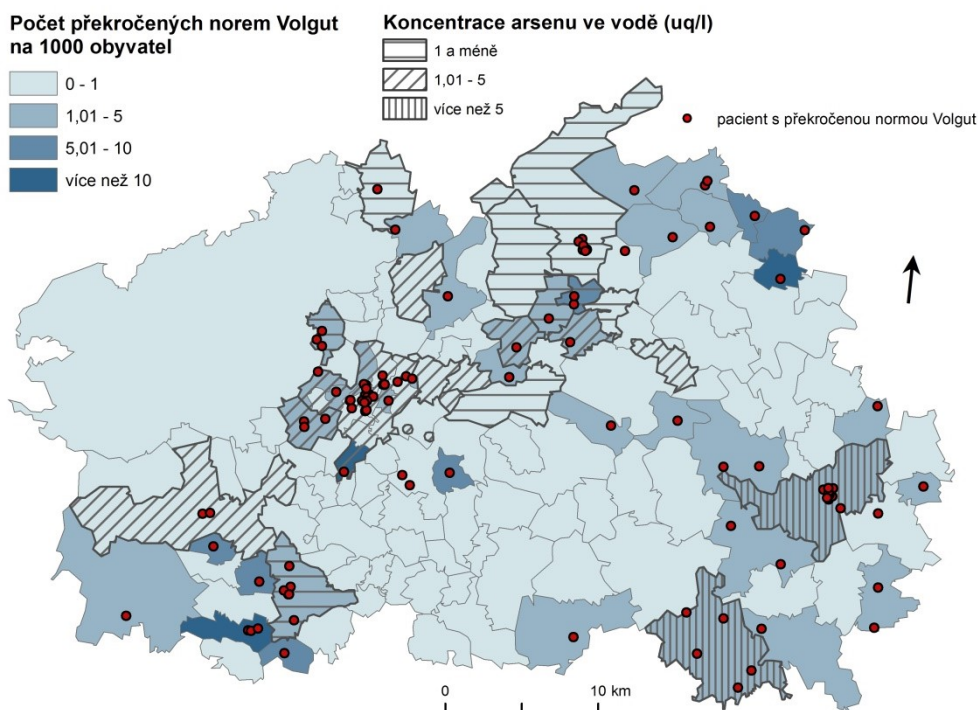
Aby mohla být vytvořena mapa znázorňující rizikové oblasti pro vznik poruch štítné žlázy, bylo by nutné objasnit větší procento ovlivnění poruch jednotlivých funkcí štítné žlázy složkami životního prostředí, tedy získat pro model více dat charakterizující životní prostředí a najít alespoň 10% objasnění. Vliv složek přírodního prostředí byl prokázán zatím velmi malý a nebylo by nijak užitečné tento vliv zobrazovat pomocí mapových výstupů. Ve vytvořeném regresním modelu se podařilo vysvětlit velkou část vzniku poruch jednotlivých parametrů štítné žlázy, ale tento výsledek je dán spíše parametry daného pacienta (věk, BMI).

Kdyby byla nalezena vyšší souvislost mezi funkcemi štítné žlázy a životním prostředím bylo by možné vytvořit mapu rizikovosti. Nejdříve by byla vypočítána regresní rovnice, na základě které by vyšly hodnoty predikce. Tyto hodnoty jsou ale transformované a ještě normované. Nejdříve by se tedy hodnoty musely odnormovat a následně odtransformovat. Poté by bylo možné vytvořit mapu.

Jelikož tedy nebyly výsledky regresního modelu vhodné pro mapové zobrazení, byly vytvořeny mapy, které ukazují jednotlivé parametry štítné žlázy společně se složkami životního prostředí.

### Volgut

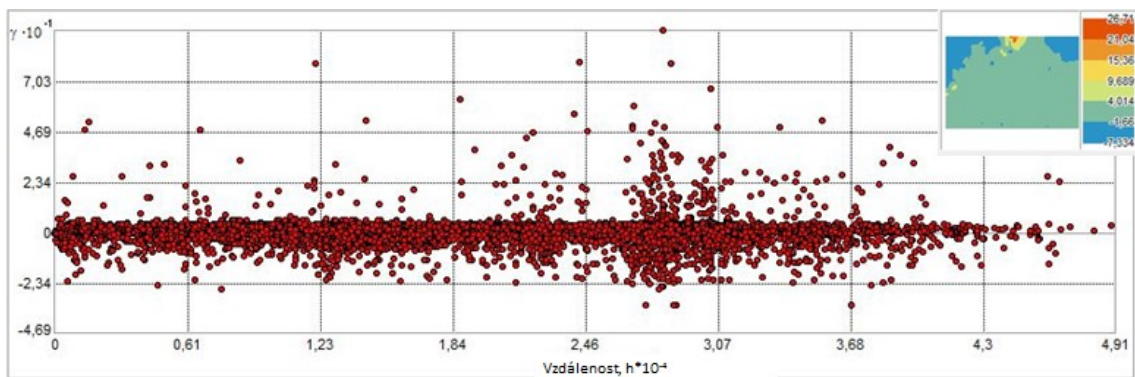
Nejprve byly vytvořeny mapy pro parametr štítné žlázy Volgut. V první komponentě modelu byl prokázán vliv koncentrace As a Pl ve vodě. Na základě toho vznikly dvě mapy (Obr. 25 , Obr. 26 ). První mapa (Obr. 25 ) znázorňuje počet překročených norem Volgut na 1000 obyvatel a koncentraci arsenu ve vodě. Data o koncentraci jednotlivých prvků ve vodě nepokrývají celé území, ale pouze obce s největším počtem obyvatel. U obcí, které mají počet překročených norem více než 10 na 1000 obyvatel, nejsou k dispozici údaje o koncentraci jednotlivých prvků ve vodě.



**Obr. 25** Koncentrace As ve vodě a překročené normy Volgut na 1000 obyvatel

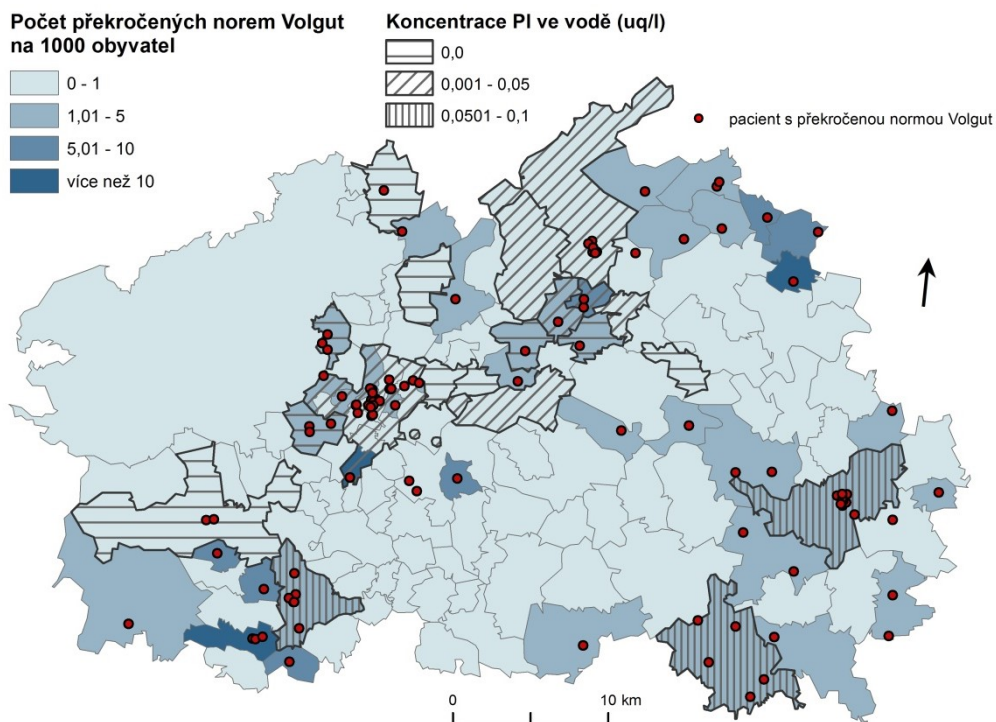


Pro koncentraci As ve vodě a hodnotu Volgut u pacientů byla vytvořena křížová kovariance (Graf 13). Tento graf zobrazuje empirické křížení pro všechny dvojice míst mezi dvěma množinami dat. Křížová kovariance je využívána k prozkoumání místních charakteristik prostorové korelace mezi dvěma množinami dat. Z tohoto grafu je patrná jen velice malá prostorová závislost těchto dvou hodnot. To potvrzuje výsledky metody O2PLS, kde byla nalezena závislost mezi těmito dvěma proměnnými, ale tato závislost je velice malá.



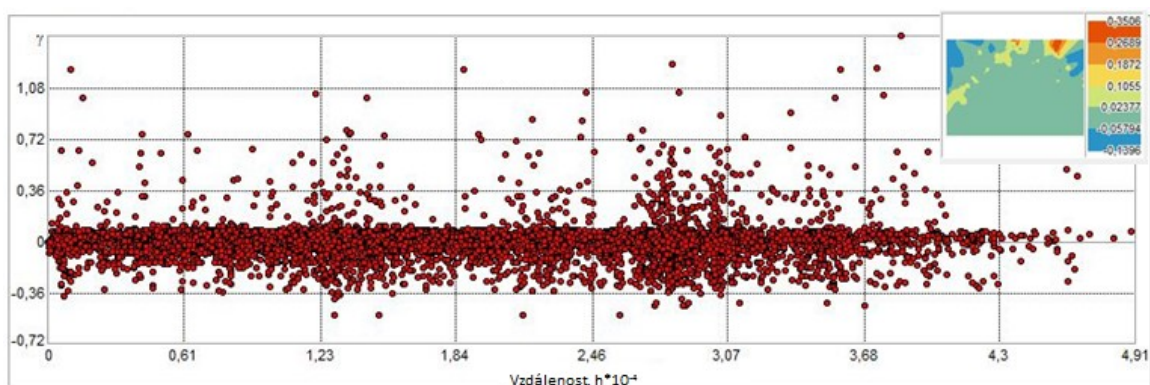
**Graf 13. Křížová kovariance As ve vodě a Volgut**

Druhá mapa (Obr. 26 ) znázorňuje koncentraci polutantů ve vodě a počet překročených norem Volgut na 1000 obyvatel. Koncentrace polutantů je velice malá. Nejvyšší hodnoty se vyskytují v obcích Březnice, Sedlčany a Petrovice. V těchto obcích je počet překročených norem Volgut na 1000 obyvatel v rozmezí od 1.01 do 5. Obcí s tímto rozmezím počtu překročených norem na 1000 obyvatel je v okrese více a je u nich různá koncentrace polutantů ve vodě. Pro obce, kde je nejvyšší počet překročených norem na 1000 obyvatel, nejsou k dispozici naměřené hodnoty o koncentraci polutantů ve vodě.



**Obr. 26** Koncentrace PI ve vodě a překročené normy Volgut na 1000 obyvatel

Pro koncentraci PI ve vodě a hodnotu Volgut u pacientů byla vytvořena křížová kovariance (Graf 14). Z tohoto grafu je patrná jen velice malá prostorová závislost těchto dvou hodnot, stejně jako u Volgut a As ve vodě (Graf 13). Pouze u vzdálenosti okolo 3 jsou vidět možné závislosti.

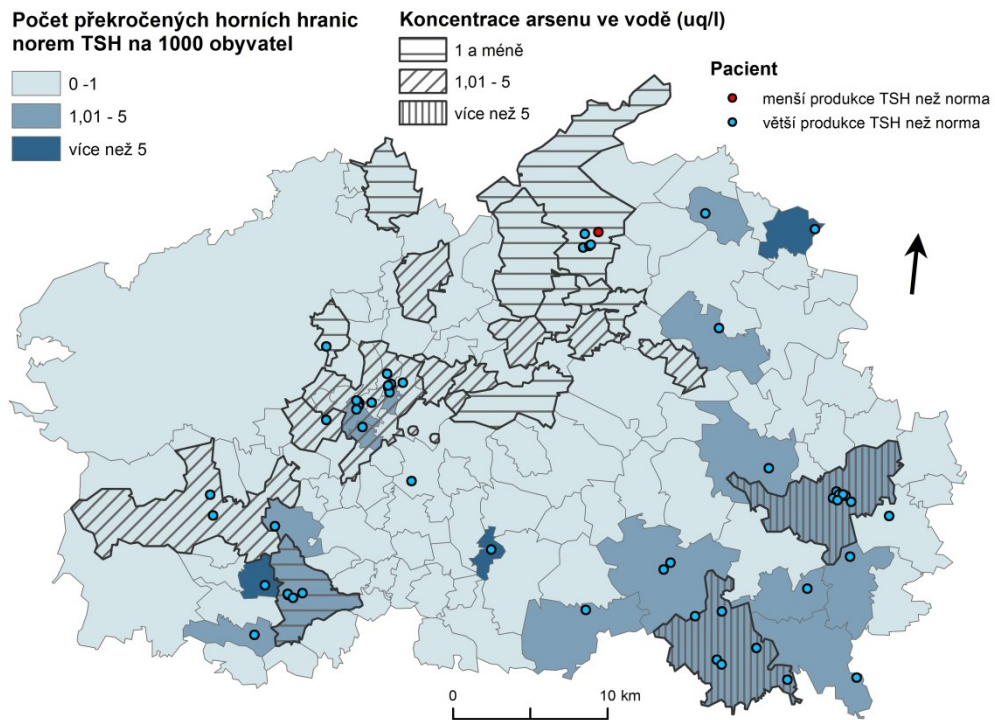


**Graf 14.** Křížová kovariance PI ve vodě a Volgut

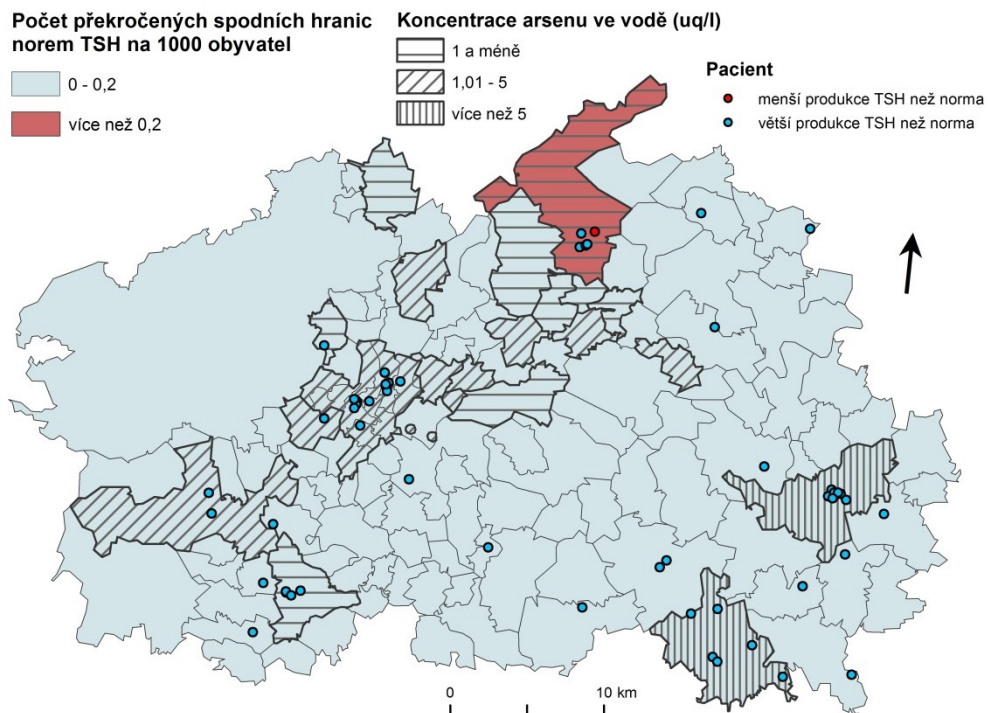
## TSH

Následně byly vytvořeny mapy (Obr. 27 a Obr. 28 ) pro parametr štítné žlázy TSH. V modelu byla prokázána korelace s koncentrací As ve vodě. Byla tedy vytvořena mapa zachycující koncentraci As ve vodě společně s překročenými dolními a horními hladinami norem TSH na 1000 obyvatel. K obcím, které mají počet překročených norem TSH na 1000

obyvatel vyšší než 5, nejsou data o koncentraci arsenu ve vodě k dispozici. V obcích kde je koncentrace arsenu ve vodě v rozmezí 1,01 až 5 uq/l se nevyskytuje zvýšené překročení norem TSH. Ve městě Dobříš je největší počet překročených spodních hranic normy TSH, zde je koncentrace As ve vodě 1 a méně uq/l.

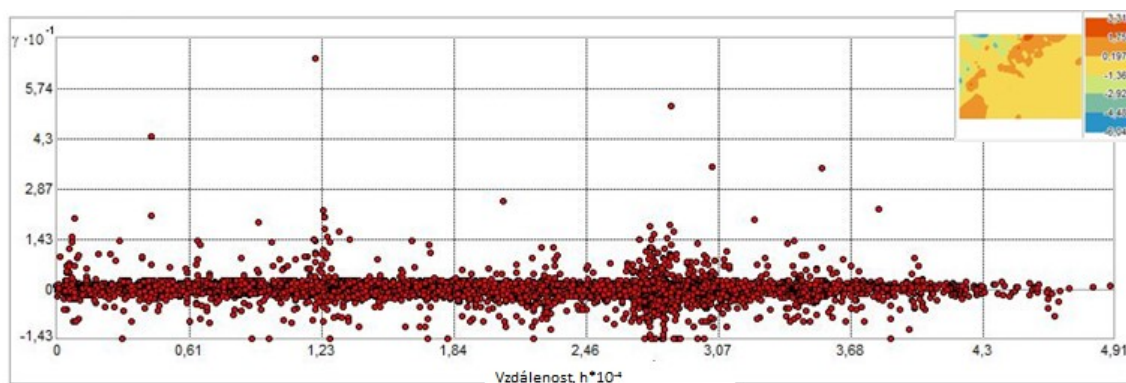


**Obr. 27** Koncentrace As ve vodě a překročené horní hranice normy TSH na 1000 obyvatel



**Obr. 28** Koncentrace As ve vodě a překročené spodní hranice normy TSH na 1000 obyvatel

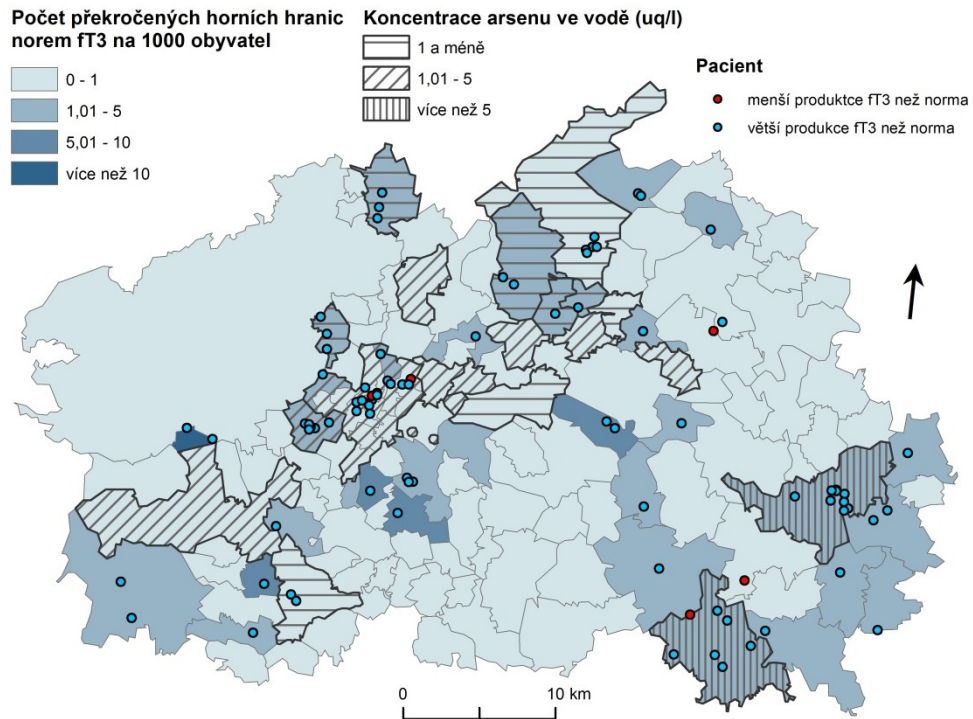
Následně byl vytvořen graf křížové kovariance (Graf 15). Tento graf znázorňuje empirické křížení pro všechny dvojice míst mezi koncentrací As ve vodě a hodnot TSH a vykládá je jako funkci vzdálenosti mezi oběma místy. Z grafu je patrná jen velice malá závislost. Změna nastává u hodnoty vzdálenosti okolo 3. Jedná se o velice podobný graf jako u koncentrace As a Volgut (Graf 13).



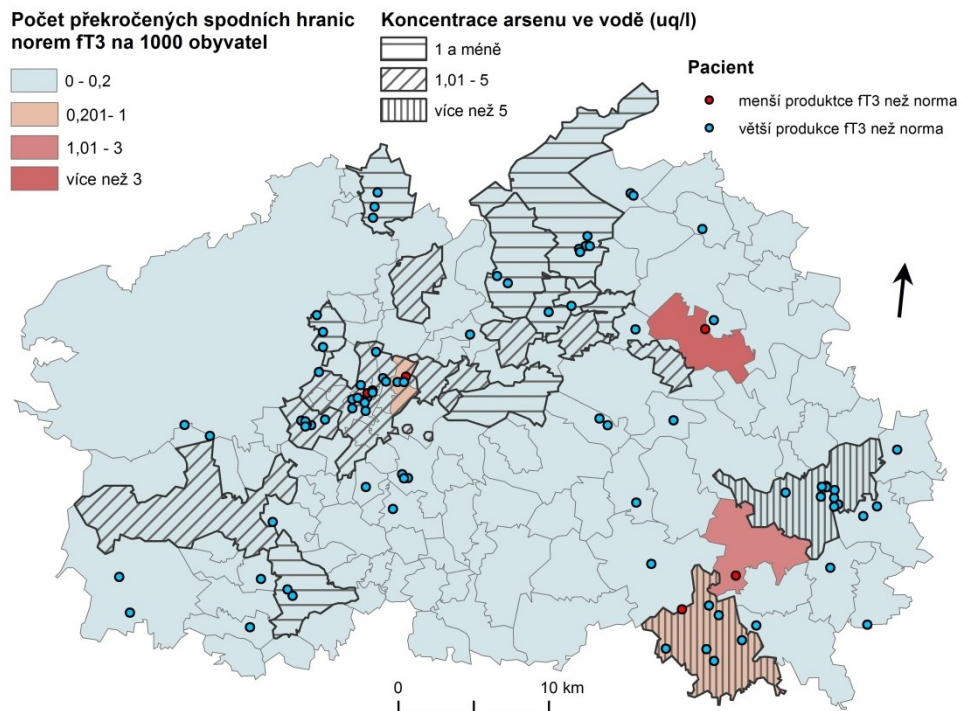
**Graf 15. Křížová kovariance As ve vodě a TSH**

### ft3

Jako poslední proměnná modelu O2PLS byl identifikován parametr ft3. U tohoto parametru měla největší vliv z prostorových proměnných koncentrace As ve vodě. Na základě tohoto byly vytvořeny mapy (Obr. 29 a Obr. 30) zaznamenávající koncentraci As ve vodě a počet překročených spodních a horních hranic norem ft3 na 1000 obyvatel. V obcích, kde je nejvyšší koncentrace As ve vodě, je počet překročených horních hranic norem 1.01 až 5 na 1000 obyvatel. K obcím, kde je překročených více než 10 horních hranic norem na 1000 obyvatel, nejsou k dispozici data o koncentraci As ve vodě. Stejně tak je tomu i u obce Borotice a Vysoký Chlumeč. Tyto obce mají největší počet překročených spodních hranic norem na 1000 obyvatel.

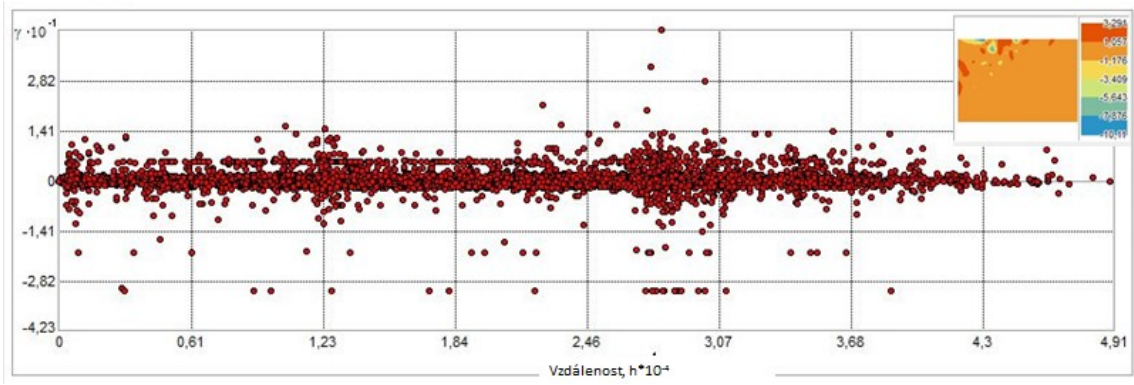


**Obr. 29** Koncentrace As ve vodě a překročené horní hranice normy FT3 na 1000 obyvatel



**Obr. 30** Koncentrace As ve vodě a překročené spodní hranice normy FT3 na 1000 obyvatel

Následně byl vytvořen graf křížové kovariance pro koncentraci As ve vodě a hodnoty FT3 (Graf 16). U všech takto zobrazených hodnot je vždy okolo vzdálenosti 3 vidět možná závislost. Jinak je závislosti vykládána jako velice slabá.



**Graf 16. Křížová kovariance As ve vodě a ft3**

## 9 Diskuse výsledků

V této diplomové práci byl navržen metodický postup pro hodnocení vlivu prostředí na laboratorní a klinické ukazatele funkce štítné žlázy. Bylo prokázáno ovlivnění parametrů štítné žlázy některými prostorovými ukazateli stavu životního prostředí (koncentrace As, Pl, Se, Pb a Cd ve vodě). Model vyloučil velké množství parametrů životního prostředí, které nebyly ve vztahu k ukazatelům funkce štítné žlázy relevantní (koncentrace Ag, Bi, Cd, Cr, Cu, Fe, In, Mo, Pb, Sb, Sn a Zn v mechu, koncentrace SO<sub>2</sub>, SPM, No<sub>x</sub>, Mn, Ni, Pb, As a PM<sub>10</sub> naměřené na stanicích ČHMÚ, koncentrace PM10 M36, SO<sub>2</sub> M4, PM25, As, Ni a Cd v imisích, dopravní koeficient, gama záření, vzdálenost od zlomu, radonový index, koncentrace Bro, Hg, Fe a Cu ve vodě za polohu bydliště i za zprůměrované hodnoty polohy bydliště a školy). V práci bylo využito množství prostorových statistických metod, které předcházely tvorbě výsledného modelu vícenásobné regrese s redukcí dimenzionality a prostorového znázornění.

Prvním z cílů diplomové práce bylo získat vhodná prostorová a zdravotnická data, která pokrývají zájmové území za stejné období. Jako období bylo zvoleno pět let od roku 2000 do roku 2005. Zdravotnická data byla získána od Endokrinologického ústavu v Praze. Bylo nutné jejich ruční zpracování. Data obsahují konkrétní adresu vyšetřených pacientů a naměřené hodnoty parametrů funkce štítné žlázy. Data, charakterizující životní prostředí, byla získána od těchto institucí: Výzkumný ústav Silva Taroucy pro krajinu a okrasné zahradnictví, Český hydrometeorologický ústav, Ředitelství silnic a dálnic, Státní zdravotní ústav, Česká geologická služba a Endokrinologický ústav. Většina získaných dat pokrývá celé zájmové území ve zvoleném období. Bohužel od SZU bylo možné získat záznamy pouze za 10 obcí a k nim 9 parametrů pitné vody. Tato data by bylo vhodnější mít k dispozici za celé území okresu, ale bohužel nebylo možné je k diplomové práci získat.

Druhým cílem diplomové práce bylo získaná data upravit pro vstup do analýz. Tato diplomová práce metodicky navazuje na práci Tomanové (2005) a Grundlocha (2009). V těchto pracích probíhala příprava a zpracování dat v programu ArcGIS Desktop. V obou pracích je program vyhodnocen jako vhodný pro přípravu dat. Tato diplomová práce daný program také využívala a s tvrzením se ztotožňuje. K přípravě dat byl navíc využit program QGIS Desktop 2.12.3, který v některých postupech zobrazení dat nabízí snadnější řešení. Vhodné je však kombinovat oba programy. U přípravy dat, zachycující parametry životního prostředí, bylo nutné využít interpolační metody. Zvolena byla metoda IDW, jelikož při aplikaci na vrstvu mechu se jevila jako nejvhodnější. Byla aplikována i na vrstvy dat z měřících stanic ČHMÚ. Takto upravená data byla podpořena i dalšími daty pro případ pozitivních výsledků v regresním

modelu. Pokud by byla data používána samostatně, bylo by vhodnější vyhledat metodu, která dokáže zohlednit i pohyby vzduchových hmot a další aspekty šíření imisí.

V práci Tomanové (2005) a Grundlocha (2009) byl dále program ArcGIS Desktop využíván pouze pro tvorbu map. Tato možnost programu byla využita k naplnění dalšího cíle práce: prostorová data popsat ve vztahu ke zdravotním aspektům funkce štítné žlázy, zdravotnická data statisticky vyhodnotit a popsat jejich výskyt vzhledem k poloze v okrese. Pomocí programu byli graficky znázorněni pacienti, rozdělení podle počtu překročených norem, a dále školy, u kterých byl v % vyjádřen počet překročených norem na žáka. Postupně byly pod tyto dvě vrstvy podkládány vrstvy parametrů životního prostředí. Z vizuálního porovnání byly identifikovány parametry vod s obsahem As, Pb, Se, poté proměnné imisních koncentrací SO<sub>2</sub>, NO<sub>2</sub>, As a PM10, a několikrát byla identifikována koncentrace Cu v mechu. V této části byly statisticky vyčísleny počty překročených norem jak souhrnně za pacienty, tak za školy. Za rizikové bylo označeno těchto 15 škol: ZŠ Nechvalice, ZŠ Petrovice u Sedlčan, ZŠ Nová Ves pod Pleší, ZŠ Rosovice, ZŠ Sedlčany, ZŠ Krásná Hora nad Vltavou, ZŠ Dublovice, ZŠ Obecnice, ZŠ Bohutín, ZŠ Malá Hraštice, ZŠ Nový Knín, ZŠ Příbram - Březové Hory, 2. ZŠ Sedlčany, ZŠ Něčín a ZŠ Jiráskovy sady. Jmenované školy se nachází převážně v severovýchodní části území okresu, a proto zde byl předpokládán zvýšený počet pacientů se zhoršenou funkcí štítné žlázy.

Dalším cílem práce bylo popsat distribuci zdravotnických dat v prostoru a odhalit pravidelnosti v prostorovém uspořádání zdravotních dat. Tato část práce byla vypracována opět pomocí programu ArcGIS Desktop. Vhodnými nástroji k mapování a popisu výskytu, konkrétně v této práci pacientů se zhoršenou funkcí štítné žlázy, jsou funkce z Spatial Statistics Tools z ArcToolboxu. Metody, které tyto nástroje nabízí, jsou velmi vhodné pro vytvoření si představy o výskytu daného jevu. V této práci byl pomocí nástrojů zobrazen geografický střed, který byl vytvořen i ve variantě váženého geografického středu, kdy vahami byl počet překročení norem u pacientů. Výsledek poukazoval na střed území okresu Příbram, ve kterém se ale nacházelo minimum pacientů. V podobné vzdálenosti od geografického středu leží tři města s největším počtem obyvatel (Příbram, Dobříš, Sedlčany). To signalizovalo, že právě zde se nachází nejvíce případů onemocnění pacientů. Zároveň byl vyvrácen předpoklad, že nejvíce pacientů je v severovýchodní části okresu. Následně se dalším využitím nástrojů z Spatial Statistics Tools prokázala užitečnost využití ArcGIS Desktop při analýze prostorových vzorů. Mapa, znázorňující počet případů na km<sup>2</sup>, byla vytvořena pomocí funkce Kerndel Density. Tato funkce ukázala nejvyšší koncentrace v okolí největších měst mimo města Rožmitál pod Třemšínem. Následně byly vypočítány metody nejbližšího souseda a Ripleyho K – funkce. Výsledky obou metod znázornily shlukové rozložení pacientů v okrese Příbram. Tyto závěry



společně s předchozími výsledky metod naznačují, že pacienti se zhoršenou funkcí štítné žlázy kopírují rozmístění obyvatelstva v okrese. Mapa koncentrace navíc neznázorňuje žádnou vyšší koncentraci mimo města, naopak ukazuje menší hustotu ve městě Rožmitál pod Třemšínem. Z tohoto důvodu byly vytvořeny kartogramy, které znázorňují počty překročených norem (celkově i u jednotlivých parametrů funkce štítné žlázy) na 1000 obyvatel pro jednotlivé obce. Zde byla vyloučena koncentrace pacientů se zhoršenou funkcí štítné žlázy ve městech. Naopak se projeví menší vesnice, které jsou převážně v jihovýchodní části okresu. V závěru byla aplikována metoda lokální Moranovo I. U parametrů štítné žlázy TSH, fT3 a fT4 byly identifikovány oblasti se sníženou hodnotou vždy v severní části a ve středu okresu. Zvýšené hodnoty byly identifikovány na spojnici měst Příbram- Dobříš a v jejím prodloužení.

Poslední cílem práce bylo navrhnout regresní model, který dává do souvislosti aspekty životního prostředí a charakteristiky funkce štítné žlázy. Tento regresní model byl navržen metodou obousměrných ortogonálních projekcí do latentní struktury (O2PLS) zahrnující hodnocení relevance prediktivních komponent metodou parciálních reziduálních součtů čtverců (Partial Residual Sum of Squares, PRESS) a dále pak relevance prediktorů hodnocených *VIP* statistikou. Do postupu byla rovněž zahrnuta křížová validace výsledků. Výsledkem uvedeného postupu je vznik vícerozměrného regresního modelu, který popisuje vztah parametrů funkce štítné žlázy Volgut, TSH a fT3 na jedné straně a prediktorů, zahrnující relevantní parametry kvality životního prostředí, na straně druhé (parametry kontaminantů ve vodách - As, Cd, Pl, Pb, Se). U ostatních parametrů životního prostředí nebyl prokázán vliv na proměnné definující funkci štítné žlázy. Dále byl z pohledu funkce štítné žlázy, podle očekávání, potvrzen vliv BMI, věku, pohlaví a jodurie. I přesto, že tento vliv je v oblasti medicíny známý, je nutno tyto proměnné do modelu zahrnout. Kdyby byly vynechány, vzrostl by podíl nevysvětlené variability a snížila by se významnost dalších relevantních prediktorů. Tímto krokem byl splněn poslední cíl diplomové práce, a to potvrdit nebo vyloučit vliv jednotlivých proměnných životního prostředí na funkci štítné žlázy.

V práci Tomanové (2005) byly korelovány hodnoty jodurie s proměnnými prostředí. Jodurie je základním parametrem, který vypovídá o obsahu jódu v moči (Velký lékařský slovník 2017), ale nejedná se o parametr, který definuje funkci štítné žlázy. Výsledky práce Tomanové (2005), které naznačují vliv gama záření a dalších geologických dat, se v této diplomové práci nepotvrdily. V práci nebyla použita stejná data a Tomanová (2005) měla k dispozici větší množství geologických charakteristik. Pro kompletní stanovení závěrů této diplomové práce by bylo vhodné rozšířit datovou základnu, např. o koncentrace thoria, draslíku nebo uranu v prostředí. V práci Grundlocha (2009) byly při využití Spearmanova korelačního koeficientu

nalezeny korelace objemu štítné žlázy s radonovým indexem, SO<sub>2</sub>, NOX, NO, Cs v mechu, intenzitě dopravy a PI ve vodě, a TSH s As, Ba, Be v mechu, PI ve vodě a intenzitě dopravy. Výsledky předkládané diplomové práce se shodují s prací Grundlocha (2009) pouze v prokázání ovlivnění PI ve vodě hormonu TSH a v menší míře Volgut. Ostatní ovlivnění proměnnými životního prostředí bylo vyloučeno. Funkci štítné žlázy není možné posuzovat pouze na základě jedné vlastnosti, stejně jako její ovlivnění nelze posuzovat pouze na základě jednoho parametru životního prostředí. Z tohoto důvodu je vhodnější volit takové metody, které dokáží zohlednit více aspektů jak lidského organismu, tak možných faktorů, které ho ovlivňují.

V práci byl sestaven regresní model, který popisuje vliv různých faktorů na Volgut, TSH a fT3. Z proměnných životního prostředí se jednalo o koncentrace As, Cd, PI, Pb, Se ve vodě. Nejvýznamnější zde byl pozitivní vliv **As** a **PI**. Kladný vliv arsenu je využíván při léčbě leukémie. V současné době se uvažuje o využití při léčbě dalších typů rakovin (Conger Krista 2010). Důležitým aspektem je velikost dávky při léčbě. Jelikož v pitné vodě nebyly nalezeny koncentrace převyšující povolenou hranici, ale pouze menší hodnoty, je tedy možné pozitivní ovlivnění menšího množství arsenu na produkci hormonu fT3 u štítné žlázy.

Cílem práce bylo upravit a rozšířit dosavadní přístupy ke statistické analýze prostorových dat v souvislosti s daty medicínálními a navrhnout vhodný metodický postup. Předchozí metodický postup byl rozšířen o množství prostorových metod, které na sebe logicky navazují a tvoří ucelený postup. Z předchozích postupů byl ponechán přístup ke zpracování dat, který byl ale také upraven. Další části práce již vycházely z obecných postupů prostorové statistiky. Přesto, že se v práci podařilo prokázat vliv některých parametrů životního prostředí, prokázané parametry neposkytovaly dostatečnou datovou základnu, aby mohla být vytvořena predikční mapa, zobrazující oblasti ovlivňující funkci štítné žlázy. Tento vliv byl velice malý a nebylo by nijak užitečné vytvářet prediktivní mapu. Od ministerstva životního prostředí bylo možné získat pouze informace za 10 obcí v okrese, v budoucnosti by bylo vhodnější mít data k dispozici za všechny obce. Zde je velký prostor pro posun vpřed. Pro tvorbu takovéto mapy by bylo nutné rozšířit datovou základnu jak z pohledu informací o dané osobě, jako např. přítomnost bisfenolu nebo ftalátů v bydlíšti pacienta, tak z pohledu parametrů životního prostředí s tím, že by bylo vhodné zaměřit se na koncentrace arsenu v geologickém podloží nebo v ovzduší, dále na pesticidy u půd a na další umělá znečišťovadla.

Navržený metodický postup by mohl být aplikován i na jiné typy zdravotních dat, vhodných pro ověření vlivu prostředí na danou nemoc. Pro tuto aplikaci by ale bylo nutné získat zdravotnická data na úrovni adres pacientů. Dané nemoci by měl být přizpůsoben i výběr parametrů životního prostředí.

## 10 Závěr

Tato diplomová práce je přínosem primárně v oblasti prostorové analýzy dat a dále také v endokrinologickém oboru. Díky této práci vznikl postup, jak zpracovávat citlivá zdravotnická data a vyvozovat z nich informace o jejich souvislosti s daty prostorovými. V práci byl vytvořen nový metodický postup analýzy prostorových a medicínálních dat. Ve srovnání s předchozími pracemi tato práce využívá větší množství prostorových statistických metod, které tvoří logickou návaznost jednotlivých částí postupu. V práci je sestaven aplikovatelný metodický postup při řešení úkolu, týkajícího se endokrinologických dat a parametrů životního prostředí.

Tento postup zahrnuje detailní popis přípravy obou typů dat. Příprava dat byla provedena v programu ArcGIS Desktop, Qgis Desktop 2.12.3 a Microsoft Excel 2010. Při přípravě dat bylo nutné využít interpolační metody. V poslední části přípravy dat byla ověřena normalita dat, aby data mohla být využita ve statistických analýzách, kde je normalita dat podmínkou. Jelikož většina dat neměla normální rozložení, byla provedena mocinná transformace, aby toto rozložení bylo získáno.

V kapitole 6 byla mapována a popisována zhoršená funkce štítné žlázy. Tato část byla vypracována v programu ArcGIS Desktop. Zde bylo navzájem vizuálně porovnáno velké množství parametrů životního prostředí a parametrů funkce štítné žlázy. Statistický popis zhoršené funkce štítné žlázy byl vytvořen v programu SPSS a Microsoft Excel 2010.

Třetí část postupu (kapitola 7) analyzovala prostorové vzory výskytu zhoršené funkce štítné žlázy. Nejprve byl určen geografický střed vážený i nevážený. Následně byla vypočítána směrodatná vzdálenost a směrodatná elipsa. Poté byla vypracována konfirmační analýza dat, která pomocí nástroje Kernel density znázornila koncentraci pacientů na km<sup>2</sup>. Následně byla vypracována Metoda nejbližšího souseda a Ripleyho K – funkce. Na základě těchto metod bylo určeno shlukové rozmístění pacientů v okrese Příbram. V poslední kroku této části byla aplikována metoda lokálního Moranova I, která určila statisticky významná seskupení prvků s nízkými nebo vysokými hodnotami a prostorové odchylky.

V poslední části metodického postupu (kapitola 8) byl za pomoci metody O2PLS vytvořen regresní model. Tento model ze 60.9 % vysvětlil vliv na objem štítné žlázy, z 10.1 % vliv na fT3 a ze 14.5 % vliv na TSH. V regresním modelu byl zaznamenán vliv parametrů životního prostředí, a to koncentrací As, Pl, Pb, Cd a Se ve vodě. Vliv ostatních proměnných životního prostředí, které byly do modelu zahrnuty, byl vyloučen. I přesto, že se v práci nepodařilo prokázat vliv více parametrů životního prostředí, podařilo se dosáhnout významného posunu v této problematice.

Takto navržená metodika by za předpokladu stejně prostorově konkrétních dat mohla být aplikována i na jiné nemoci s odlišnými parametry životního prostředí. V práci bylo použito množství metod prostorové statistiky a byla vytvořena metodika, která přinesla nové poznatky v endokrinologii, které mohou být v této oblasti využity.

## Zdroje

Agendový portál. *Endokrinologický ústav* [online]. 2016 [cit. 2016-06-22]. Dostupné z: [http://ap.mzcr.cz/endokrinologicky-ustav/informace\\_37a.html](http://ap.mzcr.cz/endokrinologicky-ustav/informace_37a.html).

ArcGIS Pro. *Spatial Statistics toolbox* [online]. 2017 [cit. 2017-04-10]. Dostupné z: <http://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/an-overview-of-the-spatial-statistics-toolbox.htm>.

ASHIZAWA, Kiyoto a kol. Prevalence of Goiter and Urinary Iodine Excretion Levels in Children Around Chernobyl 1. *The Journal of Clinical Endocrinology & Metabolism*. 1997, roč. 82, č. 10, s. 3430-3433.

BARANOVIČOVÁ, Zuzana. *Scintigrafie štítné žlázy*. Brno, 2015. Bakalářská práce, LF MU. Vedoucí práce Igor Černý.

BEDÁŇOVÁ, Iveta. *Biostatistika* [online]. 2017 [cit. 2017-04-24]. Dostupné z: <http://cit.vfu.cz/stat/index.htm>.

BENCKO, Vladimír a kol. *Statistické metody v epidemiologii*. Praha: Karolinum, 2003. ISBN 80-246-0763-8.

BIVAND, Roger ; PEBESMA, Edzer; GOMEZ-RUBIO, Virgilio. *Spatial Data Import and Export. Applied Spatial Data Analysis with R*. Springer New York. 2013. s. 83-125.

CLARKE, Bertrand; FOKOUE, Ernest; ZHANG, Hao Helen. *Principles and theory for data mining and machine learning*. Springer Science & Business Media, 2009.

CONGER, Krista. *Arsenic shows promise as cancer treatment, study finds* [online]. 2010 [cit. 2017-04-11]. Dostupné z: <https://med.stanford.edu/news/all-news/2010/07/arsenic-shows-promise-as-cancer-treatment-study-finds.html>.

ČEŘOVSKÁ, Jarmila. *Monitorování zdravotního stavu obyvatel vybraných regionů České republiky se zaměřením na tyreopatie a jejich etiologická agents*. Praha: Iga MZ ČR. Závěrečná zpráva o řešení grantu Interní grantové agentury MZ ČR. Dostupné také z: <http://kramerius.medvik.cz/search/handle/uuid:MED00159882>.

Česká geologická služba. *Státní geologická služba* [online]. 2016 [cit. 2016-06-22]. Dostupné z: <http://www.geology.cz/extranet/sgs>.

JMP. *Variable importance plot* [online]. [cit. 2017-03-04]. Dostupné z: [http://www.jmp.com/support/help/13/Variable\\_Importance\\_Plot.shtml](http://www.jmp.com/support/help/13/Variable_Importance_Plot.shtml).

Český statistický úřad. *Charakteristiky okresu Příbram* [online]. 2016 [cit. 2016-10-29]. Dostupné z: [https://www.czso.cz/csu/xs/charakteristika\\_okresu\\_pribram](https://www.czso.cz/csu/xs/charakteristika_okresu_pribram).

DRÁBOVÁ, Dana a kol. *Patnáct let od havárie Černobylu: Důsledky a poučení*. Praha, Státní úřad pro jadernou bezpečnost. 2001.

ELLIOTT, Paul a kol. *Risk of adverse birth outcomes in populations living near landfill sites*. *Bmj*. 2001, roč. 323, č. 7309, s. 363-368. ISSN 0959-8138.

EL BOUHADDANI, Said a kol. Evaluation of O2PLS in Omics data integration. In: *BMC bioinformatics*. 2016, roč. 17, č. 2, s. 11.

*Endokrinní systém* [online]. 2016 [cit. 2016-10-29]. Dostupné z: <http://www.endokrinni-system.cz/home>.

ERIKSSON, Lennart; WOLD, Svante; TRYGG, Johan. *O2PLS® for improved analysis and visualization of complex data*.

ESRI. *What is GIS?* [online]. 2016 [cit. 2016-11-9]. Dostupné z: <http://www.esri.com/what-is-gis>.

FOTHERINGHAM, Stewart; ROGERSON Peter. *Spatial analysis and GIS*. Taylor & Francis, London, 1997.

FSV ČVUT. *Regionálně geologické minimum* [online]. 2004 [cit. 2016-11-13]. Dostupné z: [http://departments.fsv.cvut.cz/k135/w\\_wwold/webkurzy/rg/regionalka.html](http://departments.fsv.cvut.cz/k135/w_wwold/webkurzy/rg/regionalka.html).

GEIZEROVÁ, Helena a kol. *Epidemiologie vybrané kapitoly pro seminární a praktická cvičení*. Praha: Karolinum, 1995. ISBN 80-7184-179-X.

Geologický park. *Bohemikum* [online]. 2016 [cit. 2016-11-13]. Dostupné z: <http://www.parkgeo.cz/bohemikum.htm>.

GRUNDLOCH, Jiří. *Geologická a environmentální interpretace dat tyreologické epidemiologie pomocí geoinformačních systémů: srovnávací lokality Krucemburk, Budišov, Radostín nad Oslavou a Dolní Rožínka*. Praha, 2009. Diplomová práce, Ústav geochemie, mineralogie a nerostných zdrojů, PŘF UK. Vedoucí práce Karel Martínek.

HENKE, Julia; PETROPOULOS, George. *A GIS-based exploration of the relationships between human health, social deprivation and ecosystem services: The case of Wales, UK*. *Applied Geography*. 2013, roč. 45, s. 77-88.

HEŘMANOVÁ, Eva. *Současné trendy v aplikacích statistických a kvantitativních metod v kontextu prostorových analýz*. Mala Ida: Podnikovohospodárska fakulta Košice Ekonomické university Bratislava. 2010, s. 23–30. ISBN 978-80-225-3016-3.

HILL, Cliff a kol. *Iodine-131 in human thyroids in Britain following Chernobyl*. Nature. 1986, roč. 321, č. 6071, s. 655-656.

HOLČÍK, Jiří a kol. *Matematická biologie: e-learningová učebnice* [online]. 1. vydání. Brno: Masarykova univerzita, 2015. ISBN 978-80-210-8095-9.

HORÁK, Jiří a kol. *Prostorové analýza nezaměstnanosti* [online]. VŠB-TU Ostrava. 2000. Dostupné z: [http://gisak.vsb.cz/GACR\\_PAN/Brozura/PAN.html](http://gisak.vsb.cz/GACR_PAN/Brozura/PAN.html).

HORÁK, Jiří. *Prostorové analýzy dat* [online]. Institut geoinformatiky, Hornickogeologická fakulta, VŠB-TU Ostrava 2002. Dostupné z: <http://gisak.vsb.cz/pad/index.htm>.

CHLUPÁČ, Ivo. *Geologická minulost České republiky*. Praha: Academia, 2011. Neživá příroda. ISBN 978-80-200-1961-5.

IHE. *Zpráva o radiační situaci na území ČSSR po havárii jaderné elektrárny Černobyl*. 1987. B.m.: Institutu hygieny a epidemiologie.

KACHLÍK, Václav. *Geologický vývoj území České republiky*. Ústav geologie a paleontologie, PŘF UK a Správa úložišť radioaktivních odpadů (SÚRAO), Praha, 2003, s. 1-64.

KNECHTSBERGER, Peter. *Radiace, nemoc a traumata z ozáření*. Radiace [online] [cit. 23. 3. 2016]. Dostupné z: [http://www.darius.cz/arceus/NT\\_radia.html](http://www.darius.cz/arceus/NT_radia.html).

KOLÁŘ, Jan. *Geografické informační systémy 10*. Vydavatelství ČVUT, 2003. ISBN 80-01-02687-6.

KURTTIO, Päivi a kol. *Fallout from the Chernobyl accident and overall cancer incidence in Finland*. Cancer epidemiology. 2013, č. 37, s. 585-592.

*Lidské tělo: srozumitelný a zevrubný průvodce po strukturách a funkcích lidského organismu*. Praha: Cesty, 1996. ISBN 80-7181-094-0.

LÖFSTEDT, Tommy. *OnPLS: Orthogonal projections to latent structures in multiblock and path model data analysis*. 2012. Disertační práce. Umeå universitet.

LSHTM. *About John Snow*. John Snow Bicentenary [online] 2016 [cit.2016-03-23]. Dostupné z: <http://johnsnowbicentenary.lshtm.ac.uk/about-john-snow/>.

MADSEN, Rasmus; LUNDSTEDT, Torbjörn; TRYGG, Johan. Chemometrics in metabolomics—a review in human disease diagnosis. Analytica chimica acta. 2010, č. 659, s. 23-33.

MALÁ, Barbora. *Klasifikace komponent získaných z dat funkční magnetické rezonance pomocí analýzy nezávislých komponent*. Brno, 2013. Diplomová práce, Přírodovědecká fakulta Masarykova univerzita. Vedoucí práce Daniel Schwarz.

MATOLÍN, Milan. *Terénní gamaspektrometrické stanovení přírodních radionuklidů, izotopů Cesia a dávkového příkonu záření gama v Příbrami a okolí*. Praha, 2000. Výzkumná zpráva oddělení užití geofyziky PŘF UK.

MELOUN, Milan a kol. Transformation in the PC-aided biochemical data analysis. *Clinical chemistry and laboratory medicine*. 2000, č. 38, s. 553-559.

MELOUN, Milan; MILITKÝ, Jiří; HILL, Martin. *Počítačová analýza vícerozměrných dat* v příkladech. Praha: Academia, 2005. ISBN 80-200-1335-0.

NETRDOVÁ, Pavlína. *Prostorová analýza dat* [online]. [cit. 2017-03-29]. Dostupné z: [https://dl2.cuni.cz/pluginfile.php/12672/mod\\_resource/content/4/Prednaska\\_SADII\\_2016\\_Pr\\_ostorova\\_analyza.pdf](https://dl2.cuni.cz/pluginfile.php/12672/mod_resource/content/4/Prednaska_SADII_2016_Pr_ostorova_analyza.pdf).

PETERKA, Miroslav; PETERKOVÁ, Renata; LIKOVSKÝ, Zbyněk. *Chernobyl: prenatal loss of four hundred male fetuses in the Czech Republic*. *Reproductive Toxicology*. 2004, roč. 18, č. 1, s. 75-79.

Portál ČHMÚ. *Historie ústavu*[online]. 2016 [cit. 2016-06-22]. Dostupné z: <http://portal.chmi.cz/o-nas/historie-ustavu>.

ŘSD ČR. *Ředitelství silnic a dálnic*[online]. 2016 [cit. 2016-06-22]. Dostupné z: <https://www.rsd.cz/wps/portal/web/rsd/Reditelstvi-silnic-a-dalnic>.

SHAHID, Syed Umair; IQBAL, Javed; HASNAIN, Ghalib. *Groundwater quality assessment and its correlation with gastroenteritis using GIS: a case study of Rawal Town, Rawalpindi, Pakistan*. *Environmental monitoring and assessment*. 2014, roč. 186, č. 11, s. 7525-7537.

SPURNÁ, Pavlína. *Současné trendy v kvantitativní analýze geografických dat se zaměřením na využití metody geograficky vážené regrese*. Praha, 2006. Diplomová práce. Katedra sociální geografie a regionálního rozvoje PŘF UK.

SPURNÁ, Pavlína. *Geograficky vážená regrese: metoda analýzy prostorové nestacionarity geografických jevů*. *Geografie – Sborník ČGS*. 2008 roč. 113, č. 2, s. 125 – 139.

SPURNÁ, Pavlína. *Prostorová autokorelace – všudypřítomný jev při analýze prostorových dat*. *Sociologický časopis*. 2008, roč. 44, č. 4, s. 767 – 787.

Státní zdravotní ústav. *Poslání ústavu*[online]. 2016 [cit. 2016-06-22]. Dostupné z: <http://www.szu.cz/poslani-ustavu>.

ŠTYCH, Přemysl a kol. *Vybrané funkce geoinformačních systémů*. Praha: Česká kosmická kancelář, 2008.



TOLLINGEROVÁ, Dana. *Geografické informační systémy*. Ostrava: VŠB-Technická univerzita. 1996. ISBN 80-7078-377-X.

TOMANOVÁ, Renáta. *Vliv geologie a radioaktivního záření na výskyt dětské tyreopatie: testovaná lokalita Příbramsko*. Praha, 2005. Rigorózní práce, Ústav geologie a paleontologie, PŘF UK. Vedoucí práce Karel Martínek.

TOMSA, Jan. *Odvození normální rozdělení* [online]. 2011 [cit. 2017-03-22]. Dostupné z: <http://mujweb.cz/jtomsa/gauss.htm>.

TONDEL, Martin a kol. *Increase of regional total cancer incidence in north Sweden due to the Chernobyl accident*. Journal of epidemiology and community health. 2004, roč. 58, č. 12, s. 1011-1016.

TRYGG, Johan; HOLMES, Elaine; LUNDSTEDT, Torbjörn. *Chemometrics in metabonomics*. Journal of proteome research. 2007, roč. 6, č. 2, s. 469-479.

TRYGG, Johan; WOLD, Svante. *O2-PLS, a two-block (X–Y) latent variable regression (LVR) method with an integral OSC filter*. Journal of Chemometrics. 2003, roč. 17, č. 1, s. 53-64.

TRYGG, Johan; WOLD, Svante. *Orthogonal projections to latent structures (O-PLS)*. Journal of chemometrics. 2002, roč. 16, č. 3, s. 119-128.

VOŽENÍLEK, Vít. *Geografické informační systémy I: pojetí, historie, základní komponenty*. Olomouc: Vydavatelství Univerzity Palackého, 2008. ISBN 80-7067-802-X.

VRTIŠKOVÁ, Marie. *Zdravotní nauky*. Praha, 2012. ISBN 978-80-87779–01–9.

VÚKOZ. *Historie* [online]. 2016 [cit. 2016-06-22]. Dostupné z: <http://www.vukoz.cz/index.php/cs/o-nas/historie>.

WANG, Wenqiao a kol. *A GIS-based spatial correlation analysis for ambient air pollution and AECOPD hospitalizations in Jinan, China*. Respiratory medicine. 2015, roč. 109, č. 3, s. 372-378.

WORLD HEALTH ORGANIZATION a kol. *Assessment of iodine deficiency disorders and monitoring their elimination: a guide for programme managers*. 2007.

## Seznam příloh

Příloha 1	CD s elektronickou podobou práce s příloženými soubory
Příloha 2	Seznam vybraných parametrů endokrinologických dat
Příloha 3	Výsledky testů normality
Příloha 4	Mocninná transformace dat
Příloha 5	Překročené normy štítné žlázy na 1000 obyvatel za obce