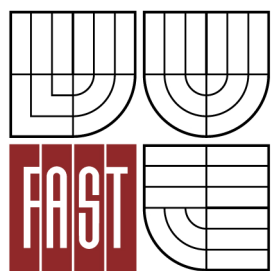




VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA STAVEBNÍ
ÚSTAV GEODÉZIE

FACULTY OF CIVIL ENGINEERING
INSTITUTE OF GEODESY

METODIKA ŘEŠENÍ MASIVNÍCH ÚLOH V GIS

METHODOLOGY FOR THE SOLUTION OF MASSIVE TASKS IN GIS

DISERTAČNÍ PRÁCE
DOCTORAL THESIS

AUTOR PRÁCE
AUTHOR

Ing. IRENA OPATŘILOVÁ

VEDOUCÍ PRÁCE
SUPERVISOR

doc. Ing. DALIBOR BARTONĚK, CSc.

BRNO 2015

Bibliografická citace VŠKP

Ing. Irena Opatřilová *Metodika řešení masivních úloh v GIS*. Brno, 2015. 95 s. Disertační práce. Vysoké učení technické v Brně, Fakulta stavební, Ústav geodézie. Vedoucí práce doc. Ing. Dalibor Bartoněk, CSc.

Abstrakt: Disertační práce se zabývá problematikou řešení masivních úloh v GIS. Tyto úlohy zpracovávají geografická data velkých objemů a různých formátů. Práce popisuje teoretický rozbor složitosti úloh a možnosti optimalizace dílčích procesů, které vedou k přijatelnému řešení. Zamýšlí se nad možností využití paralelismu v GIS, čímž lze zrychlit zpracování velkého objemu geodat. Navrhuje také způsob optimalizace procesů prostřednictvím algoritmu, který stanoví počet nutných prostředků k úspěšnému vyřešení úlohy v zadaném čase a k přiřazení procesů těmto prostředkům. Dále je zde navržen algoritmus pro optimalizaci při přípravě dat rozsáhlých GIS projektů. Algoritmy byly ověřeny v rámci výzkumného projektu, jehož cílem byla analýza povrchů terénu nad plynovody na území ČR vyjma dvou krajů. Hlavní metodou analýzy byla klasifikace obrazu ortofota, která byla dále zpřesněna filtrací z vrstev ZABAGED. Proto se práce zabývá i možnostmi zpřesnění výsledků klasifikace obrazu s využitím nástrojů GIS a stanovením chybovosti výsledků analýzy. Výstupy analýzy jsou nyní využívány pro strategické plánování údržby a rozvoje plynárenských zařízení v ČR. Výsledky práce mají obecný význam pro řešení stejné třídy úloh v GIS.

Abstract: This doctoral thesis deals with the issue of solving massive tasks in GIS. These tasks process large volumes of geographic data with different formats. The thesis describes a theoretical analysis of the complexity of tasks and the possibilities to optimize sub-processes which lead to an acceptable solution. It considers the possibility of using parallelism in GIS, which leads to an acceleration in the processing of large volumes of geographic data. It also proposes a method for the optimization of processes through an algorithm which determines the number of means necessary for the successful solution of a task at a specified time and assigns processes to these means. Additionally, there is a proposed algorithm for the optimization of the preparation of data for extensive GIS projects. The algorithms have been validated by the results of a research project, the aim of which was to analyse the terrain surface above a gas line in the Czech Republic. The primary method of analysis was the classification of an orthophoto image, which was further refined through filtration using the ZABAGED layers. Therefore, the thesis deals with the possibility of improving the results of image classification using GIS instruments as well as dealing with the determination of the error rate in analysis results. The results of the analysis are now used for the strategic planning of maintenance and the development of gas facilities in the Czech Republic. The results of the work have general importance regarding the performance of other operations of the same class in GIS.

Klíčová slova: Geografický informační systém, optimalizace, masivní úlohy, klasifikace obrazu, prostorové analýzy

Keywords: Geographic information system, optimization, massive tasks, image classification, spatial analysis

Prohlášení:

Prohlašuji, že jsem disertační práci včetně příloh zpracovala samostatně a že jsem uvedla všechny použité informační zdroje.

V Brně, dne 17. ledna 2015

.....
podpis autora

Poděkování:

Chtěla bych poděkovat svému školiteli, panu doc. Ing. Daliboru Bartoňkovi, CSc., za podporu během doktorského studia a za odborné konzultace, cenné rady a připomínky při vypracovávání disertační práce. Dále bych ráda poděkovala svým rodičům za veškerou podporu, kterou mi věnovali během mých studijních let a také bych chtěla poděkovat mému partnerovi, Jiřímu Křečkovi, který mi je v životě velikou oporou.

Obsah

1	ÚVOD.....	7
2	SOUČASNÝ STAV ŘEŠENÉ PROBLEMATIKY	9
2.1	Obecná problematika řešení složitých úloh	9
2.2	Technologie zpracování velkého množství geografických dat.....	11
2.3	Optimalizace přípravy rozsáhlých projektů v GIS.....	13
2.4	Klasifikace obrazu, její zpřesnění a stanovení chybovosti	14
2.5	Zhodnocení současného stavu řešené problematiky	17
2.6	Cíl disertační práce	19
3	TEORETICKÝ ZÁKLAD.....	20
3.1	Řešení úloh v GIS	20
3.2	Příprava dat	22
3.3	Algoritmus pro optimalizaci zpracování velkého množství geodat.....	23
3.3.1	Popis optimalizačního algoritmu	25
3.3.2	Optimalizační algoritmus v metajazyce	26
3.4	Zpřesnění výsledků klasifikace obrazu	29
3.5	Stanovení chybovosti analýzy.....	30
4	METODIKA ŘEŠENÍ.....	32
4.1	Specifikace projektu	32
4.2	Příprava dat	32
4.3	Výkonnost a řešitelnost složité úlohy	34
4.4	Navržená technologie zpracování velkého množství geografických dat.....	35
4.5	Zpřesnění klasifikace obrazu	36
4.6	Stanovení chybovosti analýzy.....	38
5	EXPERIMENTÁLNÍ VÝSLEDKY	39
5.1	Datová analýza povrchů terénu.....	39
5.1.1	Volba vhodných datových sad.....	40
5.1.2	Základní zpracovatelská jednotka	41
5.1.3	Výběr vhodného software a hardware	41
5.1.4	Fáze postupu řešení úlohy	43
5.2	Příprava dat	44
5.2.1	Úprava zdrojových dat	44
5.2.2	Příprava datových zpracovatelských jednotek	50
5.3	Zpracování dat v GIS	54
5.4	Kontrola dat	58
5.5	Distribuce výsledků a kompletace výstupu.....	60
5.6	Stanovení chybovosti výstupu analýzy.....	60
5.6.1	Analýza příčin chybovosti výsledku datové analýzy	61
5.6.2	Vyhodnocení zpřesnění výsledků klasifikace obrazu.....	63
5.6.3	Vyhodnocení absolutní chybovosti analýzy	64
5.6.4	Vyhodnocení chybovosti v závislosti na barevnosti ortofoto.....	64
5.7	Automatizace zpracování pomocí skriptů v jazyce Python	65
5.7.1	Popis koncepce automatizace procesů.....	65

5.7.2	Přehled řídicích souborů.....	67
5.7.3	Popis jednotlivých skriptů	68
5.8	Optimalizace přípravy dat a poměr automatizace k ruční práci	72
5.8.1	Vyhodnocení optimalizace přípravy dat.....	72
5.8.2	Vyhodnocení poměru automatizace k ruční práci	73
5.9	Úskalí použité technologické linky.....	75
5.10	Zkušenosti z řešení	76
6	ZÁVĚR.....	79
	SOUVISEJÍCÍ PROJEKTY	81
	SOUVISEJÍCÍ PRODUKTY	81
	PŘEHLED POUŽITÝCH ZKRATEK.....	82
	SEZNAM OBRÁZKŮ	83
	SEZNAM TABULEK	84
	LITERATURA	85
	PŘEHLED PUBLIKOVANÝCH PRACÍ STUDENTA.....	92

1 ÚVOD

Tematicky je oblast disertační práce zaměřena na problematiku řešení masivních úloh v geografických informačních systémech (GIS). GIS patří v dnešní době k významným nástrojům, díky kterým lze spravovat, analyzovat a vizualizovat geografická data. Potenciál GIS je široký a jeho rozsah využití zasahuje do mnoha oborů. K těm patří např. kartografie, státní správa a samospráva, územní rozhodování, zemědělství, životní prostředí, krizové řízení, správa inženýrských sítí a mnoho dalších.

Úlohy v GIS se vyznačují tím, že převážná část operací probíhá nad geodatabází obsahující územně vázaná data, tzv. geodata. Právě rozsah modelovaného území a typ vstupních dat výrazně odlišují procesy GIS oproti procesům, které probíhají v běžných informačních systémech. V GIS je potřeba v některých případech pracovat s velkým objemem prostorových dat, přičemž se tento objem díky dynamicky se rozvíjející dostupnosti dalších užitečných datových sad a zdokonalováním metod jejich pořizování neustále zvětšuje.

Automatizovaná tvorba kartografických děl softwarovými nástroji GIS umožňuje řešení rozsáhlých a složitých úloh, které jsou spojené s enormním objemem geodat. Rozsáhlé a složité úlohy vedou k masivním výpočtům, které jsou velmi náročné na přípravu dat, strojový čas, kapacitu paměti a výkonnost použité výpočetní techniky. Disertační práce se zabývá teoretickým rozбором složitosti úloh a možnostmi optimalizace dílčích procesů, které vedou k přijatelnému řešení.

Problematika zpracování objemných dat v GIS má své zvláštnosti, které lze stručně charakterizovat těmito faktory – data jsou vztažená ke své geografické poloze, obsahují vnitřní složitosti dané zejména počtem geografických objektů a jejich zpracování probíhá většinou pomocí prostorových analýz. Prostorové operace bývají v těchto případech časově náročné a s tím souvisí i jejich spolehlivost. V disertační práci je popsána jedna z možných metodik řešení prostorových dat značného objemu s respektováním výše uvedených faktorů.

V rozsáhlých projektech územní analýzy zpracovávaných v prostředí GIS lze při práci s velkým množstvím prostorových geodat využít dvou typů dat. Veřejně dostupná data státního mapového díla (např. ortofoto, anebo datové sady ZABAGED), anebo tematická data zpravidla privátní, která jsou poskytována příslušným správcem technické infrastruktury z jeho disponibilních zdrojů v surovém stavu a pro potřeby konkrétního projektu. Pro účel datových analýz je potřeba tato data vždy nějakým způsobem připravit.

Z hlediska značných datových objemů je téměř vyloučeno pracovat s celkovým objemem dat najednou vzhledem ke spolehlivosti, kapacitě a výkonnosti hardware a software. Proto je nutné vstupní datové sady předzpracovat a vhodně je rozčlenit na menší části. Problém spočívá v nalezení varianty takového rozčlenění, aby poměr strojočasu potřebného na zpracování úloh a času režijního (přípravné a pomocné operace vykonávané zpravidla lidským faktorem) byl co největší a celý proces v GIS probíhal efektivně, v nejkratším čase a se zachováním požadované kvality výstupů. Náročnost datové analýzy je dána především strukturou vstupních dat a její implementací do prostředí GIS. Tato fáze přípravy dat je velmi důležitým krokem při řešení složitých úloh. Proto disertační práce řeší metodiku předzpracování datových sad v rámci rozsáhlých GIS projektů s cílem dosáhnout optimálního

zpracování projektu v rámci dané informační infrastruktury (hardware, software) s uvážením lidského faktoru. Efektivní datové struktury se vytvářejí s využitím optimalizačních metod. V disertační práci je navržen vhodný způsob optimalizace rozdělení dat do menších částí.

V disertaci není řešena jen optimalizace přípravy dat, ale práce se zabývá i obecnou optimalizací zpracování velkého objemu geografických dat. Podstatou metody je hierarchický rozklad množiny procesů na elementární procesy a přiřazení prostředků těmto procesům. Dále je v práci analyzována časová náročnost masivních úloh v závislosti na typu vstupních dat a měřítkovém koeficientu, který je definován jako poměr rozsahu modelované oblasti k rozměru nejmenšího detailu ve vstupních datech.

Při produkci map pomocí technologie GIS se kladou vysoké nároky na kvalitu výstupů. K dosažení požadované kvality je zapotřebí značné množství vstupních dat a nasazení rozsáhlého repertoáru GIS nástrojů, což vede ke složitosti celého zpracovatelského procesu. V disertační práci je proto diskutována i kvalita dosažených výsledků.

Teoretické předpoklady metodiky řešení složitých úloh v GIS byly ověřeny na výzkumném projektu datové analýzy uložení plynárenských zařízení pod určitými typy terénních povrchů na území ČR. Projekt byl řešen pro společnost RWE za účelem stanovení reprodukčních hodnot plynárenských zařízení (plynovodů) a ocenění nákladů, které by bylo nutné vynaložit na vybudování nových sítí. Výsledky analýzy jsou nyní využívány pro strategické plánování údržby a rozvoje plynárenských zařízení v ČR.

Řešení tohoto projektu lze považovat za masivní úlohu, která byla zpracována v prostředí GIS a jehož hlavní metodou byla klasifikace obrazu ortofota s následnou filtrací. Proto se část disertační práce zabývá možnostmi zpřesnění klasifikace obrazu a zvýšení technické výtěžitelnosti informací z rastrového obrazu prostřednictvím nástrojů GIS. Dále je v práci popsána problematika stanovení chybovosti při automatizované klasifikaci obrazu se zpřesněním výsledků pomocí filtrace. Výsledkem projektu je ověřená technologie, jejíž hlavním cílem byla její jednoduchost, rychlost řešení s ohledem na spolehlivost, snadná kontrola dílčích procesů, kvalita výsledků dosažených za podmínek minimálních nároků na technické vybavení a personální zajištění celého projektu. Výsledky projektu mají obecný význam nejen pro tvorbu kartografických děl s podporou GIS, ale jsou obecně použitelné pro řešení úloh dané třídy.

2 SOUČASNÝ STAV ŘEŠENÉ PROBLEMATIKY

Oblast disertační práce pokrývá tato 4 klíčová témata, z kterých je zachycen a prostudován současný stav problematiky. Jde o tyto oblasti:

1. Obecná problematika řešení složitých úloh
2. Technologie zpracování velkého množství geografických dat
3. Optimalizace přípravy rozsáhlých projektů v GIS
4. Klasifikace obrazu, její zpřesnění a stanovení chybovosti.

První část se zabývá řešením složitých úloh z obecného hlediska a charakterizuje nejčastěji využívané metody pro řešení těchto úloh. Druhá část je již více zaměřena na zpracování velkého množství dat v GIS. Následuje část popisující velmi důležitou fázi u masivních úloh v GIS, a to je příprava dat a její optimalizace. Závěrečná část se zabývá hlavní metodou, která byla využita v projektu souvisejícím s experimentálními výsledky disertační práce.

2.1 *Obecná problematika řešení složitých úloh*

Problematice složitých úloh je věnována pozornost už od 70. let minulého století. Existuje celá řada publikací s touto tematikou. Z tohoto důvodu bude v této části uveden jen výčet takových prací, které mají blízký vztah k disertační práci. Pro řešení složitých úloh byla vyvinuta řada metod, které mohou být rozděleny do těchto kategorií:

- a) využití paralelních procesů v úloze, které je možné zpracovávat souběžně
- b) rozdělení úlohy na dílčí části, jejichž řešení je jednodušší a spolehlivější, než řešení celé úlohy – tyto úlohy se pak řeší postupně (sériově)
- c) využití speciálních algoritmů, které optimalizují procesy probíhající v dané úloze
- d) zavedení organizačních opatření do výpočetního procesu s cílem zjednodušit a zrychlit dobu řešení dané úlohy.

a) Paralelní procesy

Charakteristické pro tuto skupinu je, že použité metody vyžadují speciální technické vybavení. Většinou jde o mnohoprocesorové počítače, kterým se také říká akcelerátory. Do této kategorie můžeme zařadit práce: Publikace [1] popisuje speciální rozhraní sloužící ke komunikaci mezi procesory v paralelních procesech s řízeným poměrem počet výpočtů/objem komunikace. Navržené rozhraní v příspěvku slouží pro výpočet příslušejících oblastí digitálního výškového modelu. Kapitola v knize [2] zkoumá proces definice a optimalizace relativně jednoduchého algoritmu přiřazení úloh ve speciální hardwarové a softwarové architektuře a dále poukazuje na některé rozdíly v rozvoji masivně paralelních GPU (Graphics Processing Unit) architektur a tradičnějších CPU (Central Processing Unit) s jedním či více jádry. Článek [3] popisuje návrh vysokorychlostní sítě, která propojuje superpočítače spolu s dalšími grafickými zařízeními, jenž slouží jako vysoce výkonné masivně paralelní stroje. Přes tuto síť dochází k interaktivnímu prohledávání a k rychlému ukládání značného množství dat. Publikace [4], na které se podílela autorka této disertační práce, navrhuje rozhraní člověk-stroj

pro zařízení k určování polohy pomocí globálních navigačních satelitních systémů. Návrh je založen na stavovém automatu s hierarchickými stavy, přičemž některé z těchto stavů mohou nastat paralelně.

b) Rozdělení úlohy na dílčí části

Dílčí úlohy se běžně řeší sériově na výkonných počítačích. Do této kategorie můžeme zařadit především tyto práce. Příspěvek [5] představuje návrh a implementaci aplikace, jehož cílem je umožnit ve velkém rozsahu distribuci výpočtů prostřednictvím seskupení volně vázaných počítačů, které spolupracují tak, že se navenek jeví jako jeden celek. V literatuře [6] je popsáno využití multifraktálního vlnového modelu u dopravních sítí s využitím masivních datových sad. Příspěvek [7] řeší problematiku řízení geoprostorových informací. Je zde navržen mechanismus spolupráce distribuovaných uzlů, databázové uložení a řídicí strategie masivních geoprostorových informací. Ve článku [8] je popsán návrh efektivního operačního prostředí pro distribuované geoprostorové informace a jejich výpočetní analýzy. Publikace [9] představuje druh sériového algoritmu, který generuje zastínění v rozsáhlých digitálních výškových modelech. Tento algoritmus vznikl z důvodu neustále rostoucího množství dat velkých měřítek s vysokým rozlišením, které mají existující GIS algoritmy problém zpracovat. Článek [10] se zabývá efektivním zpracováním prostorových dotazů z masivního množství dat v GIS aplikacích. Prostorové dotazy jsou složeny z tzv. multi-way prostorových spojení, které počítají vztahy (např. průnik) mezi daty. Pro tento případ byla vybrána konkrétní technika výběru prostorového kruhového spojení.

c) Speciální optimalizační algoritmy

Příklady speciálních algoritmů jsou popsány v těchto pracích. Příspěvek [11] představuje novou strukturu uložení dat pro tzv. ad-hoc souhrnné dotazy. Tyto dotazy jsou velmi důležité pro dotazování v náročných aplikacích pro zpracování mračen bodů. Článek [12] zkoumá matematické modely, simulační přístupy a testbedy (testování nových vědeckých teorií) používané pro výzkum v rozsáhlém zpracování mračen bodů. Oblastí LiDARu a fotogrammetrií se zabývá dokument [13], ve kterém jsou definovány požadavky na infrastrukturu systému umožňující integraci 3D mračen bodů z heterogenních snímacích zařízení v různých časových obdobích. Tento článek navíc popisuje problematiku tzv. Out-of-core algoritmů, neboli externích paměťových algoritmů, sloužící ke zpracování dat, která jsou příliš velká na to, aby se vešla do hlavní paměti počítače současně. Tyto algoritmy musí být pak optimalizovány tak, aby efektivně načítaly a přistupovaly k datům uloženým na pomalých velkokapacitních pamětech, jako jsou např. pevné disky [93]. Příspěvek [14] řeší vylepšení vizualizační technologie pro povrch terénu ve 3D GIS prostřednictvím tzv. Level of detail (úroveň detailu) algoritmu, který využívá struktury čtyřstromů. Čtyřstrom je zástupcem hierarchických rastrových struktur. Příspěvek [15] prezentuje nový vylepšený multi-way algoritmus pro prostorové spojení, které je nazváno Multi-way R-stromové spojení a které využívá nepřímé predikáty. R-strom je stromová datová struktura používaná pro prostorové přístupové metody, jako je např. indexování vícerozměrných struktur v GIS [94].

d) Organizační opatření ve výpočetních procesech

V mnoha případech jde o metody založené na teorii grafů určené pro řešení masivních výpočtů. Článek [16] se zabývá aktuálním vývojem v oblasti výpočetní techniky, a to metodou pro výpočet koeficientů Taylorova rozvoje v oblasti kvantové fyziky. Publikace [17] představuje návrh kombinace paralelního programovacího módu MapReduce s algoritmem pro využití hrubých množin. Takto navržená kombinace je účinnou metodou dolování dat pro oblast zpracování mračen bodů. Článek [18] prezentuje efektivní algoritmus pro hydrologické modelování, založený na rastrových datových sadách terénu. Navržený algoritmus je optimalizován pro masivní data. Kapitola v knize [19] se zabývá geoprostorovými výpočty velkého množství dat. Z oblasti digitálních výškových modelů a 3D dat představuje článek [20] webovou zpracovatelskou službu (WPS), která slouží ke zpracování masivních prvotních geodat terénu v rastrovém tvaru. Příspěvek [21] představuje prostorovou informační infrastrukturu SIG, která obsahuje tři vrstvy: uživatelé-aplikace, služby a informace. Tato infrastruktura propojuje zdroje prostorových dat, výpočty, uložení dat, software, nástroje a uživatele. Článek [22] prezentuje zkušenosti s problematikou GIS ve výstavbě při přípravě staveb, ve fázi realizace a následného uvedení do provozu.

2.2 Technologie zpracování velkého množství geografických dat

V GIS je potřeba v mnoha případech pracovat s enormním objemem dat a vlivem dynamicky se rozvíjející dostupnosti dalších potřebných dat se tento objem neustále zvyšuje. Problematika zpracování velkého množství geografických dat v GIS má své zvláštnosti, které můžeme stručně charakterizovat těmito hlavními faktory:

- data jsou územně vázaná (vztahena ke geo-referenčnímu souřadnicovému systému)
- data obsahují vnitřní složitost a topologii, která je dána zejména počtem geografických objektů a jejich propojením
- převážná část procesů má charakter prostorových analýz, které jsou časově náročné a s tím souvisí i jejich spolehlivost.

Jak již bylo uvedeno výše, pro řešení úloh na matematických strojích s enormním objemem dat se v obecných systémech využívá zejména 2 osvědčených metod:

- a) ***řetězené zpracování*** (pipe-line processing) ve speciální architektuře počítače. Zpracovatelská linka je v tomto případě tvořena sérií specializovaných funkčních bloků, z nichž každý zpracovává dílčí úlohu stejného typu. Protože se doby trvání dílčích procesů mohou časově překrývat, dochází ve výsledném efektu ke zrychlení řešení celé úlohy proti zpracování na jednoprocessorovém počítači.
- b) ***rozdělení úlohy na dílčí pod-úlohy*** a jejich řešení na paralelní počítačové architektuře (multi-processing). Účinnost této metody je dána především závislostí dílčích procesů. Pokud jsou dílčí procesy na sobě nezávislé, je výkonnost celého systému maximální.

V GIS úlohách se pro řešení úloh nad rozsáhlými databázemi využívá zvláštních vlastností geografických datových sad (viz výše uvedené faktory). Vzhledem k územní vazbě dat je

možné rozdělit celkovou úlohu na dílčí pod-úlohy s minimální vzájemnou závislostí. Potom je možné s výhodou využít metody ad b). Do této skupiny patří zejména tyto práce.

Publikace [23] představuje některé paralelní přístupy k efektivní organizaci dat s využitím transputerové sítě. Transputer je jednočipový mikropočítač obsahující čtyři meziprocesorové komunikační kanály [95]. Navržený paralelismus byl využit pro výpočty viditelnosti v digitálním modelu terénu. Literatura [24] definuje inteligentní GIS jako nástroj ke zpracování různých typů geovědních znalostí souvisejících s velkými objemy dat. Představuje třívrstvou hierarchii o geovědních znalostech v GIS, přičemž první vrstva se týká dat, druhá jejich zpracováním a třetí vrstva je o znalostech z GIS pro veřejný sektor. Kapitola v knize [25] se zabývá technikami návrhů I/O (vstup/výstup) algoritmů pro efektivní zpracování obrovského množství dat v GIS. Ve článku [26] je představen temporární databázový model relačního typu se vztahy entit mezi tabulkami s komponenty. Model je implementován pomocí databázového serveru Oracle a integruje zpracovaná data s existujícími prostorovými dataseťmi pro další analýzy. Přístup má tu výhodu, že na rozdíl od tradičních prostorových analýz v GIS se pracuje s menšími sadami dat. Publikace [27] navrhuje strukturu architektury, která je složená z několika jader procesorů a slouží ke zrychlení zpracování detekce vodstva z multispektrálních satelitních snímků. V článku [28] je navržena univerzální paralelní struktura pro zpracování masivních mračen bodů pořízených LiDARem.

Ostatní práce používají kombinace obou metod a využívají zvláštnosti geografických dat. Sem mohou být zařazeny následující publikace, přičemž z nich je největší zastoupení v environmentální oblasti. Článek [29] se zabývá využitím velké databáze 3D dat, která je v GIS spojena s matematickým modelem pro předpovědi a monitorování znečištění životního prostředí. Kapitola v knize [30] pojednává o programu založeném na GIS, který slouží pro podporu rozhodování z oblasti životního prostředí. Tento systém pracuje s velkým množstvím dat a kombinuje standardní funkce GIS spojené s databází a matematickými modely environmentálních procesů. Článek [31] popisuje aplikaci navrženou na bázi GIS, která tvoří simulační model pro problematiku životního prostředí. Díky tomuto modelu lze analyzovat oblasti budoucího využití krajiny s několika scénáři. Aplikace slouží zejména pro strategická plánování bytové výstavby s dopadem na životní prostředí. Publikace [32] představuje nástroj pro vizualizaci velkého množství environmentálních dat. Tento nástroj spojuje prvky GIS, CAD a grafického software a nabízí širokou škálu výstupů od jednoduchých grafických ilustrací až po sofistikované vědecké vizualizace.

Článek [33] představuje platformu GIS pro zpracování mračen bodů z velkého objemu dat týkajícího se vodních zdrojů. Publikace [34] popisuje organizaci a řízení značného množství dat z oblastí zemědělství. Na základě GIS se snaží analyzovat rozložení a rozměry skleníků v daném území. Článek [35] řeší problematiku nestále se zvyšujícího se počtu informací a dat v dnešní době. Z tohoto pohledu rozebírá témata databází, datové organizace a datového zpracování. Nestálým nárůstem nových informací a dat se zabývá i publikace [36], jejímž tématem je využití GIS v archeologii. Článek [37] popisuje propojení geografických dat virtuální reality s 3D GIS ve webovém prostředí. Publikace [38] řeší problematiku zpracování obrazu v dálkovém průzkumu Země, kdy je potřeba transformovat různé datové sady do stejného souřadnicového systému a dále tak zkoumá techniky efektivního automatizovaného zpracování velkého množství dat. Nejbližší vztah k experimentálním výsledkům disertační

práce má kapitola v knize [39], která popisuje, jak energetické organizace využívají technologie GIS k překonávání dnešních kritických podnikatelských problémů a zajištění spolehlivé energie pro budoucnost.

2.3 *Optimalizace přípravy rozsáhlých projektů v GIS*

V obecných systémech se optimalizace procesů využívá v obou dvou výše uvedených metodách zpracování úloh s velkým objemem vstupních dat, tj. u metody paralelismu a u metody hledající společné dílčí procesy ve všech procesech v dané úloze. Využití optimalizace je v různých fázích zpracování podle povahy dané úlohy, která se má řešit. Cílem je tak zrychlení a zjednodušení všech procesů v systému.

Optimalizaci výše uvedených úloh můžeme prakticky aplikovat na dvou úrovních:

1. na úrovni makroprocesů, tj. úloh, které souvisejí s koncepcí řešení jako celku. Jde např. o přidělování úloh technickým prostředkům, tj. procesorům nebo celým počítačovým systémům.
2. na úrovni mikroprocesů, kdy se snažíme zrychlit nebo zjednodušit zpracování úloh, které již byly přiřazeny prostředkům v rámci těchto technických zařízení.

Disertační práce se zabývá optimalizací ad 1), tj. na úrovni makroprocesů. Jde o problematiku, která se řeší na začátku projektu, kdy potřebujeme odhadnout, zda zadanou úlohu nebo soubor úloh můžeme vyřešit ve stanovené lhůtě s podporou současné informační infrastruktury, kterou máme k dispozici. Jádrem problému je metoda rozdělení úlohy na dílčí procesy a přidělování dostupných prostředků pro zpracování těchto dílčích procesů.

Problematikou předzpracování dat v GIS se zabývá řada prací. Z nich budou uvedeny opět jen ty nejdůležitější studie, které mají blízký vztah k disertační práci. Z hlediska metody a přístupu je můžeme rozdělit do těchto skupin:

- a) *Optimalizační metody, založené na osvědčených heuristických algoritmech.* Publikace [40] se zabývá návrhem efektivní procedury přípravy dat pro rozsáhlé projekty. Článek [41] řeší problém zpoždování dotazů v závislosti na nestále se zvětšující objem datových skladů s reálnými daty. Navrhuje tak optimalizaci plánování dotazů pro zlepšení doby odezvy a kvality dat s ohledem na požadavky uživatelů.
- b) *Využití stromových struktur.* Sem patří literatura [42], představující efektivní externí paměťové algoritmy pro zpracování liniových segmentů v GIS s využitím obrovského množství prostorových dat. Tyto algoritmy se snaží minimalizovat problém v komunikaci mezi vstupem/výstupem a využít se dají např. pro zobrazování a modelování, překrytí map a extrahování informací z nich. Publikace [43] se zabývá metodikou pro návrh algoritmu, který má být využíván v GIS k hledání optimálních cest pro sběr vybraných kontejnerů v městských částech.
- c) *Aplikace shlukové analýzy.* Literatura [44] popisuje metodu přípravy dat v GIS k odhadování znečištění vzduchu látkou SO₂. Využívá shlukové analýzy kombinované s umělými neuronovými sítěmi. Článek [45] popisuje prostorové metody v GIS, které

analyzují tři oblasti výstavby (město, předměstí, venkov) spojené s emisemi způsobené pálením dřeva.

- d) Hierarchické členění geografických objektů. Článek [46] se zabývá problematikou určování polohy bodů ve vztahu k mnohostěnu. Práce analyzuje časovou náročnost předzpracování úlohy.
- e) Metody založené na teorii hrubých množin. V literatuře [47] je navržen algoritmus pro předzpracování dat klasifikačního modelu založeného na teorii hrubých množin, který se využívá v dolování dat. Výhodou algoritmu je, že obchází potíže se zpracováním velkého množství dat při výpočtech v procesu dolování.
- f) Aplikace metod z oblasti dolování dat. Dokument [48] navrhuje využití ontologií pro správu spojených nástrojů při předzpracování a dolování dat a popisuje MS-Analyzer, softwarový nástroj pro integrované řízení, předběžné zpracování a dolování spektrálních dat rastrového typu. Kapitola v knize [49], která se zabývá tématikou shromažďování informací a analýzou trestné činnosti, představuje analytický model Trinity Sight. Ten slouží pro prediktivní analýzy, přičemž zahrnuje faktory jako integrovanou analýzu času, prostoru, povahy události nebo potenciální hrozby.
- g) Speciální algoritmy. Zde může být zařazena práce [50], která popisuje software pro předzpracování dat v hydrologických modelech. Hydrologický model pracuje s daty digitálního výškového modelu a např. pro určení směru vodního toku využili autoři teorii grafů. Článek [51] se zabývá lokalizací archeologických nalezišť s využitím nástrojů expertního GIS. Základem metody je simulační model zájmové lokality.

Společným rysem uvedených prací jsou sofistikované postupy a optimalizační algoritmy aplikované v různých fázích předzpracování dat. Mezi nejúčinnější a současně nejlevnější způsoby, jak zefektivnit předzpracování dat, patří organizační opatření typu tzv. vnitřní normalizace, např. využívání územních číselníků, zavedení standardních názvů souborů a adresářů, respektování administrativního členění v datové a databázové struktuře vstupních dat apod.

2.4 *Klasifikace obrazu, její zpřesnění a stanovení chybovosti*

Soubor prací, zabývajících se **klasifikací obrazu a jejím zpřesněním**, je značně rozsáhlý. Mezi dvě nejvíce zastoupené skupiny řešící tuto problematiku patří DPZ a oblast počítačového vidění. Zde bude uveden pouze stručný výčet z řad publikací těchto oblastí.

- a) Dálkový průzkum Země – Zajímavý je objektový přístup k analýze obrazu v [52] využitý pro přesné mapování a monitorování v zemědělství a aplikace metody Maximum Likelihood pro segmentaci a klasifikaci multispektrálního obrazu v biologii [53]. Dále pak [54], kde je popsána metoda klasifikace obrazu na bázi hrubých a fuzzy množin a [55], která se zabývá porovnáním metod diskretních hrubých množin a metody hlavních komponent (PCA) pro klasifikaci obrazu DPZ. V této oblasti klasifikace lze dále uvést práce [56], zabývajících se vyhodnocením přesnosti klasifikace snímků oblastí lesů a bažin z družice Landsat na základě pravidel v GIS a [57], pojednávající o možnosti paralelního zpracování obrazů DPZ. Literatura [58] se

zabývá klasifikací obrazu využitím teorie grafů. Topologie detekovaných segmentů v obrazu je definována pomocí grafových algoritmů.

- b) *Počítačové vidění* – Zpřesněním výsledků klasifikace obrazu se zabývá práce [59], kde je využito Gaussovského rozložení pravděpodobnosti příslušnosti k třídám objektů. Dále publikace [60], v níž se využívá metody tzv. mravenčích kolonií pro optimalizaci klasifikace obrazu získaného zpracováním mračna bodů a článek [61], v němž je navržen algoritmus pro klasifikaci obrazu založený na metodě roje hmyzu. Podobná problematika je řešena v práci [62], a to využitím genetického algoritmu pro klasifikaci obrazu. V literatuře lze nalézt poměrně značné množství prací, které využívají pro klasifikaci obrazu metodu SVM (Support Vector Machine). Jde o práci [63], která se zabývá detekcí objektů v obrazu s využitím Bayesovských rozhodovacích pravidel a článek [64], pojednávající o klasifikaci obrazu na základě histogramu s využitím SVM. Velmi čtené je využití klasifikace obrazu v lékařství. Jako typický příklad můžeme uvést článek [65], pojednávající o hybridní metodě pro redukci šumu a detekci zvláštních objektů ve snímcích magnetické rezonance.

Práce týkající se **chybovosti výsledků klasifikace obrazu** mohou být rozděleny podle aplikačního oboru a podle použitých metod:

a) podle aplikačního oboru

- *Dálkový průzkum Země* - z hlediska aplikace je tato oblast zastoupena nejčteněji. Můžeme sem zařadit následující práce. Publikace [66] vyhodnocuje objekty z klasifikovaných snímků podle tvaru a podle polohy. Tyto objekty byly srovnány v GIS s odpovídajícími referenčními objekty na snímcích, dále byla vyhodnocena přesnost tematická a geometrická. Přesné mapování městských oblastí je důležité, ale vzhledem k rozmanitosti výškové zástavby způsobující povrchové zakrytí obtížné. Studie [67] představuje jak synergicky spojit optické a SAR data ke zlepšení mapování těchto míst povrchů. Vývoj časově konzistentní časové řady krajinného pokrytí (LCTS) ze satelitního pozorování Země je obtížné, protože pozorování z více let je získáváno za různých podmínek s různou meziroční odrazivostí povrchů. Článek [68] představujeme přístup k rozvoji změn LCTS na základě svého výzkumu, s pokrokem v detekci změn, v trénovacích množinách a v klasifikaci. Literatura [69] představuje statistické zpracování vzorků ve formě protokolu pro posouzení tematické a prostorové kvality klasifikačních map DPZ generovaných z velmi vysokých rozlišení. Článek [70] studuje analýzu přesnosti řízené klasifikace vyhodnocením satelitních snímků s ohledem pro více typů rozlišení. Cílem práce [71] je popsat nový přístup k získání map využití krajiny a krajinného pokryvu z klasifikace obrazu s vyrovnáním chyb klasifikace v oblastech pro každou třídu. Navržená metoda je doporučena zejména pro studie, jejichž cílem je poskytnout multitemporální srovnání. Článek [72] řeší problematiku mapování ekologicky významných kategorií krajinného pokryvu v spektrálně komplexním městském prostředí. Cílem práce [73] bylo vyvinout metodu automatického klasifikování nepravidelných časových řad snímků s minimem parametrizace a apriorních informací. Na rozdíl od tradičních metod pro rozšíření signatur, automatická procedura adaptivního zobecnění signatur přizpůsobí

spektrální znaky tříd k jednotlivým snímkům, a proto nevyžaduje žádné korekce obrazu. Signatury tříd jsou odvozeny z pixelů ze stabilního krajinného pokryvu v průběhu času. Studie popsaná ve článku [74] hodnotí potenciál dat získaných pro C a L pásmo ze SAR. Je zde posuzována i kvalita těchto dat, které sloužily pro stanovení oblastí výskytu ložisek železa v Amazonii. Problematikou posouzení přesnosti map krajinného pokryvu a jeho změn je předmětem článku [75]. Přesnost je zde vyjadřována chybovou maticí vypočítanou z vhodných oblastí na mapách. Literatura [76] představuje metodu faktorových grafů, která byla využita pro integraci dat získaných z různých fyzikálních typů senzorů s různým rozlišením a s různými typy digitálních GIS dat.

- Potravinářský průmysl (testování kvality potravin) a zemědělství (předpověď úrody v dané oblasti) – Literatura [77] posuzuje předběžné a v sezóně dálkově snímané veličiny pro předpovídání výnosů kukuřice a sóji ve Spojených státech. Publikace [78] vyhodnocuje přesnost dvou základních technik klasifikace obrazu pro využití krajiny v tropických oblastech. Konkrétně se jedná o klasifikátory Maximum Likelihood a nejbližšího souseda. Kvalita hyperspektrálního snímání jako techniky v potravinářském průmyslu pro klasifikaci a třídění, detekci vad a nemocí, vizualizaci chemických vlastností a vyhodnocení celkové kvality masa, ryb, ovoce, zeleniny a dalších potravinářských výrobků, je předmětem článku [79].
- Lékařství a biologie - Cílem práce [80] bylo odhadnout několik vhodných výrazných bodů na měkkých tkáních obličeje, které by sloužily jako referenční body pro metrické obličejové analýzy ve 3D. Ve článku [81] je navržena automatická metoda k segmentaci a klasifikaci jader prsních buněk s cílem poskytnout kvantitativní posuzování a diagnostiku.
- Ekologie - Literatura [82] popisuje použití kombinace metod klasifikace obrazu k modelování vhodných stanovišť bělokura horského.
- Geodézie - Článek [83] prezentuje zkoumání objektivě orientovaných snímků barevných ortofotomap s různou úrovní komprese a analyzuje účinky ztráty dat způsobené kompresí na výsledky klasifikace obrazu u těchto snímků.

b) podle nejčastěji použitých metod

- klasické metody *vylepšení obrazu* (filtrace, transformace jasu, diskrétní konvoluce) [66]
- využití metody *Support vector machine* (SVM) [71]
- kombinace *různých způsobů pořízení snímků* povrchu Země [67], [82]
- zpracování *časové posloupnosti snímků* [68], [73]
- využití *teorie grafů* (faktory grafu) [76]
- aplikace *korelační analýzy s maticí chyb* [75]
- využití *metrických prostorů a teorie tolerance* [84]. Tato publikace má bezprostřední vztah k disertační práci.

2.5 Zhodnocení současného stavu řešené problematiky

V tabulce 1 je výčet a hodnocení nejčastěji používaných metod týkající se problematiky disertační práce. Konkrétně jsou zde hodnoceny metody pro dvě základní oblasti, a to obecné řešení složitých úloh a optimalizace přípravy dat rozsáhlých projektů. Hodnocení je zde uvedeno ze tří hledisek – složitost úlohy, její účinnost a vhodnost pro GIS. Složitost úlohy je hodnocena slovně, ostatní dva parametry jsou porovnány číselným stupněm 1 až 3, přičemž čím vyšší číselná hodnota, tím je metoda účinnější/vhodnější pro GIS.

Oblast problému	Metoda (princip) řešení	Kritéria hodnocení řešení		
		Složitost metody	Účinnost, výkonnost	Vhodnost pro GIS
Obecné řešení složitých úloh	Využití paralelismu	Velmi vysoká	3	3
	Zjednodušení rozdělením	Vysoká	1	3
	Speciální algoritmy	Střední	1	2
	Organizace procesů	Velmi vysoká	2	2
Optimalizace přípravy dat rozsáhlých projektů	Heuristika	Velmi vysoká	2	3
	Stromové struktury (hierarchie jevů nebo GO)	Střední	1	1
	Shluková analýza	Střední	2	3
	Hrubé množiny	Střední	1	1
	Dolování dat	Vysoká	1	3
	Speciální algoritmy	Střední	3	2

Tab. 1 – Hodnocení metod řešení

Z tabulky vyplývají následující závěry k jednotlivým metodám:

Obecné řešení složitých úloh

- *Využití paralelismu* – Cílem metody je rozdělit vstupní úlohu na nezávislé procesy. Toto rozdělení je nejobtížnějším článkem metody, protože jde o NP-úplný problém, který dosud nebyl vyřešen efektivním algoritmem. Pokud se úlohu podaří rozdělit, je výkonnost metody vysoká. V územně vázaných systémech je tohle rozdělení jednodušší než v obecných systémech, protože úloha se dá rozdělit podle územního členění (pokud se areály nepřekrývají, tvoří tak nezávislé oblasti pro zpracování).

- *Zjednodušení rozdělením* – Složitě entity je vhodné rozdělit (vyhlazením, generalizací) na menší z důvodu jednoduššího zpracování úlohy vzhledem k možnostem současného HW a SW. Složitost úlohy pak vychází z míry zjednodušení.
- *Speciální algoritmy* – Snahou je rozdělit obrovské množství vstupních dat tak, aby zpracovatelský segment byl na nejrychlejším médiu (paměť počítače, SSD rychlý disk). Ze zdrojových dat se tak vybírají jen data relevantní, přičemž dochází ke snížení objemu dat.
- *Organizace procesů* – Snahou je vhodně uspořádat dílčí procesy tak, aby celková doba zpracování kompletní úlohy byla co nejkratší. Algoritmicky jde o velmi obtížný problém, jehož podstatou je přidělování prostředků procesům.

Optimalizace přípravy dat rozsáhlých projektů

- *Heuristika* – Metoda vybírá z nadbytečného množství dat jen ta data, která budou postačovat ke kvalitnímu výsledku zpracování. Jde o kombinatorickou úlohu, při níž dochází k prohledávání velkého stavového prostoru. Jednou z oblastí využití je např. laserové skenování a zpracování mračen bodů.
- *Stromové struktury* – Vstupní data jsou hierarchicky rozdělena, např. podle dat reprezentující geografické objekty (GO). Ze vstupních dat lze vybírat podle stromové struktury v závislosti na úrovni detailu, tzn., čím je vyšší detail, tím je úloha složitější a výkonnost úlohy nižší.
- *Shluková analýza* – Využití této metody je např. pro územní analýzy, kde je cílem zjednodušit vyhodnocení složitých jevů nebo geografických objektů (sociologické jevy, výskyt epidemií aj.). Slouží tak nejen ke zpřehlednění výstupů, ale i k objasnění příčin jevů.
- *Hrubé množiny* – Souvisí s přesností a aproximací kvality výstupů. Podle toho jsou pak vybírány kvalitní vstupy. Řeší se zde stanovení kritéria, zda úlohu znovu zpracovat, anebo je výstup postačující.
- *Dolování dat* – Při této metodě dochází k doplnění chybějící informace na základě vstupních dat. V mnoha případech je zde využití predikce, a tudíž výsledek má pravděpodobnostní charakter. Spolehlivost metody je nižší, protože nová doplněná informace je získána nepřímou a účinnost metody je ovlivněná kvalitou predikce.
- *Speciální algoritmy* – V tomto případě jde o procesy, které pro danou (dílčí) úlohu vyberou veškerá potřebná data pro další zpracování.

Vhodné metody pro řešení složitých úloh v GIS

Z hodnot v tabulce 1 a ze stručného slovního hodnocení metod vyplývají následující závěry. Mezi nejúčinnější metodu pro řešení složitých úloh v GIS bude patřit paralelismus, jenž spočívá v rozdělení úlohy do pokud možno nezávislých procesů, které lze zpracovávat současně na několika prostředcích (počítačích). K tomu bude nejprve zapotřebí stanovit optimální dílčí zpracovatelskou jednotku z hlediska rozsahu území a potažmo i objemu vstupních dat. K jejímu určení bude vhodné použít metody experimentu nebo heuristiky. Dále bude zapotřebí navrhnout a využít speciální algoritmy pro optimalizaci přípravy dat tak, aby v následujících krocích zpracování úlohy byla využívána pouze vybraná potřebná data.

2.6 *Cíl disertační práce*

V předchozí podkapitole byl zhodnocen současný stav problematiky a vytipovány vhodné metody pro řešení problémů v těchto oblastech:

- Obecná problematika řešení složitých úloh
- Technologie zpracování velkého množství geografických dat
- Optimalizace přípravy rozsáhlých projektů v GIS
- Klasifikace obrazu, její zpřesnění a stanovení chybovosti.

V těchto oblastech jsou v mnoha případech aplikovány metody, které nevykazují optimální poměr cena/výkon/doba řešení. Ve své práci se pokusím tuto situaci zlepšit návrhem nového řešení.

Cílem disertační práce je navrhnout vhodné nebo inovovat stávající metody pro řešení výše uvedených problémů a ověřit jejich efektivitu v konkrétní aplikaci v GIS. Verifikace metodiky je realizována na projektu pro společnost RWE, jehož cílem bylo zpracovat analýzu povrchů nad průběhem plynovodů na celém území ČR vyjma dvou krajů. Podstatou řešení projektu bylo navrhnout novou metodu na bázi klasifikace obrazu s velmi vysokým rozlišením na rozsáhlém území s cílem dosáhnout co nejlepšího výsledku. Charakteristikou tohoto procesu je vytěžení informací a převod rastru na vektorový tvar obrazu, který umožňuje přesnější výsledky prováděných operací.

3 TEORETICKÝ ZÁKLAD

Tato kapitola je kvůli přehlednosti rozdělena do pěti částí. První část popisuje obecný teoretický základ u řešení masivních úloh v GIS. Druhá část se zabývá předzpracováním geografických dat. Třetí část kapitoly řeší algoritmus pro optimalizaci velkého množství geodat. Čtvrtá část zkoumá problematiku zpřesnění klasifikace obrazu a pátá část kapitoly obsahuje návrh na stanovení chybovosti výsledků klasifikace.

3.1 Řešení úloh v GIS

Řešení úloh v GIS je možné matematicky popsat touto formulí:

$$R = (I, S, Q, Y, \delta, \varphi, \psi), \quad (1)$$

kde

$$I \subseteq (A_1 \times A_2 \times \dots \times A_l), \quad (2)$$

I ... množina vstupních datových sad, A_i jsou dílčí datové sady

l ... počet dílčích datových sad

S ... množina vnitřních stavů GIS

Q ... množina přípustných výsledků (výsledných vrstev), které vyhovují požadované kvalitě

Y ... množina výstupů z GIS

δ ... koeficient územní podrobnosti daný vztahem:

$$\delta = \frac{P_{tot}}{P_{det}}, \quad (3)$$

kde

P_{tot} ... celková plocha modelovaného území, které se řeší v rámci dané úlohy

P_{det} ... plocha detailu, který je popsán ve vstupních vrstvách I

φ ... zobrazení (vstupní predikát), které má charakter přípravy vstupních dat:

$$\varphi: I \rightarrow S, \quad (4)$$

ψ ... zobrazení (výstupní predikát):

$$\psi: S \cap Q \rightarrow Y. \quad (5)$$

Operace $\gamma: S \cap Q$ představuje proces kvality výstupu (kartografického díla), který lze vyjádřit vztahem:

$$\gamma: S \cap Q = f(C, P, T, L, A), \quad (6)$$

kde (podle ISO 19 113):

C ... úplnost vstupních dat

P ... polohová přesnost

T ... časová přesnost (aktuálnost)

L ... logická (topologická) přesnost

A ... atributová (tematická) přesnost.

Časová náročnost dané úlohy v GIS z teoretického hlediska je dána vztahem:

$$T = f(\delta, \varphi, \psi, l). \quad (7)$$

V této rovnici δ představuje měřítko zobrazení z reality do digitální geodatabáze, φ reprezentuje složitost přípravy vstupních dat, tj. transformaci vstupních vrstev do vnitřních stavů GIS, ψ má charakter vnitřních procesů v GIS, tj. nástrojů pro územní operace, analýzy, editace, kontrolu kvality γ apod. a l je počet typů vstupních vrstev – viz vztah (2).

Pokusme se nyní o rozbor časové náročnosti úloh v GIS podle vztahu (7). Parametry v této rovnici můžeme rozdělit podle míry ovlivnění ze strany uživatele do 2 skupin:

1. vstupní φ a výstupní ψ predikáty jsou dány charakterem úlohy a uživatel je může ovlivnit jen z malé části
2. koeficienty δ a l naopak můžeme ovlivnit dosti značně. Parametr n je dán úspěšností výstupu a jeho kvalitou, na koeficientu δ závisí nejen doba zpracování, ale i spolehlivost vnitřních procesů ψ .

Z hlediska dob trvání jednotlivých procesů můžeme rovnici (7) přepsat také do tohoto tvaru:

$$T = T_{Ia} + T_{Im} + T_{Oa} + T_{Om}, \quad (8)$$

kde

T_{Ia} ... je čas přípravy dat automatizovaně

T_{Im} ... je doba přípravy dat, která probíhá poloautomaticky nebo ručně (vstupní kontroly)

T_{Oa} ... je čas zpracování dat a výpočtu výsledků automatizovaně

T_{Om} ... je čas zpracování dat a výpočtu výsledků poloautomaticky nebo ručně (výstupní kontroly).

V rámci experimentálních výsledků disertační práce se podařilo empiricky stanovit hodnotu funkce v rovnici (8). Čas automatizované přípravy dat T_{Ia} je vyjádřen vztahem (28), přičemž doba poloautomatické a ruční přípravy dat T_{Im} je závislá na čase T_{Ia} (viz obr. 30). Čas automatizovaného zpracování dat T_{Oa} je dán rovnicí (29) a doba zpracování dat poloautomaticky nebo ručně T_{Om} je závislá na čase T_{Oa} (viz obr. 32).

3.2 Příprava dat

Hlavním cílem této fáze je upravit vstupní data tak, aby byla zajištěna:

1. řešitelnost
2. spolehlivost všech procesů, které budou v GIS probíhat.

Pokud jsou vstupní data příliš rozsáhlá (objemná), není možné výše uvedené požadavky zajistit. V tomto případě je zapotřebí datové sady rozdělit na menší části podle vhodně stanovených kritérií.

Nechť neprázdná množina $U \neq \emptyset$ je tzv. universum discursu a X je jeho podmnožinou ($X \subseteq U$). Množina U reprezentuje v našem případě celé řešené území a podmnožina X je část řešeného území. Relace ekvivalence R rozdělí množinu U do podmnožin:

$$U / R = \{X_1, X_2, \dots, X_n\} \quad (9)$$

tak, že pro každé i, j platí:

- 1) $X_i \subseteq U, X_i \neq \emptyset \dots$ všechny podmnožiny jsou neprázdné
- 2) $X_i \cap X_j \neq \emptyset \dots$ průnik všech podmnožin je prázdný
- 3) $\bigcup_{i=1,2,\dots,n} X_i = U \dots$ sjednocení všech podmnožin je právě celá množina U .

Za relaci ekvivalence R můžeme považovat územní rozdělení ČR podle administrativního uspořádání, která rozdělí toto území do nižších správních celků (kraje, okresy atd.). Pokud jsou splněny podmínky ad 1) – 3), má každé dílčí území z hlediska teorie množin charakter třídy X_i .

Předpokládejme, že existuje jiná relace $S \subseteq R$ taková, že definuje rozklad podmnožin $X_i \subseteq X$ do ekvivalentních tříd X_{ij} se stejnými vlastnostmi ad 1) – 3):

$$X_i / S = \{X_{i1}, X_{i2}, \dots, X_{im}\}, \quad (10)$$

pro všechna $i = 1, 2, \dots, n$.

Pak systém relací $\{R, S\}$ představuje spolu s množinami U, X hierarchický rozklad množiny U . V našem případě relace S definuje další rozdělení územních celků, reprezentovaných třídami X_i do nižších územních celků podle věcného principu. Kritérium tohoto rozdělení je skutečnost, zda v daném území jsou nebo nejsou relevantní geografické objekty, tj. prvky, které jsou hlavním předmětem zpracování v daném projektu GIS. Relace S tak rozdělí areály do podtříd X_{ij} pro $j = 1, 2, \dots, m$ podle vztahu (10). Hierarchický rozklad množiny $X \subseteq U$ do ekvivalentních tříd je výsledkem kombinace kritérií rozdělení podle:

- 1) administrativního uspořádání území v první hierarchické úrovni
- 2) věcného uspořádání, tj. podle existence geografických objektů v území ve druhé hierarchické úrovni.

Nechť X je množina geografických objektů v reálném světě a množina Y obraz množiny X v digitální geo-databázi. Pak zobrazení $\varphi: X \rightarrow Y$ musí mít tyto vlastnosti: (11)

- 1) φ je jedno-jednoznačné
- 2) φ je spojité
- 3) existuje inverzní zobrazení $\varphi^{-1}: Y \rightarrow X$, které je také spojité.

Pak zobrazení φ je tzv. homeomorfismus, který je základní vlastností topologických vztahů mezi geografickými objekty. V našem případě jsou geografické objekty (množina X) vstupní datové sady a množina Y pak obrazy těchto datových sad v GIS projektu.

Na základě koeficientu δ , definovaného v předchozí podkapitole, bude rozhodnuto o tom, podle jakých dílčích územních celků v rámci administrativního uspořádání ČR (relace R , tj. rozklad v 1. hierarchické úrovni) budou rozděleny vstupní datové sady projektu.

3.3 Algoritmus pro optimalizaci zpracování velkého množství geodat

Rozdělením vstupních dat do menších celků proces přípravy dat zcela nekončí. V další fázi musíme zajistit, aby celý projekt byl vyřešen v zadaném čase (termínu). Vhodnou metodou je optimalizace všech procesů, které budou v GIS následovat. Z hlediska efektivity je užitečné využít možného paralelismu v GIS a přidělit dílčí datové sady dostupným prostředkům (počítačům). Problém je řešen v této podkapitole.

Nechť P je množina n procesů $p_i \in P$, $i = 1, 2, \dots, n$ a M množina r prostředků $m_j \in M$, $j = 1, 2, \dots, r$. Definujme přiřazení \Leftrightarrow procesů prostředkům ($p_i \Leftrightarrow m_j$ znamená, že proces p_i je přiřazen prostředku m_j) jako zobrazení:

$$p_i \Leftrightarrow m_j \in S: P \times M \rightarrow S, p_i \in P \wedge m_j \in M, \quad (12)$$

kde S je uspořádaná množina kombinací $(p_i, m_j) \in S$, $i = 1, 2, \dots, n$, $j = 1, 2, \dots, r$. Přičemž platí:

$$\begin{aligned} &(\text{pro } \forall p_i \in P \exists m_j \in M / \text{je-li } (p_i, m_j) \wedge (p_k, m_l) \in S \Rightarrow \text{pro } \forall i, k \in \{1, 2, \dots, n\}, (i \neq k)) \wedge \\ &(\text{pro } \forall m_j \in M \exists p_i \in P / \text{je-li } (p_i, m_j) \wedge (p_i, m_k) \in S \Rightarrow \text{pro } \forall j, k \in \{1, 2, \dots, r\}, (j \neq k)) \end{aligned} \quad (13)$$

a

$$\text{card}(S) = \text{card}(P) = n. \quad (14)$$

Z rovnic (13) a (14) vyplývá, že všechny procesy $p_i \in P$ jsou obsaženy v kombinaci $(p_i, m_j) \in S$ právě jednou. Dále z těchto rovnic plyne, že prostředky $m_j \in M$ se v kombinaci $(p_i, m_j) \in S$ mohou opakovat nebo tam nemusí být obsaženy všechny (Pozn. Symbol \Rightarrow ve vztahu (13) vyjadřuje implikaci).

Dále definujeme dobu trvání procesu na daném prostředku jako zobrazení $S \rightarrow T$ s vlastností:

$$\text{pro } \forall s_i \in S \exists t_{ij} \in T / \text{je-li } s_i \in (p_i, m_j) \rightarrow p_i \Leftrightarrow m_j \text{ pro } \forall i \in \{1, 2, \dots, n\}. \quad (15)$$

Úlohou je sestavit optimální rozvrh přiřazení procesů prostředkům tak, aby celková doba řešení T_{tot} byla minimální. Účelová funkce, která se bude optimalizovat, je dána vztahem:

$$\sum_{i=1}^n t_{ij} \leq T_{tot}, \quad (16)$$

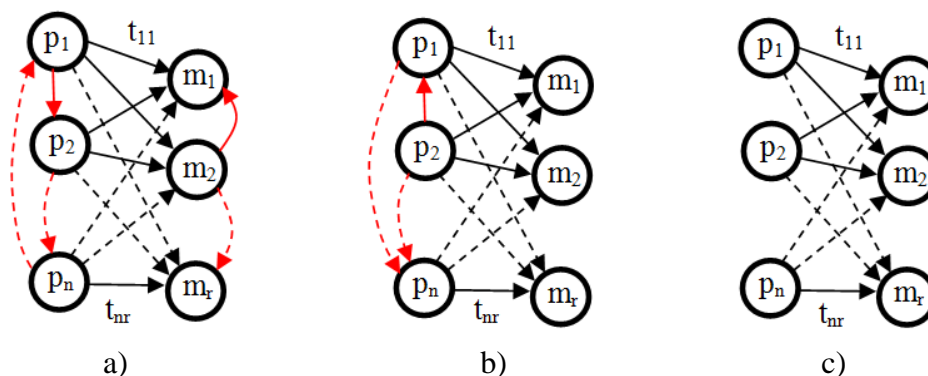
kde

t_{ij} je doba trvání procesu p_i na prostředku m_j

T_{tot} je celková doba řešení úlohy, která se skládá ze všech procesů $p_i \in P, i = 1, 2, \dots, n$.

Obtížnost optimalizace je dána stupněm závislosti mezi procesy a prostředky. Prakticky mohou nastat tyto varianty závislosti, které jsou znázorněné v grafu doby zpracování procesů prostředky na obr. 1 (hrany, reprezentující závislost, jsou v grafu označeny červeně):

- procesy i prostředky jsou navzájem závislé (obr. 1a)
- procesy jsou navzájem závislé, prostředky jsou nezávislé (obr. 1b)
- procesy i prostředky jsou navzájem nezávislé (obr. 1c).

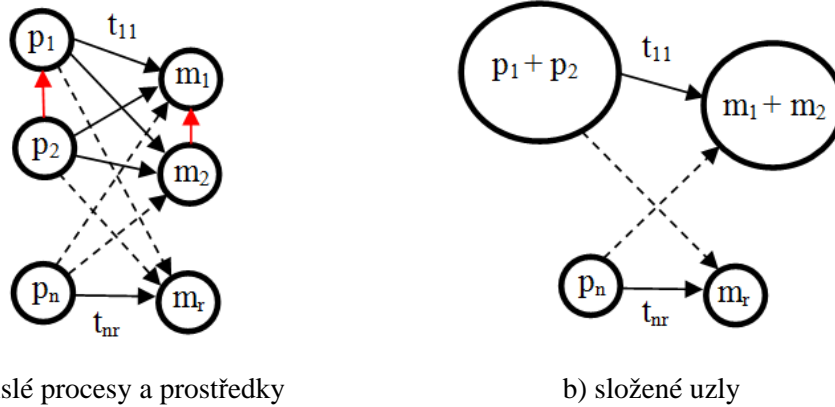


Obr. 1 – Graf doby zpracování procesů prostředky: a) procesy i prostředky jsou navzájem závislé, b) procesy jsou navzájem závislé, prostředky jsou nezávislé, c) procesy i prostředky jsou navzájem nezávislé (časy jsou v grafu pro přehlednost uvedeny jen u krajních hran)

Je zřejmé, že nejjednodušší varianta je ad c), nejsložitější je varianta ad a). Tento případ lze zjednodušit např. tím, že pokud jsou procesy nebo prostředky navzájem závislé, sloučí se všechny závislé procesy nebo prostředky do skupin (viz obr. 2) a na tyto skupiny se aplikuje optimalizační algoritmus. V rámci závislých skupin se optimalizace řeší sekvenčně (postupně). Algoritmus byl navržen především pro nezávislé skupiny i procesy, viz obr. 1 c). Tento případ se v systémech GIS vyskytuje častěji než v jiných oborech.

<i>Procesy/prostředky</i>	m_1	m_2	...	m_r
p_1	t_{11}	t_{12}	...	t_{1r}
p_2	
...	
p_n	t_{n1}	t_{n2}	...	t_{nr}

Tab. 2 – Mapa procesů



Obr. 2 – Sloučení závislých procesů a prostředků do složených uzlů

Na základě grafu na obr. 1 sestavíme mapu procesů M , jejíž řádky odpovídají procesům, sloupce prostředkům a prvky matice m_{ij} reprezentují dobu zpracování. Ukázka obecné mapy procesů je v tab. 2.

3.3.1 Popis optimalizačního algoritmu

Nejdříve definujeme procesy, které se v systému vyskytují, a určíme prostředky, na nichž je možné dané procesy zpracovávat. Potom je zapotřebí stanovit doby zpracování t_{ij} i -tého procesu p_i zpracovávaném na j -tém prostředku m_j – viz vztah (15). Tyto doby se zjistí buď kvalifikovaným odhadem, nebo experimentálně.

Dále se zjišťuje vzájemná závislost mezi procesy a prostředky. Tento vztah lze zjistit např. orientovaným grafem (viz obr. 1). Pokud mezi procesy nebo prostředky existují závislosti, lze tento problém vyřešit sloučením závislých uzlů tak, že mezi sloučenými a ostatními uzly už nejsou žádné závislosti (viz obr. 2). Pokud jsou procesy a prostředky navzájem nezávislé, můžeme sestavit mapu procesů a prostředků. Záhlaví řádků této mapy obsahuje jednotlivé procesy p_i , záhlaví sloupců prostředky m_j , do buněk pak zapíšeme doby trvání procesu t_{ij} na daném prostředku (viz tab. 2). V této mapě uspořádáme řádky podle součtu časů sestupně a sloupce podle součtů časů ve sloupcích vzestupně tak, aby platilo:

$$\sum_{j=1}^r t_{1j} \geq \sum_{j=1}^r t_{2j} \geq \dots \geq \sum_{j=1}^r t_{nj}, \sum_{i=1}^n t_{i1} \leq \sum_{i=1}^n t_{i2} \leq \dots \leq \sum_{i=1}^n t_{ir}. \quad (17)$$

Prakticky to znamená, že časově náročnější proces má v mapě nižší index řádku než proces, který trvá kratší dobu. Výkonnější prostředek pak má nižší index sloupce než prostředek méně výkonný. Smyslem tohoto uspořádání je, aby nejnáročnější proces byl řešen pokud možno na nejvýkonnějším prostředku. Na tomto principu je založen optimalizační algoritmus.

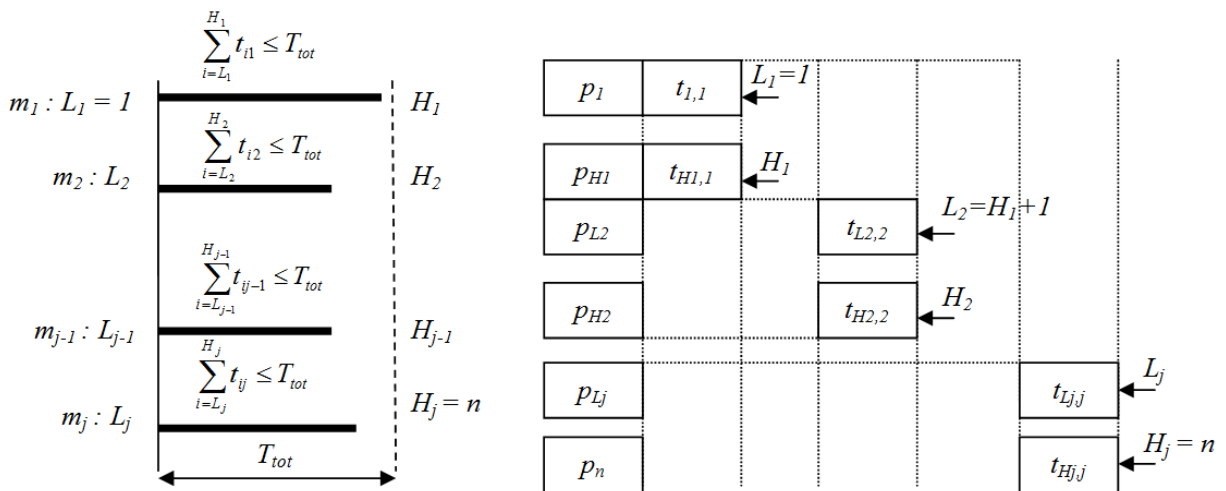
Po sestavení mapy ji musíme otestovat, zda je korektní. Může nastat případ, kdy některý proces p_i nepůjde řešit ani na jednom z prostředků, které jsou v mapě uvedeny. V tomto případě se do příslušného pole zapíše hodnota $t_{ij} = \infty$, tzv. strojového nekonečna, což je maximální číslo, které lze v počítači zobrazit. Před vlastním algoritmem musíme zkontrolovat, aby každý řádek obsahoval alespoň jeden údaj, který je menší než toto strojové

nekonečno. Pokud některý řádek obsahuje samá strojová nekonečna, musíme přidat do mapy takový prostředek, který daný proces vyřeší v konečném čase, tj. platí: $t_{ij} < \infty$.

Jádrem optimalizační procedury je dvojnásobný cyklus, kde vnější cyklus probíhá po řádcích (index i) a vnitřní cyklus po sloupcích (index j). V každém j -tém sloupci se stanoví dolní mez L_j a horní mez H_j . V těchto mezích se po řádcích v daném sloupci sčítají hodnoty t_{ij} i -tého procesu p_i a při každém součtu se testuje, zda součet mezi L_j a H_j nepřesahuje hodnotu celkové doby T_{tot} řešení úlohy. Pokud ne, pokračuje se v součtu dob t_{ij} v daném sloupci. Pokud ano, přejde se na další sloupec ($j = j + 1$), tj. musí se použít další prostředek m_j .

Nejdříve se testuje, zda index j nepřekročil hodnotu m_r , tj. počet prostředků, které máme k dispozici. Pokud ano, pak musíme přidat nový prostředek a zařadit jej do mapy procesů. Pak se spustí dvojitý cyklus znovu. Pokud ne, tak se nastaví dolní a horní mez L_j a H_j řádku v dalším j -tém sloupci a postupně se sčítají hodnoty t_{ij} mezi těmito mezemi, dokud součet nepřesáhne hodnotu T_{tot} . Dojdeme-li na konec řádků, tj. $L_j = n$, algoritmus končí a aktuální index j (číslo sloupce v matici procesů a prostředků) udává optimální počet prostředků, který je postačující pro vyřešení celé úlohy.

Protože předpokládáme, že procesy jsou na sobě navzájem nezávislé, můžeme je spouštět na určených prostředcích paralelně. Situace je znázorněna na obr. 3. Časová náročnost algoritmu je v nejhorším případě (tj. kdy index sloupce j dosáhne hodnoty r) úměrná součinu $n \times r$, tj. $O(nr)$.



Obr. 3 – Paralelní zpracování procesů na prostředcích

3.3.2 Optimalizační algoritmus v metajazyce

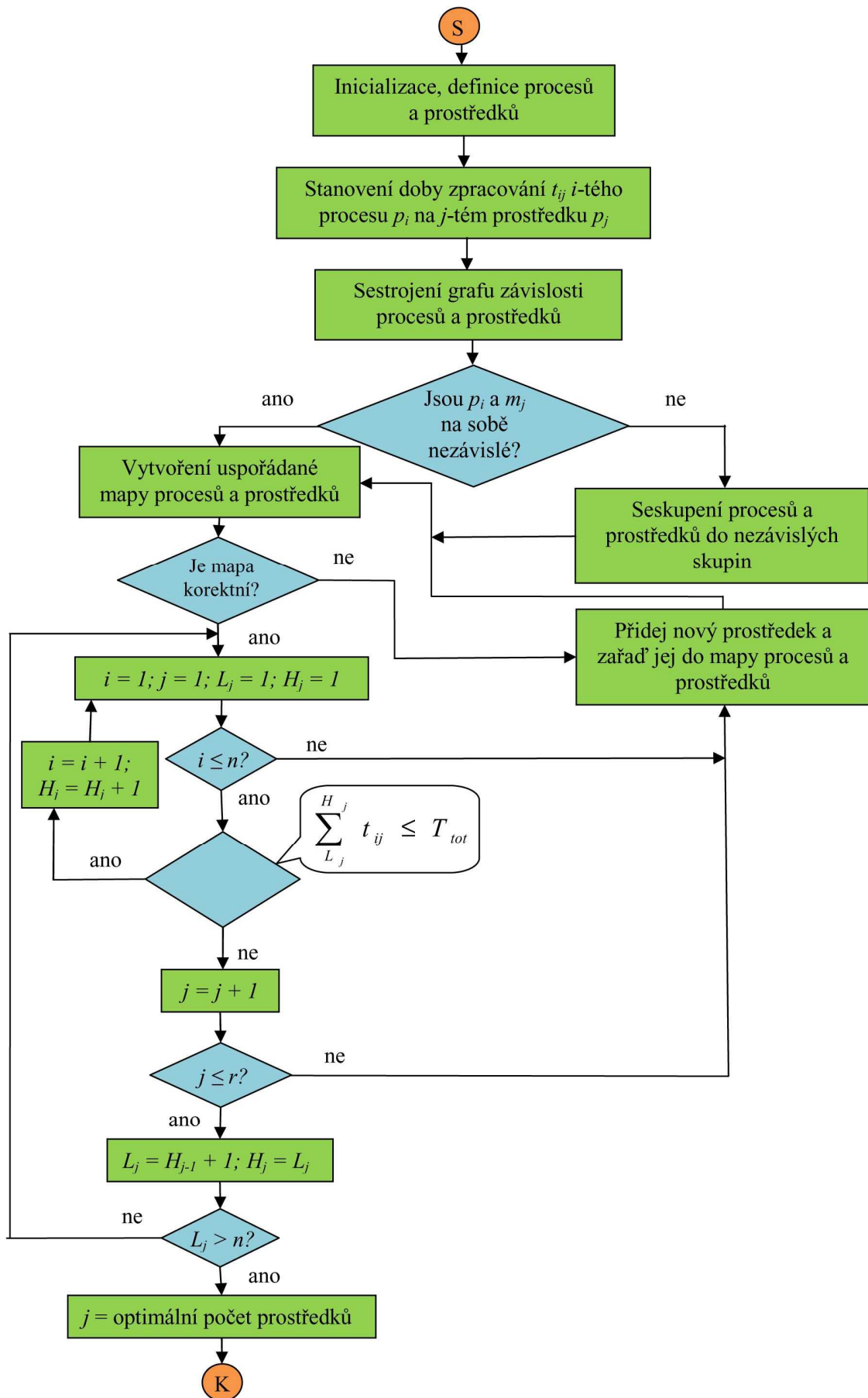
Následuje popis navrženého algoritmus v metajazyce, tj. v symbolické formě. Algoritmus pracuje v těchto krocích:

1. Inicializace, definice procesů $p_i \in P$ a prostředků $m_i \in M$
2. Stanovení doby zpracování t_{ij} i -tého procesu p_i zpracovávaném na j -tém prostředku m_j (experimentálně) – viz vztah (15)
3. Sestrojení grafu závislosti procesů a prostředků – viz obr. 1.

4. Test: Jsou p_i ($i = 1, 2, \dots, n$) a m_j ($j = 1, 2, \dots, r$) na sobě nezávislé?
- ANO: Vytvoření uspořádané mapy procesů a prostředků takto:
 - Uspořádání řádků mapy podle součtů v každém i -tém řádku sestupně

$$S_i = \sum_{j=1}^r t_{ij}$$
 - Uspořádání sloupců mapy podle součtů v každém j -tém sloupci vzestupně $S_j = \sum_{i=1}^n t_{ij}$
 - Pokračování krokem 5.
 - NE: Seskupení procesů a prostředků do nezávislých skupin. Tyto skupiny pak tvoří složený uzel v grafu procesů a prostředků – viz obr. 2. Návrat na krok 4a.
5. Test: Je mapa procesů korektní?
- ANO: Jdi na krok 6.
 - NE: Jdi na krok 6d.
6. Optimalizace přiřazení procesů prostředkům:
- $i = 1$ {nastavení indexu řádku}, $j = 1$ {nastavení indexu sloupce}
 - $L_j = 1$ {nastavení dolní meze j -tého sloupce}, $H_j = 1$ {nastavení horní meze j -tého sloupce}
 - Pokud ($i \leq n$), *tak*: {cyklus přes všechny řádky}
 - Test: ($\sum_{L_j}^{H_j} t_{ij} \leq T_{tot}$)?
 - ANO: {pokračuje se v součtu ve stejném sloupci}
 - $i = i + 1$ {inkrement indexu řádku i }
 - $H_j = H_j + 1$ {inkrement H_j }
 - Jdi na krok 6c.
 - NE: {přechod na další sloupec – pokud existuje}
 - $j = j + 1$ {inkrement indexu sloupce j }
 - Test: ($j > r$)? { r je počet prostředků}
 - ANO: Jdi na krok 6d.
 - NE:
 - $L_j = H_{j-1} + 1$ {inkrement dolní meze dalšího sloupce L_j }
 - $H_j = L_j$ {nastavení horní meze dalšího sloupce L_j }
 - Test: ($L_j > n$)?
 - ANO: Konec algoritmu.
 - NE: Jdi na krok 6c.
 - Přidej nový prostředek a zařaď jej do mapy procesů a prostředků. Jdi na krok 5.

Vývojový diagram algoritmu optimalizačního procesu je znázorněn na obr. 4. Algoritmus byl ověřen v rámci experimentálních výsledků disertační práce (viz podkapitola 5.1.3)



Obr. 4 – Algoritmus optimalizace přiřazení procesů prostředkům

3.4 Zpřesnění výsledků klasifikace obrazu

V této části je navržena metoda pro zlepšení kvality výsledků klasifikace a určení klíčových faktorů, které tuto kvalitu významně ovlivňují. K posouzení kvality výsledků klasifikace ortofota využijeme teorii hrubých množin. Příslušnost prvku k hrubé množině je zprostředkována pomocí speciální relace ekvivalence, tzv. relace nerozlišitelnosti. Původní definice hrubých množin je předložena v publikaci [85].

Mějme universum prvků U , množinu atributů A množiny hodnot V , kterých mohou atributy a_i z množiny A nabývat. Uvažujme funkci g , která určuje hodnoty atributů pro prvky universa U :

$$g: U \times A \rightarrow V. \quad (18)$$

V našem případě množina U reprezentuje ortofoto a množina A jednotlivé zóny ortofota, tj. geografické areály (les, půda, komunikace apod.).

Pomocí funkce g je zavedena relace nerozlišitelnosti $RE(A)$ takto:

$$x_1, x_2 \in U, (x_1 RE(A) x_2) \Leftrightarrow (g(x_1, a_i) = g(x_2, a_i)). \quad (19)$$

Na základě relace nerozlišitelnosti můžeme určit, které prvky universa U a s jakou kvalitativní jistotou patří do podmnožiny $X \subseteq U$, která je předmětem našeho zájmu. Tato relace v našem případě rozlišuje, které pixely ortofota patří do určité zóny.

Definujme dolní aproximace jako popis objektů, které s jistotou patří do podmnožiny X . Dolní aproximace se také nazývá pozitivní oblastí $Posi_{RE}(X)$.

$$Posi_{RE}(X) = \bigcup \{Y \mid (Y \in U / RE) \text{ AND } (Y \subseteq X)\}, \quad (20)$$

kde U/RE je faktorová množina zkonstruovaná na množině U podle relace $RE(A)$ a Y je prvek této faktorové množiny.

Definujme horní aproximace jako množinu prvků z U , které mohou patřit do podmnožiny X . Označuje se $Poss_{RE}(X)$ a je definována takto:

$$Poss_{RE}(X) = \bigcup \{Y \mid (Y \in U / RE) \text{ AND } (Y \cap X \neq \emptyset)\}. \quad (21)$$

Smysl zavedení obou definic spočívá v tom, že některé pixely ortofota můžeme s jistotou zařadit do určité zóny (jde o dolní aproximaci), zatímco existují i takové pixely, které nelze s jistotou zařadit do žádné zóny (horní aproximace).

Množinový rozdíl mezi horní ($Poss_{RE}(X)$) a dolní ($Posi_{RE}(X)$) aproximací X se nazývá hraniční množina $Bound_{RE}(X)$ a lze ji vyjádřit následovně:

$$Bound_{RE}(X) = Poss_{RE}(X) - Posi_{RE}(X) \quad (22)$$

Hrubá množina je podmnožina X universa U , která je definována pomocí horní a dolní aproximace ($Poss_{RE}(X)$, $Posi_{RE}(X)$) a pro kterou platí:

$$Bound_{RE}(X) = \emptyset \quad (23)$$

Praktický smysl této definice spočívá v tom, že každý pixel ortofota musíme s určitou jistotou zařadit do nějaké zóny (neexistují nezařazené pixely). Pomocí hrubé množiny můžeme definovat přibližnou přesnost $\alpha_{RE}(X)$, s jakou nalezená aproximace reprezentuje vybranou množinu X :

$$\alpha_{RE}(X) = \frac{card(Posi_{RE}(X))}{card(Poss_{RE}(X))} \quad (24)$$

Aproximace ve vztahu (24) je v našem případě měřítkem kvality klasifikace, tj. do jaké míry odpovídá zařazení daného pixelu do příslušné třídy. Tato přesnost je ovlivněna především těmito faktory:

- kvalitou vstupních dat (v našem případě ortofota)
- zvolenou metodou automatické klasifikace
- kvalitou trénovací množiny (pokud jde o řízenou klasifikaci)
- doplňkovými zpřesňujícími operacemi (filtrace, kontrola apod.).

Tyto faktory jsou podrobněji rozebrány v podkapitole 4.5.

3.5 Stanovení chybovosti analýzy

Tato podkapitola řeší vyhodnocení přesnosti výsledků klasifikace obrazu ortofota. V GIS je pojem přesnosti posuzován podle tří hlavních hledisek:

1. polohová přesnost geografických objektů
2. časová přesnost (souvisí s aktualizací dat)
3. tematická (atributová) přesnost.

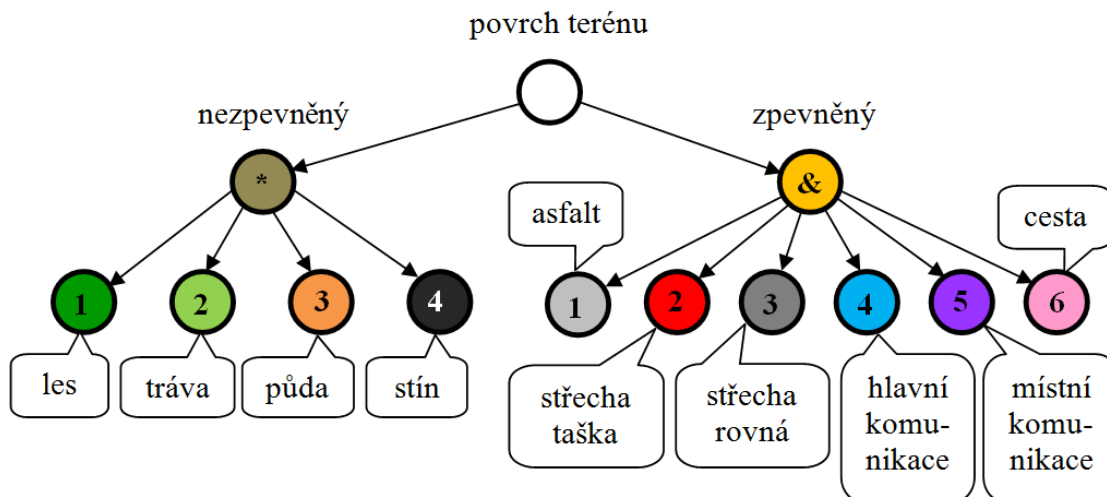
Polohová a časová přesnost se modeluje pomocí metrických prostorů [84]. Tyto přesnosti jsou dány použitím nejaktuálnějších dostupných dat od poskytovatele. V našem případě bude podrobněji řešena tematická přesnost. V této souvislosti lze definovat metrický prostor podle definice (9) a s ní souvisejících podmínek 1) – 3).

Zde za relaci ekvivalence R můžeme považovat trénovací množinu, která má n vzorků reprezentující různé typy zemského povrchu. Pokud jsou splněny podmínky ad 1) – 3), má každé dílčí území z hlediska teorie množin charakter třídy X_i . Každé třídě pak můžeme přiřadit název podle druhu povrchu, který reprezentuje (např. les, půda, asphalt apod.). Na množině X pak definujeme tematickou (atributovou) metriku d_a takto:

$$\begin{aligned} d_a(x_i, x_j) &= 0, \text{ jsou-li } x_i, x_j \text{ ze stejné třídy} \\ d_a(x_i, x_j) &= k, \end{aligned} \quad (25)$$

kde $k \in N$, jsou-li x_i, x_j z různé třídy, N je množina přirozených čísel.

Prvek $x_i \in X_i$ můžeme v našem případě považovat za pixel ortofota. Vzdálenost d_a mezi pixely, které patří do různých tříd, pak vypočítáme pomocí klasifikačního stromu. Strom má 1 kořen, který se hierarchicky větví, přičemž počet hierarchických stupňů není omezen. Listy stromu reprezentují jednotlivé klasifikační třídy X_i . Strom si sestavuje uživatel účelově podle charakteru úlohy, kterou řeší. V našem případě je strom znázorněn na obr. 5.



Obr. 5 – Klasifikační strom

Vnitřní vrcholy stromu, které se dále větví, mají přiřazen zvláštní symbol (*, &, ...). Listy stromu jsou očíslovány přirozenými čísly v pořadí hran, které vycházejí ze stejného uzlu. Na základě klasifikačního stromu můžeme pro každou třídu odvodit jednoznačný řetězcový kód, s jehož pomocí je možné klasifikační strom rekonstruovat. Např. pro třídu *půda* je řetězcový kód *3, pro třídu *cesta* kód &6. Atributová metrika je v rámci podstromu rovna 1. Například v obr. 5 mají třídy *les*, *tráva*, *půda*, *stín* vzájemnou vzdálenost 1. Vzdálenost tříd přes n vnitřních uzlů je rovna n , např. *půda-asfalt*, nebo *tráva-střecha rovná*, mají vzdálenost 2. Smyslem zavedení tematické metriky spočívá v tom, že pokud identifikujeme např. místo *trávy* *les*, jde o menší chybu, než kdybychom místo *trávy* identifikovali např. *asfalt*. Tato metrika byla použita pro analýzu chybovosti klasifikace – viz podkapitola 4.6 a 5.6.3.

4 METODIKA ŘEŠENÍ

Teoretická východiska disertační práce byla ověřena na projektu zabývajícím se klasifikací údajů o uložení plynárenských zařízení pod určitými typy povrchů terénu. Tento projekt je specifický svým územním rozsahem, objemem zpracovávaných dat, relativně krátkou dobou řešení a snahou o co největší přesnost výstupu. Rozsah zpracovávaného území pokrýval téměř celé území našeho státu. Objem zpracovávaných dat se pohyboval ve stovkách GB. Doba řešení byla stanovena na několik týdnů. Přesnost výsledku byla očekávána v řádech maximálně několika jednotek metru. Celý tento proces bylo potřeba co nejvíce automatizovat. Tyto charakteristiky řadí projekt mezi masivní úlohy se závazným termínem dokončení. Vzhledem k těmto vlastnostem bylo nutné navrhnout a ověřit metodiku řešení masivních úloh v GIS.

4.1 *Specifikace projektu*

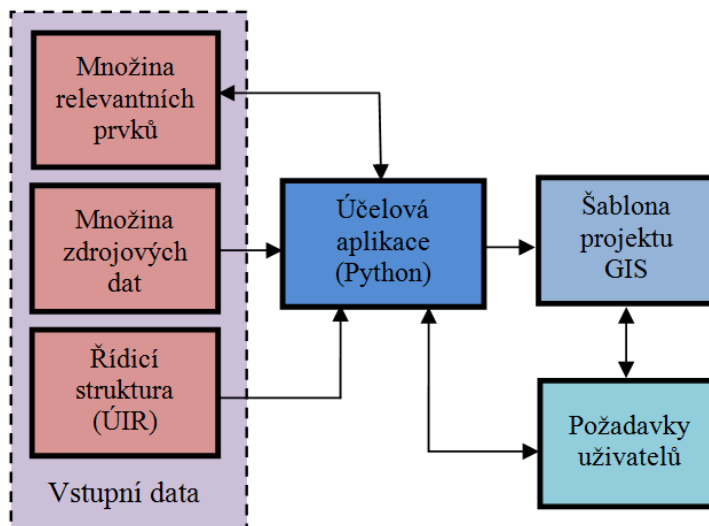
Cílem projektu bylo klasifikování druhů povrchů terénu nad průběhem trasy plynového vedení a jeho závěrečné statistické vyhodnocení. Tyto údaje jsou nyní využívány za účelem odhadu nákladů, které je nutné vynaložit na vybudování nových plynárenských sítí pod předmětnými povrchy. Plynářská společnost potřebovala znát zejména základní rozdělení povrchu terénu na zpevněný a nezpevněný povrch, dále pak co možná nejpodrobnější rozdělení povrchů včetně určení zásahu trasy plynovodu do hlavní a místní komunikace. Všechna data měla obsahovat jednak informaci o druhu povrchu terénu, ale i příslušnost k územní jednotce části obce.

Při prvotní fázi zkoumání vhodných dat pro analýzu druhů povrchů musel být brán zřetel na rozsah zpracovávaného území, které pokrývá téměř celé území státu. Proto bylo potřeba uvažovat za vhodná taková data, která se vyskytují po celé ČR. Z tohoto důvodu bylo vybráno za hlavní zdrojovou datovou sadu pro analýzu ortofoto, z kterého se dá pomocí klasifikace obrazu získat informace o typu povrchu. Po důkladnější analýze bylo shledáno ortofoto za jediný zdroj dat jako nedostačující, zejména proto, že z něj nelze spolehlivě získat informaci o hlavní a místní komunikaci. Proto byl za druhou zdrojovou datovou sadu vybrán ZABAGED, a to konkrétně liniový zakres komunikací, který zpřesňuje výsledky klasifikace obrazu nad ortofotem a navíc poskytuje důležitou informaci o druhu komunikace (více o vhodnosti dat v podkapitole 5.1.1). Za základní metodu řešení lze tedy považovat klasifikaci obrazu ortofota s rozlišením 0,25 m/pixel, která je následně zpřesněna vybranými vrstvami ZABAGED. Celý proces je řešen pomocí řady funkčních nástrojů v GIS tvořící analytický celek.

4.2 *Příprava dat*

Příprava dat hraje velmi důležitou roli v řešení masivních úloh. Zdrojová data mohou být různých datových typů (vektor, rastr, tabulková databáze) a je nutné je upravit před hlavním zpracováním. Tím se rozumí odstranit topologické chyby a jiné nedostatky, ale také vhodně rozdělit velké množství dat na menší segmenty pro zpracování. Proto je hlavním cílem

kvalitní přípravy datové struktury projektu rozdělení a uspořádání datových sad do vhodných menších celků a jejich uložení do vhodné projektové struktury prostředí GIS. Na obr. 6 je znázorněno obecné schéma přípravy dat. V našem případě bylo jádrem řešení vytvoření účelové softwarové aplikace ve skriptovacím jazyku Python s podporou knihoven ArcGIS ESRI.



Obr. 6 – Schéma předzpracování vstupních dat projektu

Tato softwarová aplikace třídí a vybírá zdrojová data do speciální datové struktury šablony projektu, která je předem vytvořena v prostředí ArcGIS. Aplikace tak realizuje zobrazení ze vztahu (11) $\varphi: X \rightarrow Y$, kde množina X reprezentuje zdroj vstupních dat a množina Y digitální geodatabázi v rámci šablony projektu. Zdrojová data softwarová aplikace rozděluje na základě řídicí struktury – územně identifikačního registru (ÚIR). Tento proces reprezentuje zobrazení R (relaci ekvivalence), popsané v podkapitole 3.2, tzv. územní hledisko rozkladu prvků do tříd. V rozsáhlých projektech mnohdy nastanou případy, kdy toto rozdělení je vhodné upravit vypuštěním redundantních dat. Např. optimalizaci pokrytí území ortofotem je výhodné realizovat pouze na datově společných (vzájemně překrytých) oblastech. Proto je účelné dané dílčí územní části získané rozkladem podle relace R (ÚIR) rozložit dále podle tematického hlediska. Dostáváme tak hierarchický rozklad vstupní množiny datových sad, kde kritérium druhého stupně rozkladu je množina relevantních prvků, která reprezentuje zobrazení S – viz podkapitola 3.2.

Pro korektní automatizované zpracování je nutné zajistit, aby zobrazení φ realizované softwarovou aplikací v Pythonu bylo homeomorfismus. To znamená, že každý řádek v tabulce ÚIR musí reprezentovat právě jediný územní celek (areál) a naopak. Předzpracování dat tedy probíhá tak, že nejdříve se ze zdrojových datových sad vyberou data relevantních prvků podle územního hlediska (relace ekvivalence R) a následně se vybírají další data podle kritérií relevantních prvků (relace ekvivalence S). Vývojový diagram algoritmu předzpracování je na obr. 29.

4.3 Výkonnost a řešitelnost složité úlohy

Výkonnost technologie a řešitelnost složité úlohy v GIS je ovlivněna především těmito faktory:

- parametry technického vybavení – hardware (HW)
- výpočetní mohutnost vybraného programového vybavení – software (SW)
- zvolená metodika zpracování celé úlohy
- organizace celého zpracovatelského procesu a podmínky, v nichž probíhá řešení (úroveň informační infrastruktury)
- počet geografických objektů ve vstupních datech a jejich geometrická a topologická složitost.

Míra vlivu faktorů na dílčí doby zpracování podle níže uvedeného vztahu (26) je uvedena v tabulce 3. Tato míra vlivu je ohodnocena třemi stupni: nízký, střední a vysoký vliv.

Faktory/doba	HW	SW	Metodika	Organizace	Počet GO
$T_{rež}$	střední	střední	střední	střední	střední
T_{pa}	vysoký	vysoký	nízký	nízký	vysoký
T_{pm}	střední	střední	střední	střední	střední

Tab. 3 – Vliv faktorů na dobu řešení úloh s enormním objemem dat

Celková doba řešení úlohy je dána vztahem:

$$T_c = T_{rež} + T_{pa} + T_{pm}, \quad (26)$$

kde $T_{rež}$... režijní čas (příprava dat)

T_{pa} ... doba automatizovaného zpracování

T_{pm} ... doba manuálního zpracování.

Dalším významným faktorem, který sice zásadně neovlivňuje výkonnost, ale souvisí s řešitelností úlohy, je spolehlivost. Dlouhá doba řešení T_c zvyšuje riziko nespolehlivosti a má značné nároky na parametry HW a SW. Podle tabulky 3 lze tuto dobu zkrátit kvalitním HW, SW a snížením počtu geografických objektů. Protože HW a SW mají svoje technická omezení, jeví se jako nejvhodnější krok, vedoucí ke zkrácení celkové doby řešení a současně zvýšení spolehlivosti procesů, rozdělit vstupní data do menších částí. Z hlediska efektivnosti je zapotřebí, aby doba režie nepřesáhla dobu zpracování, tj. aby poměr doby processingu $T_p = T_{pa} + T_{pm}$ k době režie byl větší než 1:

$$\eta = \frac{\sum_{i=1}^n T_{p_i}}{\sum_{i=1}^n T_{rež_i}} > 1, \quad (27)$$

kde $T_{rež_i}$... čas režie

T_{p_i} ... doba zpracování dílčího bloku vstupních dat

n ... počet bloků, na které je objem vstupních dat rozdělen.

Obě doby závisí na objemu zpracovávaných dat podle empiricky určeného vztahu:

$$T_{rež} = 0,002x^2 + 0,003x + 0,04, \quad (28)$$

$$T_p = 0,02e^{0,31x}, \quad (29)$$

kde x je objem zpracovávaných dat.

Z rovnic (28) a (29) je zřejmé, že režijní čas roste s počtem bloků, zatímco čas zpracování je na počtu bloků relativně nezávislý. Tyto vztahy byly určeny v rámci experimentálních výsledků disertační práce (viz obr. 30 a obr. 32). Dosazením těchto rovnic do vztahu (27) dostaneme:

$$\eta = \frac{\sum_{i=1}^n T_{p_i}}{\sum_{i=1}^n T_{rež_i}} = \frac{\sum_{i=1}^n 0,02e^{0,31x_i}}{\sum_{i=1}^n (0,002x_i^2 + 0,003x_i + 0,04)} \quad (30)$$

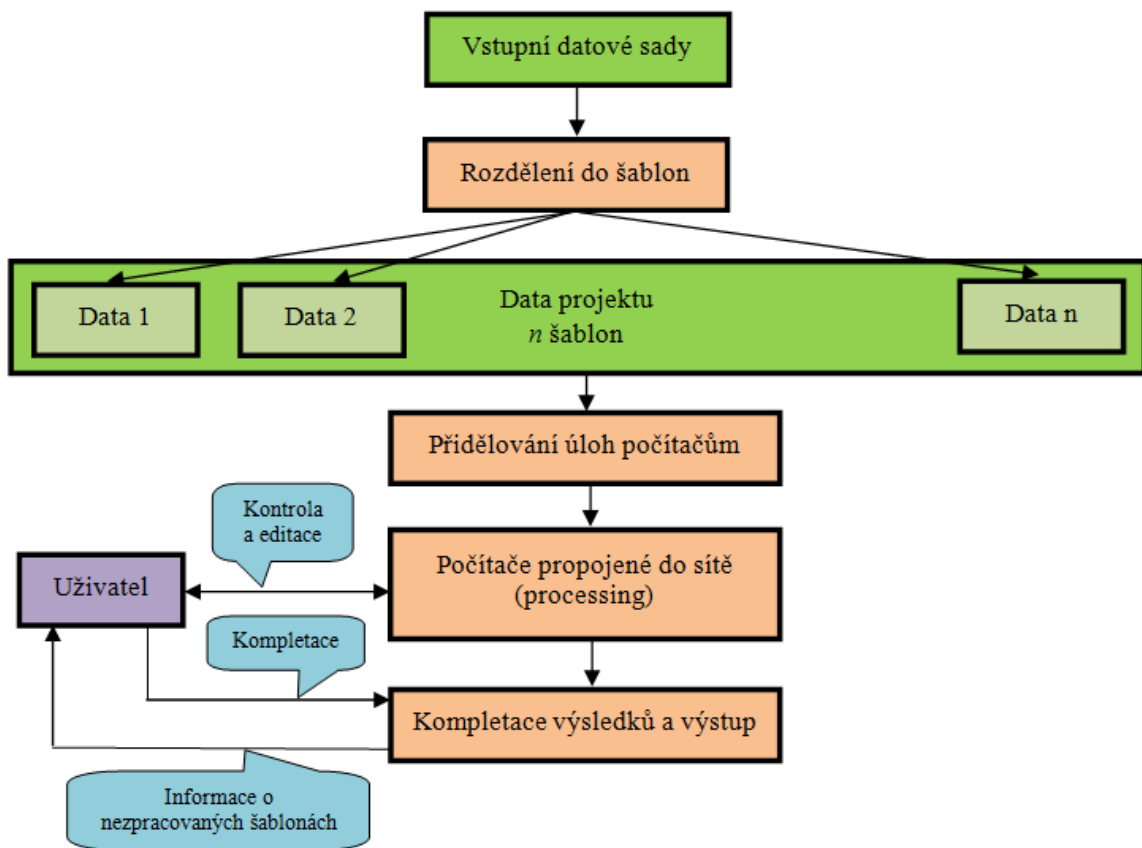
Ze vztahu je patrné, že závislost doby zpracování na objemu dat je exponenciální, zatímco závislost doby režie je kvadratická. Největší poměr η tedy dostaneme při zpracování celého objemu dat, ale v tomto případě hrozí největší riziko selhání spolehlivosti. Každý výpadek během dlouhodobého zpracování přináší velké časové ztráty a vede k výraznému snížení výkonnosti.

4.4 Navržená technologie zpracování velkého množství geografických dat

Realizovaná technologie zpracování značného objemu dat projektu, podrobněji popsaného v experimentálních výsledcích práce, je založena na poloautomatickém principu. Blokové schéma postupu zpracování je na obr. 7. Podstatou je rozdělení dat ze vstupních datových sad do samostatných bloků, tzv. šablon, podle územního a tematického principu. Územní rozdělení je provedeno podle hranic administrativního členění ČR, podkladem pro tematické rozdělení dat je zakres průběhu plynovodu.

Toto rozdělení vstupních dat se provádí automatizovaně podle účelové procedury v jazyku Python s podporou knihoven ESRI. V této fázi je nejprve nutné stanovit optimální územní jednotku zpracování jako prvek šablony projektu (viz *Data i* na obr. 7). Datové rozdělení podle územního principu je na základě koeficientu δ (viz vztah 3). Velikost jednotlivých bloků dat je diskutována v kapitole 5.1.2.

Tímto způsobem se data celého projektu rozdělí do dílčích šablon, které se pak zpracovávají nezávisle v počítačích propojených do sítě. Přidělování dílčích úloh jednotlivým počítačům je vyřešeno algoritmem, který je blíže popsán v podkapitole 3.3. Díky němu se určí počet nutných prostředků (počítačů) k vyřešení úlohy ve stanovené lhůtě a k těmto prostředkům se přidělí jednotlivé dílčí úlohy ke zpracování (tzv. processing šablon).



Obr. 7 – Harmonogram polo-automatických procesů ve velkém GIS projektu

Hlavním procesem je klasifikace obrazu ortofota pro účel určení druhu terénního povrchu v místě plynárenských zařízení s následnou filtrací vybranými vrstvami vektorové datové sady ZABAGED. Tento proces probíhal zcela automatizovaně v účelové proceduře vytvořené v jazyku Python. Po skončení tohoto procesu následuje vizuální kontrola uživatelem. Na závěr se dílčí výsledky z jednotlivých datových projektů kompletují do společného výstupu - filtrovatelné kontingenční tabulky. Tento proces probíhá automatizovaně opět v proceduře v jazyku Python.

4.5 Zpřesnění klasifikace obrazu

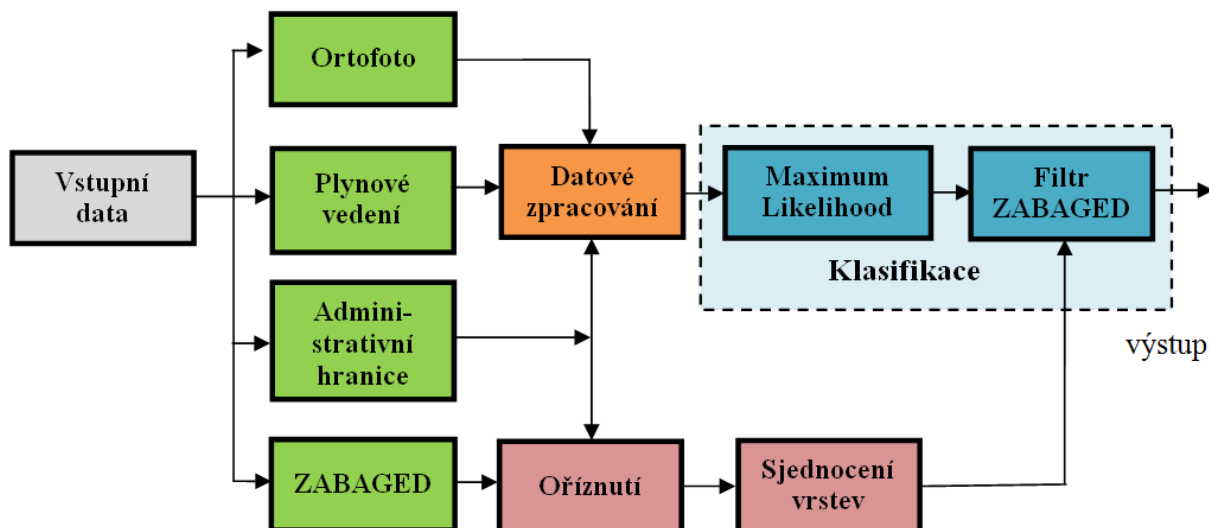
Zjednodušený celkový proces klasifikace povrchů je znázorněn na blokovém schématu na obr. 8. Vstupní datové sady projektu byly uloženy na serveru a obsahovaly:

- ortofoto snímky ČR
- průběh plynovodu
- průběh os vybraných typů komunikací ze ZABAGED
- průběhy hranic území podle administrativního členění státu.

V rámci přípravy dat se ze vstupních dat zkopírovaly datové sady do šablony projektu, která byla členěná podle územního principu. Vlastní zpracování dat probíhalo v prostředí ArcGIS. Základní metodou řešení datové analýzy v projektu byla klasifikace obrazu ortofota s

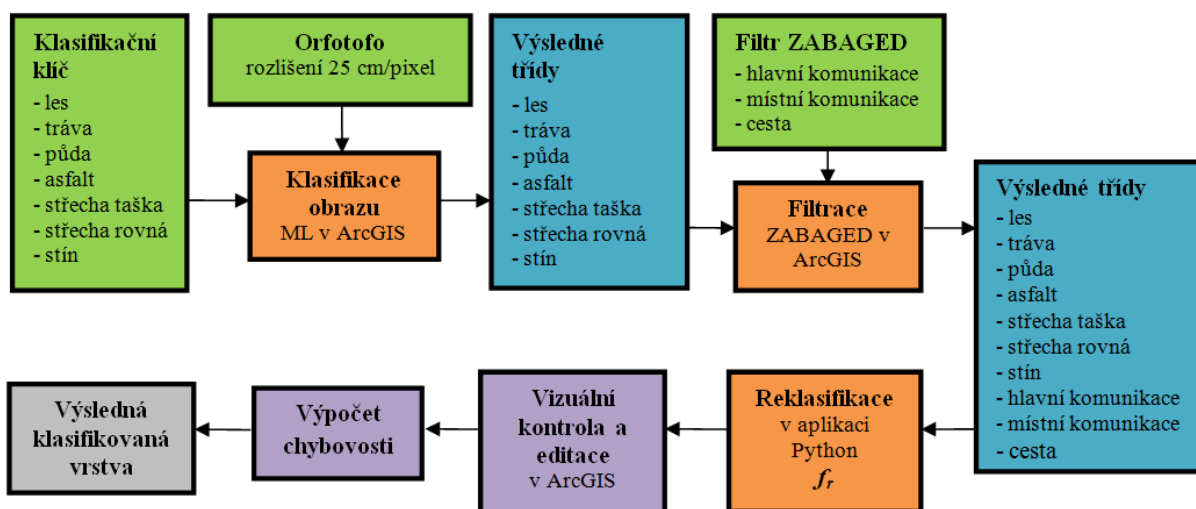
rozlišením 0,25 m/pixel, která probíhala v několika fázích. V těchto fázích docházelo k postupnému zpřesňování výsledků, viz obr. 9:

1. automatická klasifikace metodou Maximum Likelihood (ML) v ArcGIS
2. automatická filtrace výsledku vybranými vrstvami ZABAGED v ArcGIS
3. automatická reklasifikace pomocí účelové aplikace v jazyku Python
4. manuální kontrola výsledků a výpočet chybovosti.



Obr. 8 – Blokové schéma klasifikace povrchů

Výsledkem klasifikace obrazu metodou Maximum Likelihood bylo 7 tříd povrchů. Protože průběh plynovodu v mnoha případech zasahuje do komunikací, byl jako další krok zpřesnění výsledků zvolen filtr ZABAGED s vrstvami komunikací. Filtr byl realizován obalovou zónou kolem os komunikací. Šířky obalových zón byly určeny na základě přímého měření vybraných vzorků komunikací v každé projektové datové územní jednotce v ArcGIS. Filtrací byly rozšířeny třídy z klasifikace obrazu o třídy žádaných komunikací (hlavní a místní) a nezanedbatelným přínosem bylo také zpětné zpřesnění třídy zpevněného povrchu, konkrétně povrchu typu asphalt, protože hlavní i místní komunikace jsou tvořeny právě tímto povrchem.



Obr. 9 – Postupné zpřesňování výsledků klasifikace obrazu ortofota

Klasifikace povrchů pak probíhala v místě nad plynárenskými zařízeními tří typů:

- hlavní řad (územní oblast intravilán a extravilán)
- přípojky (územní oblast intravilán)
- vysokotlak (územní oblast extravilán).

Každý typ těchto zařízení se vyznačuje tím, že je převážně uložen pod určitou třídou povrchu. Např. vysokotlak probíhá v extravilánu, kde je obvykle nezpevněný povrch, přípojky jsou v intravilánu většinou uloženy pod zpevněným povrchem apod. Této skutečnosti bylo využito k dalšímu zpřesnění výsledků klasifikace pomocí speciální účelové aplikace v jazyce Python. Další mírné zpřesnění výsledků klasifikace bylo dosaženo až závěrečnou vizuální kontrolou s následnou ruční editací v prostředí ArcGIS.

4.6 Stanovení chybovosti analýzy

Součástí závěrečných fází projektu byla vizuální kontrola výsledků v mapovém okně ArcGIS a stanovení absolutní chybovosti klasifikace. Chybovost určíme na základě poznatků z podkapitoly 3.5 a s využitím následujících úvah vyplývajících z obrázku 9.

Výsledek klasifikačního procesu může být vyjádřen vztahem:

$$C = f_r(A \cup B_1 \cup B_2 \cap Z) \quad (31)$$

kde C ... je množina prvků ve výsledné vrstvě klasifikace

A ... množina prvků s nezpevněným povrchem

B_1 ... množina prvků zpevněného povrchu bez komunikací

B_2 ... množina prvků v komunikacích

Z ... množina prvků ZABAGED

f_r ... funkce reprezentující reklasifikaci podle typů plynárenských zařízení.

Prvkem zde rozumíme rozlišitelnou jednotku, např. pixel ortofota (v rastrové vrstvě) nebo jednotkovou délku plynového vedení (ve vektorové vrstvě). Chybovost pak vypočítáme podle vztahu:

$$e = \frac{\sum_{i=1}^k n_i d_a(c_i, c_j)}{\text{card}(C)} \quad (32)$$

kde n_i ... počet chybně zařazených prvků $c_i \in C$ (po vizuální kontrole byl místo c_i jako správný určen prvek $c_j \in C$)

d_a ... atributová vzdálenost prvků c_i, c_j (viz podkapitola 3.5)

k ... počet segmentů (skupin prvků), kde byla nalezena chyba

$\text{card}(C)$... počet prvků množiny C – viz vztah (31).

5 EXPERIMENTÁLNÍ VÝSLEDKY

Experimentální výsledky disertační práce souvisí s výše zmíněným projektem klasifikování druhů povrchů terénu nad průběhem trasy plynového vedení. Tento projekt byl řešen pro společnost GasNet, s.r.o., která je součástí koncernu RWE ČR. Projekt lze považovat za masivní úlohu řešenou v prostředí GIS z důvodu následujících charakteristik:

- Rozsah zpracovávaného území byl o rozloze celé ČR, vyjma dvou krajů – rozloha území byla 68312 km²
- Objem zdrojových dat byl přibližně 500 GB, objem zpracovávaných dat okolo 1 TB
- Přesnost výstupu byla očekávána okolo několik jednotek metru, maximální chybovost byla stanovena do 5 %
- Analýza celého území byla stanovena na 16 týdnů.

Hlavními cíly projektu bylo:

- a) vytvoření klasifikace/statistiky údajů o uložení plynárenských zařízení pod určitými typy povrchů terénu v určených lokalitách tak, aby se tyto údaje mohly následně využít za účelem stanovení reprodukčních hodnot plynárenských zařízení a odhadu nákladů, které by bylo nutné vynaložit na vybudování nových plynárenských sítí pod předmětnými povrchy
- b) vytvoření a ověření použitelnosti elektronického programu/počítačového postupu, jež by v budoucnosti umožnil automatizovaný přístup k vytvoření klasifikace podle bodu a) na územní regiony.

Analýza povrchů měla být provedena ve třech variantách pro průběh plynovodu typu vysokotlak, hlavní řad a přípojky. Byly požadovány 3 klasifikační kategorie výstupů pro každý typ plynovodu:

- A. 2 třídy druhu povrchu: *zpevněný* a *nezpevněný*
- B. 5 tříd druhu povrchu: *asfalt*, *hlavní komunikace*, *místní komunikace*, *nevázaný* (štěrk, zelená plocha a další nezpevněné povrchy), *neznámý*
- C. 10 tříd druhu povrchu: *les*, *travní porost*, *holá půda*, *asfalt*, *střecha – taška*, *střecha – rovná*, *stín*, *hlavní komunikace*, *místní komunikace*, *cesta*. To znamená analýza úseků plynovodu pod všemi typy povrchů, které jsou z datových sad klasifikovatelné.

Výstupem měla být strukturovaná tabulka ve formátu XLS s klasifikovanými údaji typů povrchů nad průběhem trasy plynovodu a grafický výstup strukturovaných klasifikovaných údajů ve formátu SHP pro vysokotlak, hlavní řad a přípojky.

5.1 Datová analýza povrchů terénu

Řešitelnost úlohy byla nejprve otestována v rámci pilotního projektu [86]. Výsledky projektu stanovily výběr vhodných datových sad pro zpracování úlohy, dílčí zpracovatelskou

jednotku, vhodný software a hardware, a nakonec byl zkonkretizován kompletní postup jednotlivých fází řešení úlohy.

5.1.1 Volba vhodných datových sad

Vzhledem k zadání a předpokládanému cíli projektu (automatická klasifikace povrchů na rozsáhlém území ČR) bylo zapotřebí, aby použité vstupní datové sady pokrývaly pokud možno celé zájmové území. Proto bylo jako první testováno pro řešení úlohy **ortofoto**. V průběhu řešení se ukázalo, že klasifikace obrazu na základě ortofota dat je nedostatečná a nevyhovující pro daný účel. Problém klasifikace obrazu nastává především v místech městské zástavby a všude tam, kde je povrch zastíněn budovami, stromy či jinými objekty. Tento případ je velmi častý především v místě městských komunikací. Proto byla řešena možnost zpřesnění výsledku klasifikace dalšími dostupnými vstupními daty se zákresem komunikací, a to s využitím datových sad **ZABAGED**. Jedná se konkrétně o liniové vektorové soubory zákresu dálnic, silnic, ulic a cest pro modelované lokality. Výhodou těchto dat je, že pokrývají celé území ČR a dále, že obsahují některé důležité atributy jako např. druh komunikace. To bylo důležité pro další zpřesnění klasifikace ve smyslu rozlišení komunikací např. na typy silnice 1. – 3. třídy a komunikace místní. Nevýhodou je to, že zakres obsahuje jen linie reprezentující osy zmíněných komunikací bez atributu šířky. Tento údaj je nezbytný pro prostorové analýzy s obalovými zónami (buffer) s následným překrytím. Právě v místech překrytí je možné použít reklasifikaci, a tím přesněji identifikovat druh povrchu. Hlavním problémem v této fázi bylo stanovení velikosti bufferu, který reprezentuje šířku dané komunikace. Optimální hodnota bufferu byla určena sadou nezávislých měření v obrazu ortofota v různých lokalitách a jejich následným zprůměrováním. Tyto hodnoty byly u tříd hlavních komunikací srovnány s parametry uváděnými v ČSN 736101 – Projektování silnic a dálnic. Hodnoty bufferů odpovídaly těmto parametrům jen v některých případech. Proto bylo rozhodnuto, že se velikosti šířek hlavních komunikací budou odměřovat ručně pro každou zpracovatelskou jednotku, stejně tak jako to bylo nutné u komunikací místních.

Uvažovalo se i o možném zpřesnění klasifikace využitím jiných dostupných datových sad, např. katastrální mapou, datovou sadou silniční databanky či ortofotem s pásmem blízkého infra záření. Nevýhodou ovšem je, že při využití katastrálních map by se do řešení vnášely chyby identifikace druhu pozemku, které nastávají v případě, kde není stav katastrální mapy v souladu se skutečným stavem v terénu, např. sloučení nádvoří a budovy, problém vyznačení parcel u některých komunikací (komunikace a nepevněný zelený pás sloučeny do jedné parcely), zakres vnitřní kresby, aj. Data silniční databanky neposkytují pro řešení více potřebných údajů, než ZABAGED. Využití ortofoto s pásmem blízkého infra naráželo na problém obtížnější dostupnosti a značného zpoplatnění dat pro celé řešené území. Navíc zvyšování počtu vstupních podkladů (datových sad) vede k výraznému zvýšení doby a složitosti řešení. Proto bylo snahou optimalizovat počet vstupních běžně dostupných datových sad na takový nezbytně nutný počet, který zajišťuje přijatelnou kvalitu a vypovídací schopnost výsledků i v relativně rozsáhlém území.

5.1.2 Základní zpracovatelská jednotka

Rozdělení vstupních dat v GIS musí být nejen v souladu s rovnicí (30), ale také by mělo logicky respektovat územní a tematické členění. Podle pilotního projektu bylo na základě koeficientu δ (viz vztah 3) stanoveno optimální rozdělení dat na datové jednotky projektu o objemu $\sim 2,2$ GB, které odpovídají rozsahu území o rozloze ~ 368 km². Taková rozloha odpovídá administrativnímu členění ČR na tzv. obce s rozšířenou působností (ORP). Zpracovávané území analýzy pokrývalo celkem 188 ORP. Optimální poměr η pro celkový objem dat $x = 500$ GB (viz vzorec 30) vychází pro tyto hodnoty: $n = 200$, $x_i = 2,5$ GB, což odpovídá zvolené datové projektové jednotce.

Pro území s průměrnou rozlohou ORP je hodnota $\delta \approx 6 \cdot 10^9$. Tato hodnota koeficientu orientačně určuje řešitelnost úlohy vzhledem ke spolehlivosti zpracování a k době řešení. Použití menší územní jednotky vede k nadměrnému zvýšení cyklů dílčích procesů (zvyšuje se režijní čas procesu), naopak použití větší územní jednotky značně zvyšuje časovou a prostorovou náročnost operací a výrazně snižuje technickou spolehlivost zpracování s ohledem na limity výpočetní techniky (např. hodnota koeficientu pro celé řešené území projektu vychází na $\delta \approx 1,1 \cdot 10^{12}$).

Každá zpracovatelská jednotka ORP byla tvořena stejnou adresářovou strukturou. Předlohou této struktury byla prázdná šablona projektu ORP, která byla nakopírována na příslušné místo na disku a následně byla naplněna zdrojovými daty (viz podkapitola 5.2.2).

5.1.3 Výběr vhodného software a hardware

Vhodný software

Vzhledem k tomu, že plynárenská zařízení jsou geografické objekty s územně vázanými informacemi, nestačila k úspěšnému řešení jen prostá klasifikace obrazu, ale bylo zapotřebí i aplikace prostorových analýz na vstupní data. Z tohoto hlediska se jako optimální programové prostředí jevílo prostředí GIS s integrovaným modulem pro klasifikaci obrazu. Z tohoto důvodu byl využit pro realizaci projektu software ArcGIS verze 10.0 (32-bit) od firmy ESRI, který má k dispozici modul Image Classification umožňující klasifikaci obrazu. ArcGIS patří v dnešní době k jednomu z nejrozšířenějších profesionálních programů pro tvorbu a správu GIS. Protože bylo potřeba jednotlivé kroky úlohy co nejvíce automatizovat, byl dále využíván aplikační software Python 2.6 (32-bit) pro tvorbu skriptů, které využívaly DLL knihovny ArcGIS 10.0.

Vhodný hardware

Pro úspěšné zpracování úlohy s velkým objemem dat bylo potřeba vytvořit počítačovou síť s výkonnými jednotkami. K určení počtu vhodných prostředků pro řešení úlohy a k přiřazení procesů jednotlivým prostředkům byl využit výše uvedený optimalizační algoritmus (viz podkapitola 3.3). Postup algoritmu je možné rozdělit do těchto fází:

1. určení celkové doby řešení T_{tot}
2. identifikace dílčích procesů

3. počáteční výběr prostředků
4. aplikace optimalizačního algoritmu.

Výsledkem je optimální počet takových prostředků, které zaručují vyřešení celé úlohy v požadovaném termínu. Předpokládáme, že každý dílčí proces v mapě procesů t_{ij} (viz tab. 2) se skládá z 3 částí:

$$t_{ij} = t_{ijp} + t_{ija} + t_{ijm}, \quad (33)$$

kde t_{ijp} je doba přípravy, která zahrnuje: získání vstupních dat, tvorbu dílčích šablon projektu, tvorbu klasifikačních klíčů, návrh a vývoj skriptů v jazyku Python a čas nutný pro organizaci práce celého projektu. Dále t_{ija} je doba automatizovaného a t_{ijm} doba ručního zpracování procesu p_i na prostředku m_j . Automatizovaným procesem se rozumí např. běh procedur v jazyku Python, do ručního zpracování můžeme zařadit vizuální kontroly dílčích i konečných výsledků v prostředí ArcGIS.

Nejdříve bylo nutné stanovit celkovou dobu řešení úlohy T_{tot} . Všechny časové údaje budeme převádět na jednotku 1 hodiny. Vycházelo se z požadavků zadavatele, který očekával výsledky do 16 týdnů, tj. přibližně do 4 měsíců. Z této doby musíme odečíst čas přípravy t_{ijp} , který byl odhadnutý na dobu trvání 1 měsíce. Pro zbývající 3 měsíce odvodíme efektivní čas řešení $\sum_{i,j}^n t_{ija} + t_{ijm}$ v hodinách následovně:

Předpokládáme-li, že příprava dat bude probíhat z celkové doby řešení čtvrtinu času (tj. 20 dní z 16 pracovních týdnů), pak nám na vlastní zpracování zbývá 60 pracovních dní. Každý den bude probíhat manuální a automatizované zpracování procesů. V pilotním projektu bylo zjištěno, že automatizované procesy zabírají 46 % a manuální procesy 54 % celkové doby řešení (viz podkapitola 5.8.2). Uvažujeme-li, že denně budeme manuálně zpracovávat úlohy maximálně po dobu 4 hodin, pak vzhledem k poměru 54:46 zbývá na automatizované zpracování cca 3,5 hodiny. Celková denní kapacita tak činí $4 + 3,5 = 7,5$ hodin. Při 60 pracovních dnech je to $60 * 7,5 = 450$ hodin.

ID procesu	Počet objektů [km]	t_{ija} (automatizace) [h:m:s]	t_{ijm} (manuálně) [h:m:s]	$t_{ija} + t_{ijm}$ [h:m:s]
1	1493,4	16:10:43	4:30:00	20:40:43
2	1295,6	14:02:08	4:30:00	18:32:08
3	1134,6	12:17:29	4:30:00	16:47:29
186	70,0	0:45:30	4:30:00	5:15:30
187	61,8	0:40:10	4:30:00	5:10:10
188	35,3	0:22:57	4:30:00	4:52:57

Tab. 4 – Určení dob zpracování jednotlivých procesů v závislosti na počtu prvků (ilustrativní část)

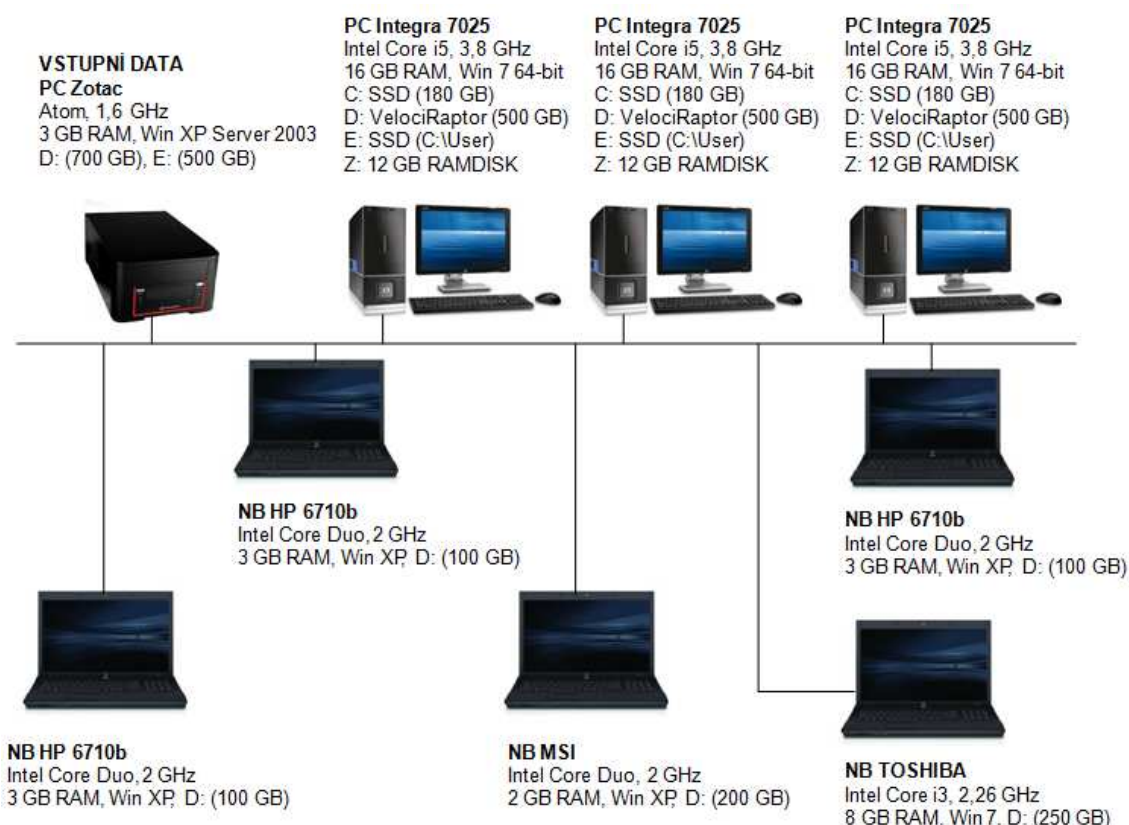
V další části musíme identifikovat procesy. Celý projekt byl rozdělen na územním principu do 188 nezávislých procesů, reprezentující území ORP. Doba zpracování v těchto dílčích územních celcích závisí jednak na jejich rozloze a na počtu prvků (geografických objektů), s nimiž se provádějí požadované operace. Příslušné doby zpracování daných územních celků byly určeny experimentálně v rámci pilotního projektu. Prvotní návrh s výchozími hodnotami

v závislosti na počtu geografických objektů, tj. délce zpracovávaného plynovodu v příslušném ORP, je v tabulce 4. Tato tabulka obsahuje jen vybranou část, celá tabulka má 188 řádků.

Následoval krok stanovení vhodných prostředků pro zpracování úlohy. Mezi vhodné prostředky pro zpracování byly vybrány 2 typy počítačů:

- výkonný stroj typu PC s parametry Intel Core i5 (3.8 GHz), 16 GB RAM, NVIDIA GTX650 (2 GB), HD typu SSD a VelociRaptor
- notebook s parametry Intel Core Duo (2 GHz), 3 GB RAM.

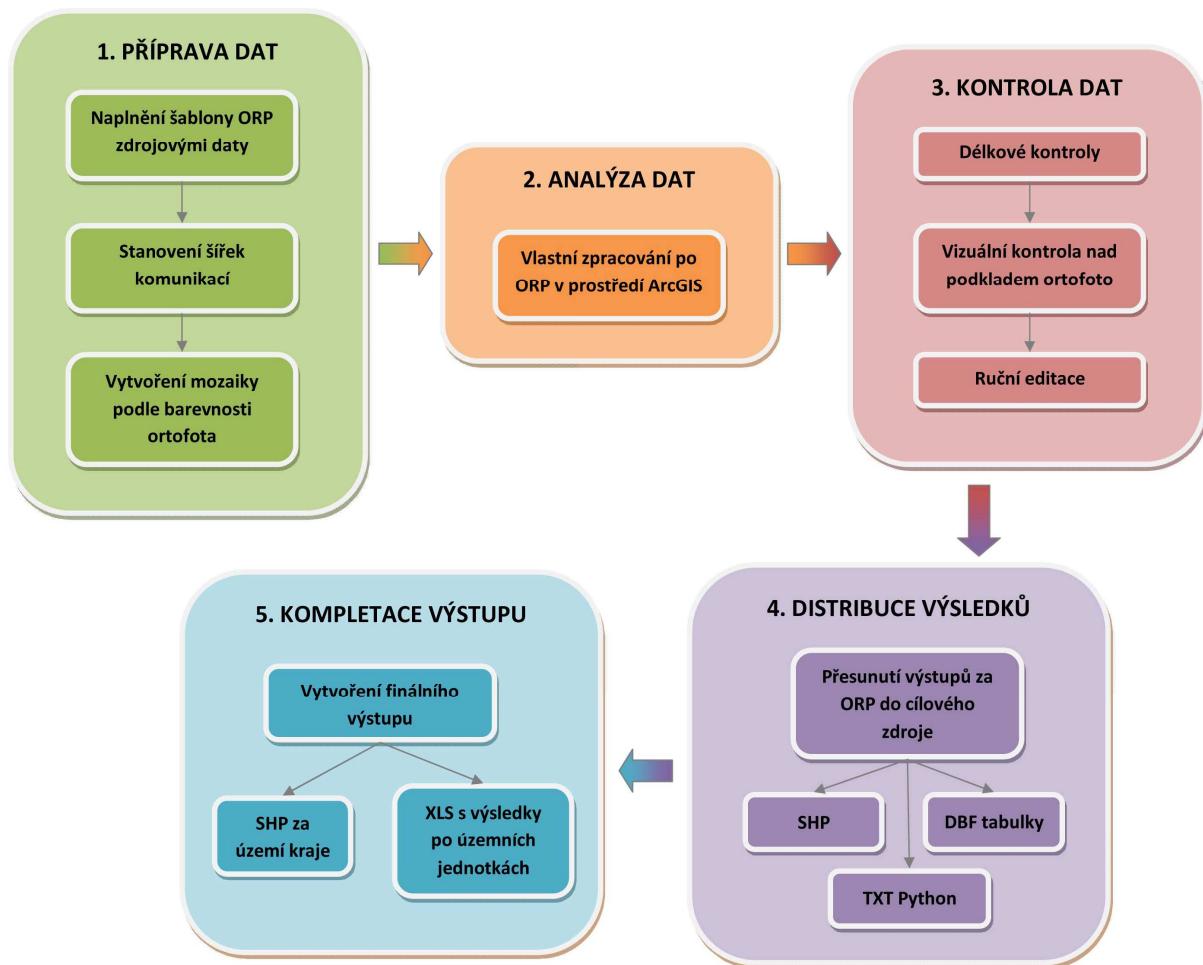
Aplikací optimalizačního algoritmu bylo zjištěno, že minimální počet prostředků pro celkovou dobu řešení je použití 3 výkonných PC a 1 notebooku. Notebooků bylo nakonec použito více, ale sloužily spíše jen jako záložní zdroje, nebo jako prostředek ke vzdálenému přístupu k PC. Schéma počítačové sítě a parametry jednotlivých počítačů jsou na obr. 10. Počítače byly zapojeny do sítě WLAN 100 Mb/s. Síť byla tvořena PC se vstupními datovými sadami, dále třemi výkonnými počítači typu PC a pěti notebooky. Na všech počítačích byl nainstalován ArcGIS. PC se vstupními daty byl zabezpečen a zpřístupněn pouze řešitelům projektu, aby byla ochráněna autorská práva k datům. Téměř veškerá příprava dat a jejich následné hlavní zpracování probíhalo na výkonných počítačích.



Obr. 10 – Schéma zapojení a parametry strojů v počítačové síti

5.1.4 Fáze postupu řešení úlohy

V rámci pilotního projektu byl navržen pracovní postup řešení celé úlohy. Jednotlivé fáze postupu jsou znázorněny na obr. 11 a jsou podrobně popsány v následujících podkapitolách.



Obr. 11 – Fáze pracovního postupu řešení úlohy

5.2 Příprava dat

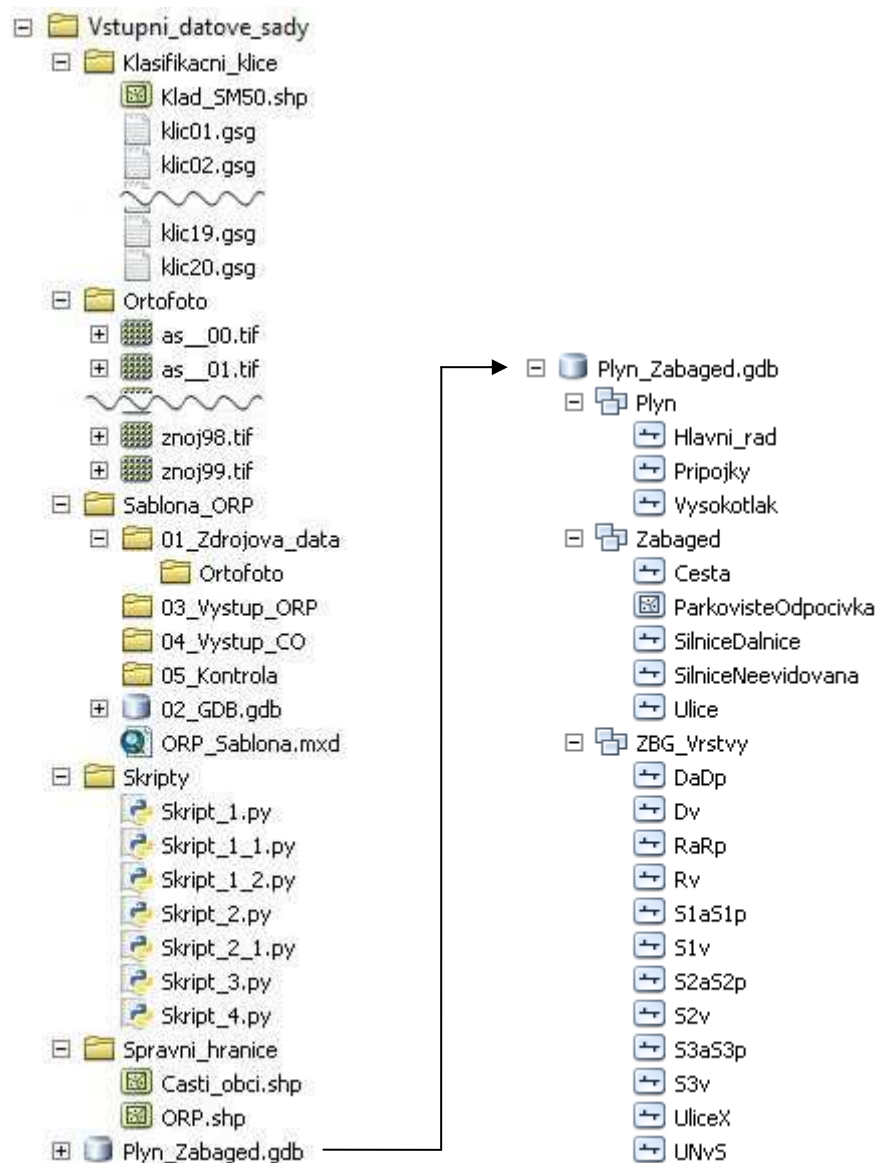
Příprava dat je velmi důležitou fází při řešení masivních úloh před hlavním zpracováním dat. Tato fáze se skládala ze dvou kroků. Po získání veškerých potřebných dat pro analýzu bylo nejprve potřeba tato data upravit a dále je nachystat pro následné zpracování.

5.2.1 Úprava zdrojových dat

Mezi zdrojová data patřily:

- Vektorový zákres plynového vedení ve formátu DGN (9 výkresů, 270 MB, poskytovatel RWE)
- Rastrové ortofoto snímky ČR v měřítku 1: 5 000 s rozlišením 0,25 m/pixel ve formátu TIFF (16 313 mapových listů, 480 GB, poskytovatel RWE)
- Vektorový zákres os komunikací ZABAGED ve formátu SHP (5 datových vrstev, 1,6 GB, poskytovatel ČÚZK prostřednictvím RWE)
- Vektorové datové soubory hranic správních jednotek ve formátu SHP (5 datových sad, 290 MB, poskytovatel ČÚZK a RWE).

Zákres plynového vedení a vrstvy ZABAGED byly v rámci úpravy převedeny na datové typy feature class (FC, třída prvků) do společné geodatabáze. Soubory hranic správních jednotek byly upraveny do dvou řídicích souborů ÚIR. Upravená zdrojová data byla uložena ve struktuře, která je znázorněna na obr. 12.



Obr. 12 – Struktura uložení upravených zdrojových dat

Plynové vedení

Zákres plynového vedení pro hlavní řad, přípojky a vysokotlak byl v různých DGN výkresech. Proto bylo nejprve potřeba tyto výkresy převést na datový typ FC, z kterých se následně vybírala data pro 3 typy plynodůů. Vznikly tak 3 liniové vrstvy plynového vedení, které byly následně podrobeny topologické kontrole v ArcGIS. Vzhledem k charakteru datové analýzy byly kontrolovány pouze ty chyby, které by měly vliv na finální výsledek analýzy. Proto byla analyzována tři topologická pravidla, která jsou znázorněna v tabulce 5. Celkový počet opravených chyb byl 263. Následně byly opravené vrstvy uloženy do zdrojové geodatabáze do datasetu *Plyn*. Struktura této geodatabáze je na obr. 12.

Topologické pravidlo	Obrázek	Popis	Celkový počet chyb
Různé linie se nesmí překrývat		modrá – 1. linie žlutá – 2. linie červená – překryt 1. a 2. linie	110
Linie nesmí překrývat sama sebe		modrá linie v červených místech překrývá sama sebe	7
Linie nesmí sama sebe protínat		modrá linie sama sebe protíná či překrývá, průnik červeně	146

Tab. 5 – Topologická kontrola zákresu plynového vedení

ZABAGED

Využité datové vrstvy os komunikací ZABAGED v původním formátu SHP byly převedeny a uloženy do zdrojové geodatabáze do datasetu *Zabaged* (viz obr. 12). Následně byly z FC *SilniceDalnice* a *Ulice* vybrány konkrétní zákresy jednotlivých tříd komunikací a uloženy jako nové FC do datasetu *ZBG_Vrstvy*. Jde např. o silnice 1., 2. a 3. třídy. Z vybraných vrstev datasetu *Zabaged* a z vrstev datasetu *ZBG_Vrstvy* byl poté tvořen filtr pro výstup z klasifikace obrazu (viz podkapitola 5.3).

Správní hranice

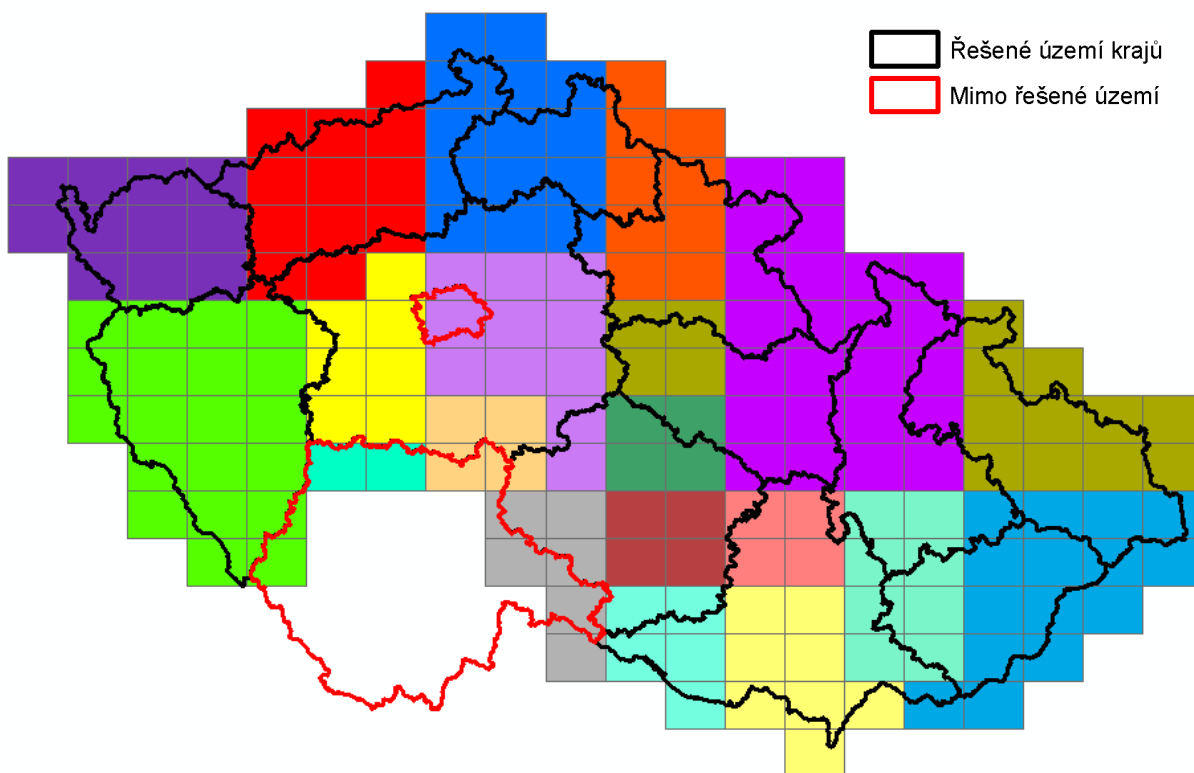
Mezi použitá zdrojová data s vektorovým zákresem hranic správních územních celků patřily vrstvy krajů, okresů, ORP, obcí a částí obcí. Problémem bylo, že tyto vrstvy nebyly ve všech případech navzájem atributově propojeny a chyběla zde číselníková vazba. Například vrstva ORP obsahovala atribut vyjadřující příslušnost ke krajům, ale chyběla zde číselníková vazba k okresům. Vrstva částí obcí měla atributy příslušnosti ke krajům i okresům, naopak zcela chyběla vazba na ORP. Proto data musela být upravena a vznikly tak dva řídicí soubory:

- *ORP.shp* – Zdrojový SHP soubor s hranicemi ORP byl automatizovaně upraven pomocí účelově vytvořeného skriptu. Při této úpravě se doplnil do tohoto souboru potřebný atribut s vazbou na okresy. Práce skriptu spočívala v tom, že si vytvořil do paměti počítače polygony reprezentující území jednotlivých ORP a okresů a tyto polygony navzájem porovnával.
- *Casti_obci.shp* – Původní zdrojový SHP soubor s hranicemi částí obcí obsahoval duplicitní řádky příslušející k jedné části obce. Duplicita se v tabulce nacházela v případech, kdy daná část obce není ve skutečnosti jedno celistvé území, ale je složena ze dvou či více disjunktních areálů. Tato duplicita byla pro další operace nežádoucí a musela být odstraněna. Vymazáním duplicit se sjednotily nesourodé areály částí obcí. Nad zpracovávaným územím bylo tak odstraněno 120 duplicit. Dále byla do souboru doplněna chybějící vazba na ORP, opět automatizovaně pomocí účelově vytvořeného skriptu. Skript tentokrát nevyužíval geometrického porovnávání jednotlivých polygonů, ale pracoval na principu porovnávání vazeb tří číselníků mezi soubory s hranicemi ORP, obcí a částí obcí.

Ortofoto a klasifikační klíče

Po zhlédnutí získaných snímků ortofoto, které pokrývaly celé zpracovávané území, bylo zjištěno, že nejsou stejné škály barevnosti. Tato situace zkomplikovala řešení úlohy, protože nemohlo být použito univerzálního klasifikačního klíče při klasifikaci obrazu ortofota. Různá barevnost ortofota vznikla díky technologii při jeho vzniku. Mapové dílo Ortofoto ČR tvoří ČÚZK ve spolupráci s VGHMÚř a je pravidelně aktualizováno. Fáze barevného vyrovnání snímků ortofoto lze rozdělit na dva základní kroky. Nejprve se provádí etapa barevného vyrovnání mezi snímky pocházejícími z různých snímkových letů v rámci jednotlivých bloků a mezi bloky navzájem, zejména pokud byly pořízeny v časovém odstupu několika týdnů až měsíců (při změně vzrůstu a zralosti vegetace, při výrazné změně výšky Slunce nad obzorem, při místním zastínění území stíny způsobenými vysokou oblačností, při opakovaném snímkování částí zastíněnými lokálními mraky v době původního snímkování). Následuje automatické barevné vyrovnání celých oblastí, resp. celého pásma snímkaného v příslušném roce [88].

Mapové listy, které byly získány poskytovatelem, pocházely z let snímkování 2010 – 2012. Po detailnějším zkoumání bylo nad zájmovým územím nalezeno 20 odlišných oblastí barevnosti, které však byly různých rozměrů (viz obr. 13). Dále bylo zjištěno, že hranice těchto oblastí kopírují hranice kladu listů Státní mapy v měřítku 1: 50 000 (SM50). Z tohoto kladu je odvozen klad mapových listů ortofoto v měřítku 1: 5 000.



Obr. 13 – Oblasti barevnosti ortofota podle kladu mapových listů SM50

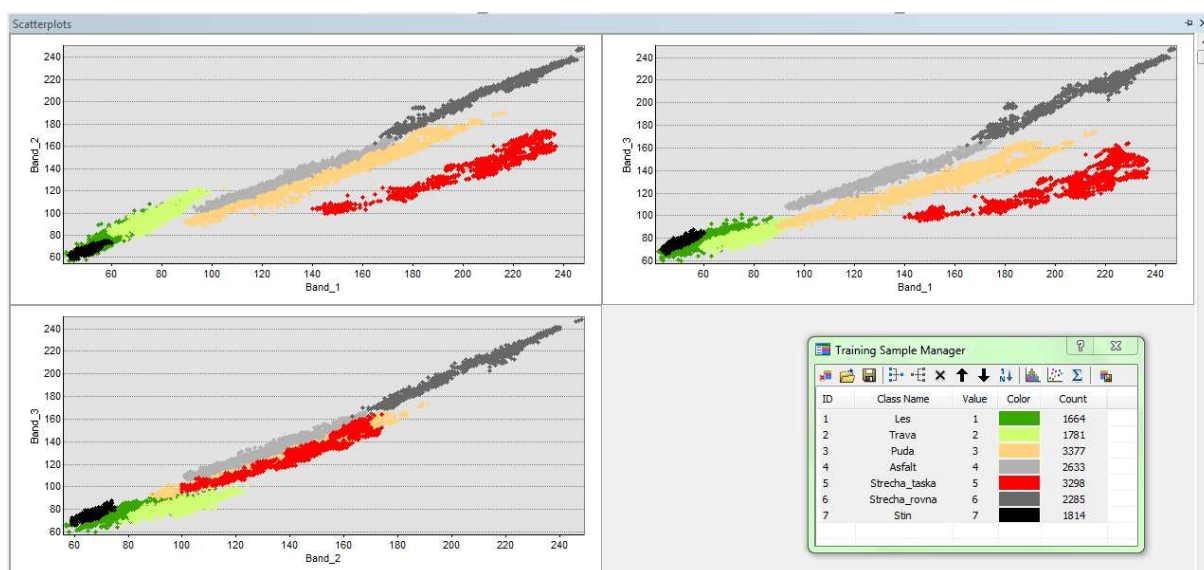
Pro těchto 20 oblastí barevnosti byly v prostředí ArcGIS manuálně vytvořeny trénovací množiny pro každou z těchto oblastí a z nich byly následně spočítány klasifikační klíče. Snahou bylo tvořit trénovací množiny o velikosti 100 – 130 pixelů, přičemž pro každou třídu

bylo vždy vytvořeno minimálně 15 množin. Klasifikačních tříd bylo použito 7 a jejich výčet je v tabulce č. 6.

Druh	Třída
Les	1
Travní porost	2
Holá půda	3
Asfalt	4
Střecha - taška	5
Střecha - rovná	6
Stín	7

Tab. 6 – Klasifikační třídy pro klasifikaci obrazu ortofota

Pro tvorbu trénovacích množin se v prostředí ArcGIS používal nástroj Training Sample Manager. Tento nástroj dokáže vypočítat statistiky pro vybrané trénovací množiny nad daným snímkem, jedná se zejména o znázornění tzv. histogramů a rozptylogramů. Histogramy dávají uživateli informaci o rozložení a četnosti barevných hodnot pixelů trénovacích množin v aktivním snímku pro každé barevné pásmo zvlášť. Rozptylogram znázorňuje srovnání hodnot pixelů množin stejných klasifikačních tříd v závislosti na dvou barevných pásmech. Ukázka rozptylogramů při vyhodnocení trénovacích množin pro kombinace barevných pásem RGB je na obr. 14. Pomocí těchto statistických grafů je uživatel seznámen s kvalitou trénovacích množin. Pokud množiny reprezentují různé třídy, neměly by se v grafech překrývat.

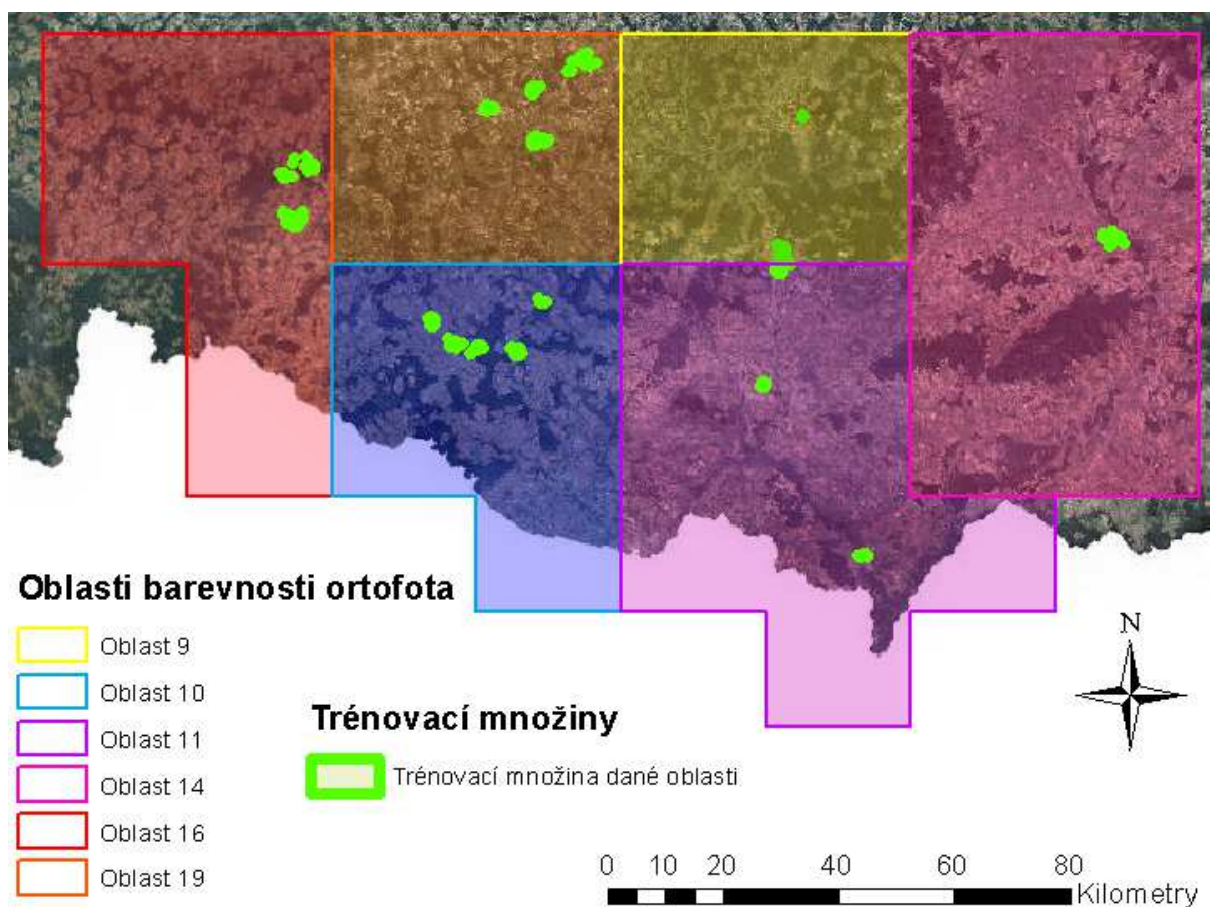


Obr. 14 – Ukázka vyhodnocení rozptylogramů pro vybrané trénovací množiny

Při tvorbě trénovacích množin se v prostředí ArcGIS vyskytl problém, při kterém docházelo k opětovné havárii programu. Konkrétně se jednalo o práci s manažerem pro tvorbu trénovacích množin. Potíže s havárií aplikace se vyskytly vždy v případě, kdy bylo potřeba z důvodu editace pracovat s vytvořenými množinami, které byly opětovně načteny do manažera po novém spuštění ArcGIS. Trénovací množiny manažer sice načtl, ale nedalo se s nimi dále pracovat, např. jim spočítat statistiku (histogram, rozptylogram), protože program ArcGIS přestal pracovat. Bylo zjištěno, že pokud jsou množiny jedné třídy od sebe vzdáleny

více jak cca 6 km, program zhavaruje. Pokud jsou množiny od sebe vzdálené třeba 10 km v rámci jiných tříd, program bez problémů statistiku spočítá. Potíže se vyskytly tedy pouze v případech vzdálených množin v rámci jedné třídy.

Tyto problémy byly řešeny i s technickou podporou softwaru ArcGIS pro ČR a z důvodu nenalezeného řešení byl dále tento problém odeslán k technické podpoře ESRI. Tato programová chyba, kterou lze nalézt pod označením NIM097046 na internetových stránkách technické podpory společnosti ESRI, nebyla doposud opravena. Situace proto byla řešena tak, že trénovací množiny jedné třídy byly vybírány ve svém blízkém okolí, tzn., do oblasti v rozsahu maximálně 5 km. Ukázka rozmístění trénovacích množin u šesti vybraných oblastí barevnosti je na obr. 15.



Obr. 15 – Rozmístění trénovacích množin v oblastech barevnosti ortofota

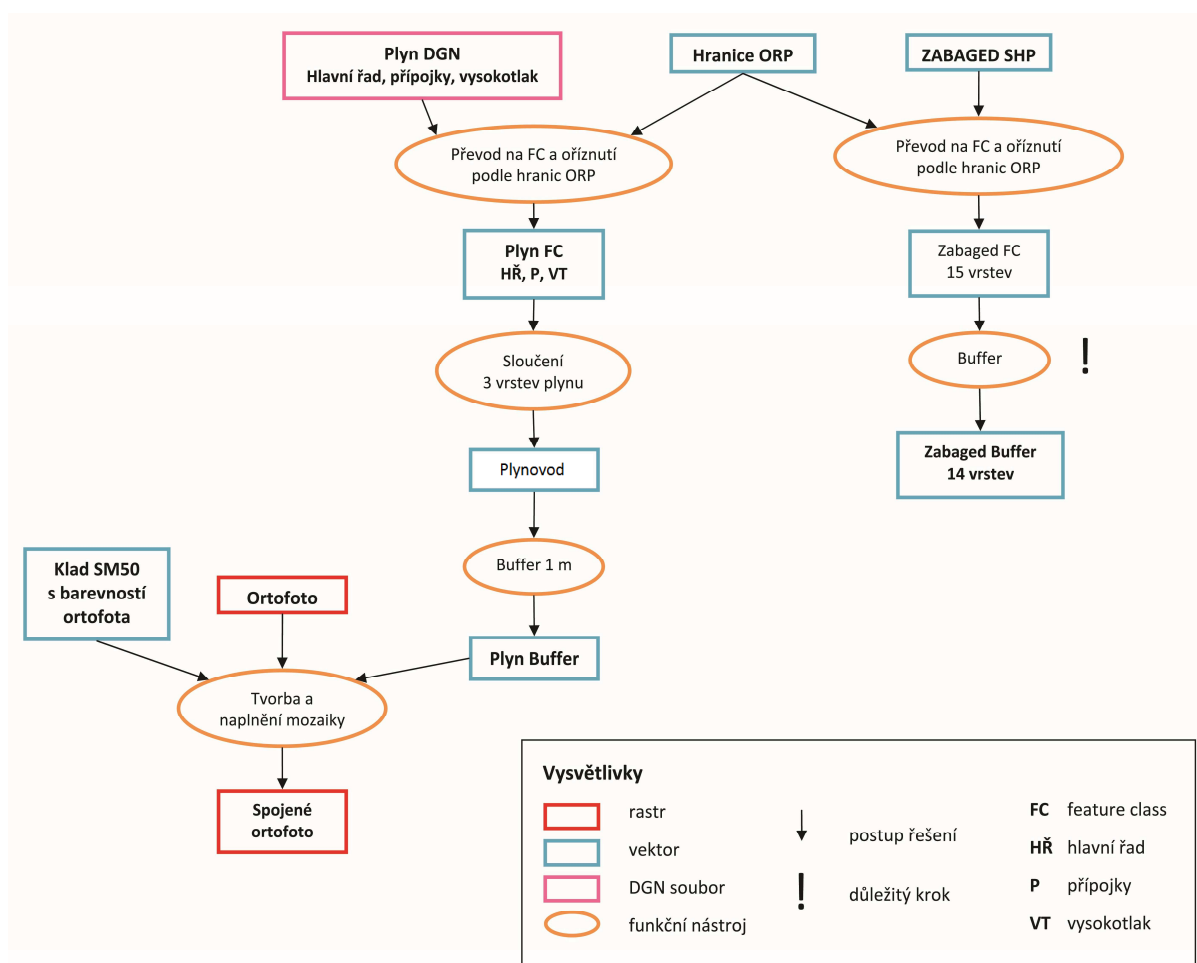
Uvažovalo se i o kalibraci oblastí barevnosti ortofota podle vybrané referenční oblasti barevnosti. Všechna ortofota by pak měla stejnou škálu barevnosti a mohl by tak být použit jeden univerzální klasifikační klíč. Tato varianta však byla shledána jako nereálná, vzhledem k obtížnosti kalibrace díky velkým rozdílům v barevnosti jednotlivých listů ortofota způsobených zejména odlišnými podmínkami při snímkování území.

5.2.2 Příprava datových zpracovatelských jednotek

Jakmile byla provedena kompletní úprava zdrojových dat a byly vytvořeny klasifikační klíče, mohlo se přistoupit k přípravě datových zpracovatelských jednotek před vlastním zpracováním. Nejprve bylo potřeba vytvořit strukturu adresářů po územních jednotkách kraje, okresu a ORP. Tato struktura sloužila ke snadnější orientaci při manuální práci s jednotlivými zpracovatelskými jednotkami. Následovaly tři hlavní kroky přípravy datových jednotek:

1. Naplnění šablony ORP zdrojovými daty
2. Stanovení skutečných šířek komunikací
3. Vytvoření mozaiky podle barevnosti ortofota.

Zjednodušené schéma datové přípravy je na obr. 16.



Obr. 16 – Zjednodušené schéma řešení přípravy dat pro analýzu

Naplnění šablony ORP zdrojovými daty

Algoritmus předzpracování vstupních dat s cílem naplnit šablonu zdrojovými daty je znázorněn na obr. 29 a jeho proces byl následující. Do adresáře vybraného ORP se nakopírovala šablona prázdného GIS projektu. Následně byla vytvořena maska typu polygon představující území příslušného ORP a dále také masky představující území částí obcí. Poté

byly ze vstupních datových sad kopírovány vrstvy zákresu průběhu plynu do formátu FC do geodatabáze projektu ORP, stejný proces následoval s 15 vrstvami komunikací ZABAGED. Plyn a ZABAGED byly při procesu kopírování rovnou ořezány podle hranic ORP prostřednictvím masky.

ZABAGED byl dále upravován tak, že se kolem zákresu linií komunikací vytvořily obalové polygony funkcí *Buffer*. V této chvíli byly velikosti bufferu kolem jednotlivých kategorií komunikací nastaveny jako implicitní hodnoty, které byly zjištěny předchozím ručním přeměněním a zprůměrováním šířek komunikací ve čtyřech vybraných ORP v prostředí ArcGIS. Pozn. u vrstvy *ParkovisteOdpočivka* se obalový polygon nevytvářel.

Plyn byl sjednocen z hlavního řádu, přípojek a vysokotlaku do jedné třídy a kolem něj byl vytvořen buffer velikosti 1 m. Velikost bufferu byla stanovena na základě těchto dvou parametrů: rozlišení ortofota (25 cm) a polohová přesnost ortofota a ZABAGED, která podle sdělení pracovníků Zeměměřického úřadu činí cca 35 cm. Vrstva bufferu kolem plynovodu měla dále sloužit jako podklad pro výběr ortofota, nad nímž je zákres plynovodu. Bylo to z toho důvodu, aby se předešlo zbytečnému kopírování velkého objemu dat ortofota, nad kterým nevede plynovod (viz tematická selekce podle zobrazení *S* – podkapitola 3.2).

Výběr ortofoto pokrývající průběh plynovodu v příslušném ORP a nakopírování mapových listů ze vstupních sad do šablony projektu, probíhal následujícím způsobem. Nejprve byly zjištěny souřadnice severozápadního rohu mapového listu ortofota, které se dají zjistit z umístovacího souboru TFW. Tento soubor musí mít každé georeferencované ortofoto. Ze souřadnic severozápadního rohu ortofota byly spočítány souřadnice ostatních rohů kladu mapového listu příslušného ortofota ze znalosti rozměru kladu 2 x 2,5 km. Ze souřadnic rohů mapového listu byl následně virtuálně vytvořen polygon kladu. Ten byl podroben testování, zda zasahuje do polygonu upravené vrstvy bufferu kolem plynu. Testováním prošla všechna ortofota ze vstupních datových sad. Protože test virtuálních polygonů probíhal pouze v paměti počítače, snížila se tím časová náročnost. Všechna nakopírovaná ortofota v šabloně projektu byla následně spojena do jednotné mozaiky, zatím bez rozlišení jejich barevnosti.

Na obrázku 17 je vidět adresářová struktura pro Plzeňský kraj, ve kterém jsou jednotlivé okresy. V adresářích okresů jsou příslušné ORP. Pro detailnější ukázkou je zde vidět struktura projektu pro ORP Rokycany po fázi naplnění šablony zdrojovými daty. Všechny výše uvedené operace této fáze probíhaly automatizovaně a bylo poměrně složité je naprogramovat.

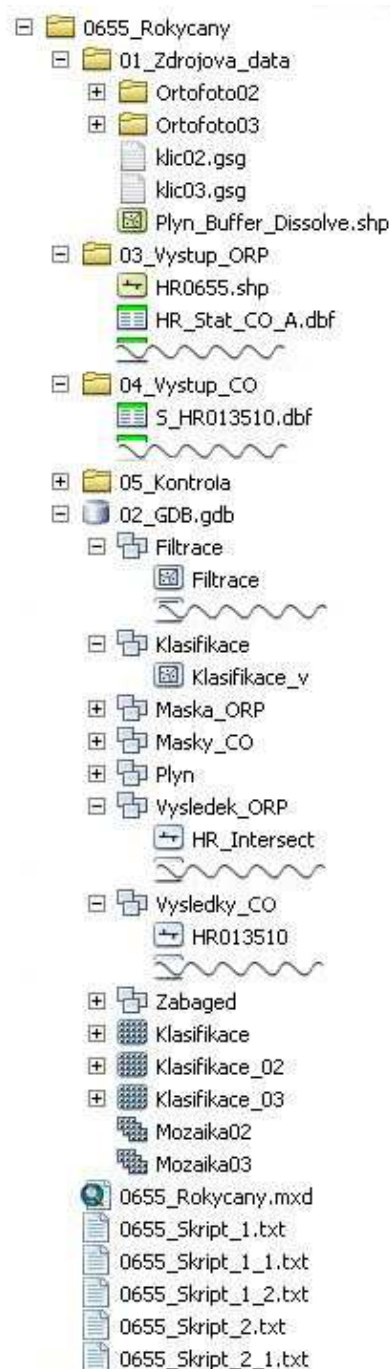
Stanovení skutečných šířek komunikací

Bylo zjištěno, že velikosti šířek komunikací nejsou ve všech ORP stejné a tudíž je nelze jednoduše parametrizovat. Rozdíly ve velikosti defaultních hodnot bufferů oproti skutečnosti by vnesly chyby do analýzy. Proto bylo rozhodnuto, že musí proběhnout v rámci procesu přípravy dat zjištění skutečných šířek komunikací v daném ORP. Šířky byly poté stanoveny jako průměrné hodnoty z manuálního odměřování nad zákresem linií ZABAGED. Během měření bylo pořízeno tolik vzorků pro každé ORP, aby tvořily dostatečnou reprezentaci z hlediska statistiky. Jako podklad pro tuto činnost sloužila vytvořená mozaika z ortofota z předchozí fáze. Poté došlo k automatickému přegenerování bufferů kolem komunikací podle

nových skutečných šířek. Pokud byla v daném ORP shledána skutečná šířka bufferu jako vyhovující z nastavených defaultních hodnot, tento parametr nebyl přeměřován a zůstal pro tento konkrétní typ komunikace zachován.



Obr. 17 – Stromová struktura adresářů Plzeňského kraje včetně detailní struktury projektu ORP Rokycany po fázi naplnění šablony zdrojovými daty

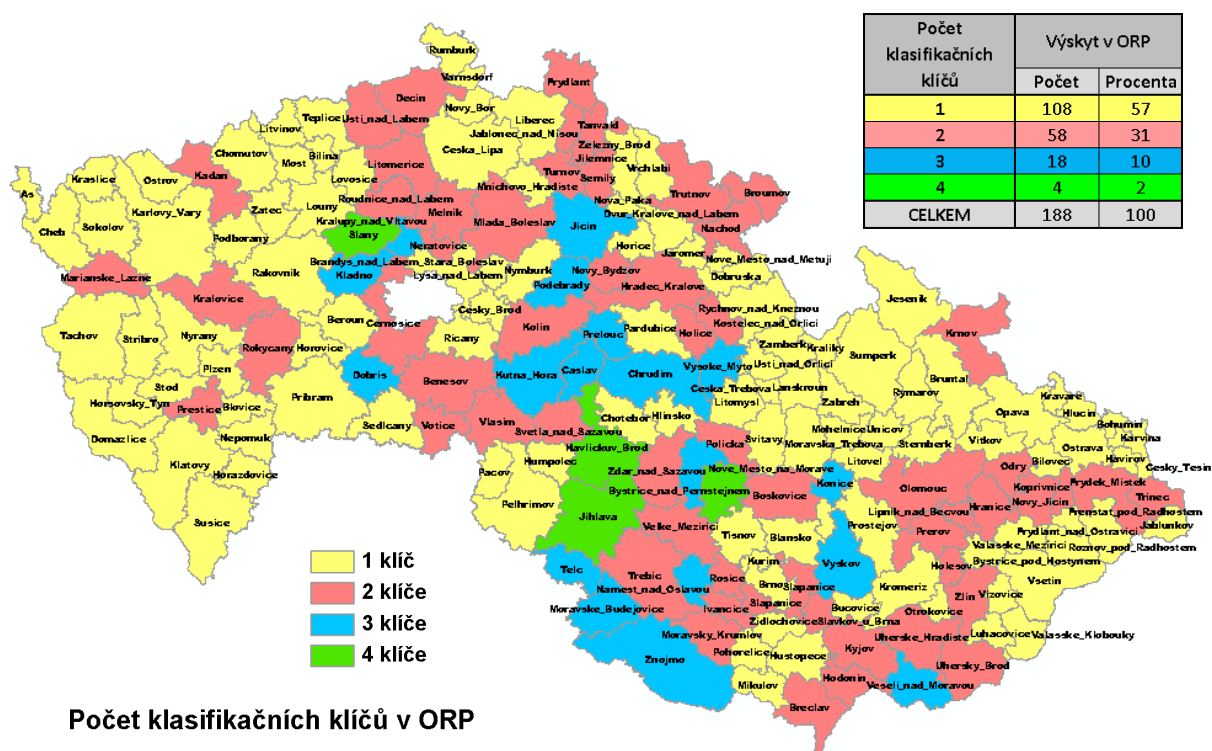


Obr. 18 – Stromová struktura projektu ORP Rokycany po fázi hlavní analýzy dat

Vytvoření mozaiky podle barevnosti ortofota

Z důvodu nestejně barevnosti ortofota bylo potřeba v dalším kroku přípravy vytvořit mozaiky spojených rastrů se stejnou barevností nad daným ORP. Jako poklad pro výběr ortofota k tvorbě mozaik sloužil upravený klad SM50 ve formátu SHP, který obsahoval dvě důležité informace. Každý mapový list měl k sobě přiřazen svůj název, který tvoří 4 textové znaky a z kterého se odvozuje klad ortofota, a dále označení klasifikačního klíče podle oblasti barevnosti, do které klad spadal. Výběr probíhal tak, že se porovnával název souborů ortofota umístěných v adresáři naplněné šablony ORP s názvem v souboru s kladem SM50. Podle toho byla kopírována vybraná ortofota do nových adresářů, z kterých se následně tvořila spojená mozaika stejné barevnosti. Poté byly zkopírovány ze vstupních datových sad potřebné vybrané klasifikační klíče do šablony ORP a dále byla smazána původní celistvá mozaika z ortofota nad celým ORP bez rozlišení barevnosti. Nakonec byl smazán i původní adresář s kompletním ortofoto pro celé ORP.

V případě, že území řešeného ORP spadalo pouze do jedné oblasti barevnosti ortofota, původní celistvá mozaika byla pouze přejmenována podle dané oblasti barevnosti a do šablony byl nakopírován příslušný klasifikační klíč.

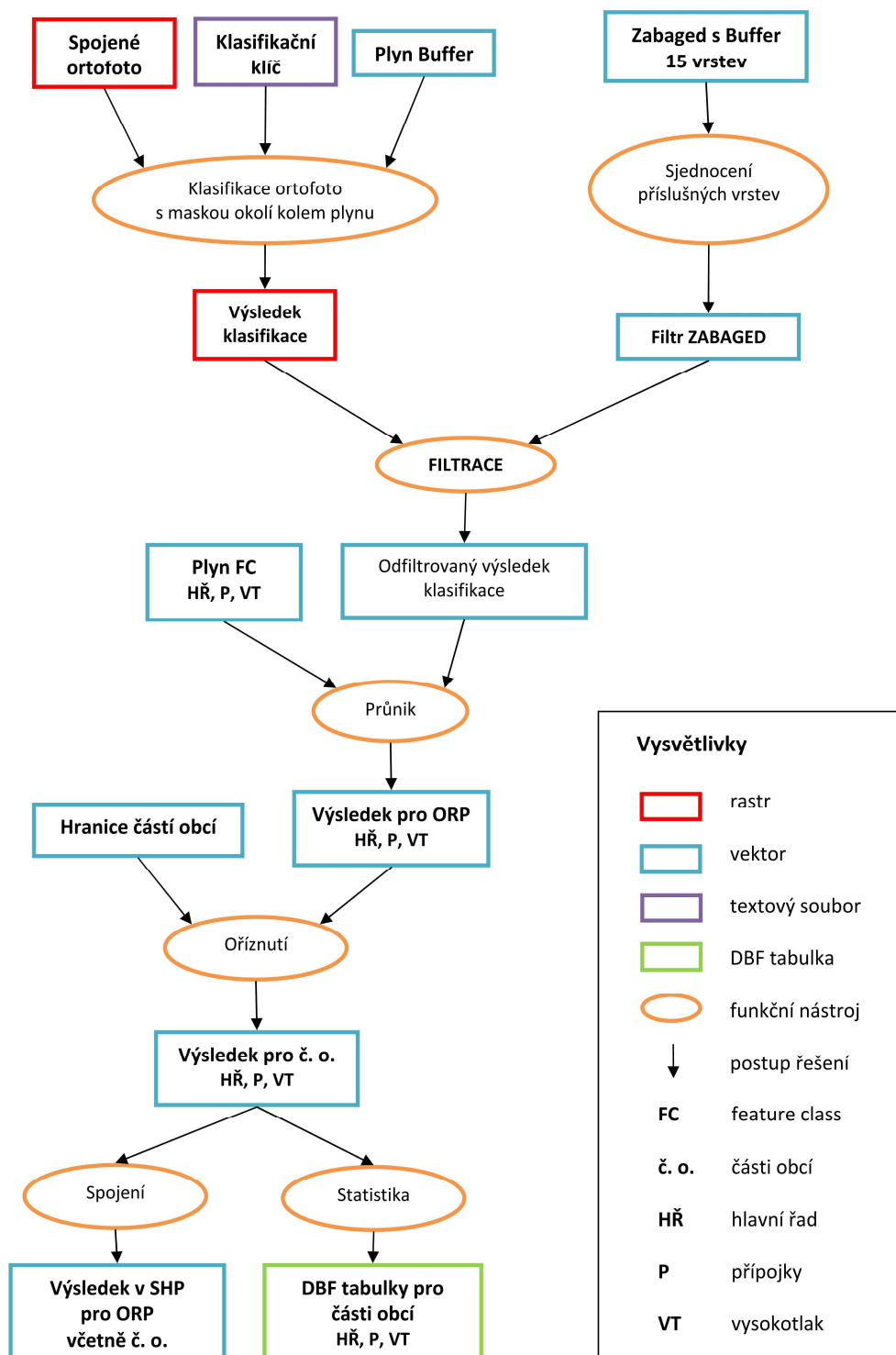


Obr. 19 – Počet potřebných klasifikačních klíčů pro jednotlivá ORP

Celý tento proces probíhat automatizovaně. Na obr. 19 je vidět, do kolika oblastí barevnosti ortofota spadala jednotlivá území ORP, respektive kolik klasifikačních klíčů bylo potřeba pro řešení. Nejvíce je zastoupena skupina s jedním klasifikačním klíčem, objevila se zde však i taková ORP, která zasahovala do čtyř oblastí barevnosti ortofota.

5.3 Zpracování dat v GIS

V této fázi dochází k vlastnímu zpracování analýzy nad územím příslušného ORP. Vše probíhá automatizovaně v prostředí ArcGIS. Ukázka naplněného projektu ve stromové struktuře po hlavní analýze pro ORP Rokycany je na obr. 18. Schéma řešení hlavní analýzy je znázorněné na obrázku 20 a je podrobněji popsáno v následujícím textu.

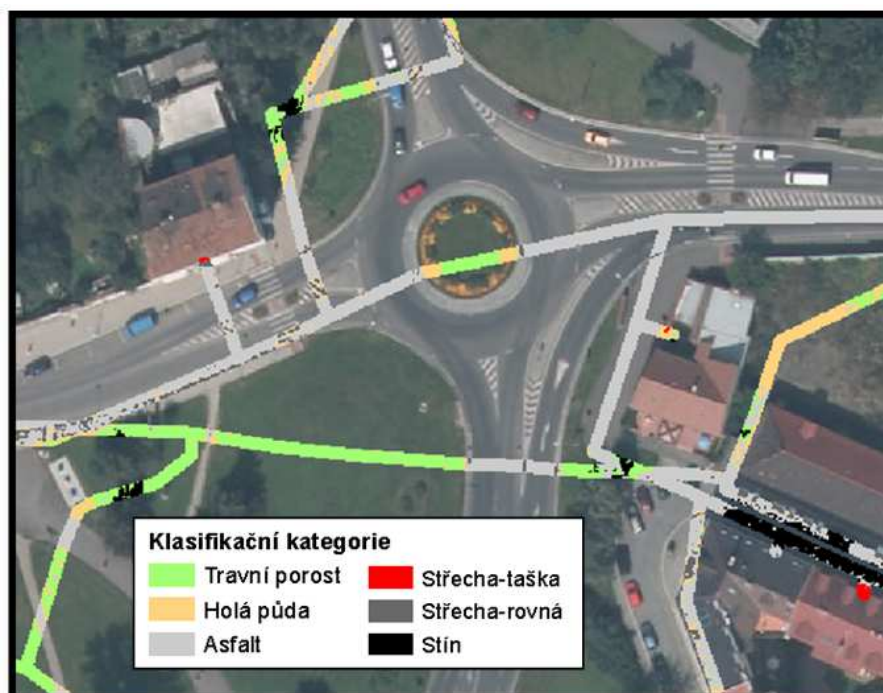


Obr. 20 – Schéma řešení analýzy na základě klasifikace obrazu ortofoto a filtrace vrstvami ze ZABAGED

Klasifikace obrazu

Nad mozaikami z ortofota stejné barevnosti, které pokrývaly oblast vedení plynovodu s obalovou zónou 1 m od trasy vedení, byla provedena klasifikace obrazu. Konkrétně byl zvolen za typ řízené klasifikace *Maximum Likelihood*. Tomuto kroku předcházela náročná tvorba klasifikačních klíčů pro oblasti barevnosti ortofota (viz podkapitola 5.2.1). Při tvorbě klíčů byla vytvořena metodika tvorby trénovacích množin pro 7 základních klasifikačních tříd.

Při klasifikaci obrazu byl automatizovaně přiřazen konkrétní mozaice, která spadala do dané oblasti barevnosti, příslušný klasifikační klíč. Klasifikace obrazu byla prováděna pouze pro oblast bufferu 1 metru kolem plynu a mohly nastat dva případy. Buď byl proveden proces klasifikace pouze jednou, a to v případě, že celá oblast daného ORP spadala do jedné oblasti barevnosti ortofoto. Výsledkem byla jedna rastrová vrstva. Anebo klasifikace byla opakována tolikrát, kolik bylo částí oblastí plynovodu spadající do dané oblasti barevnosti (tzn. nejvíce čtyřikrát). V tomto případě bylo výsledkem více rastrových vrstev, které bylo potřeba spojit. Ukázka detailu výsledku procesu klasifikace obrazu nad průběhem bufferu kolem plynovodu je na obrázku 21.



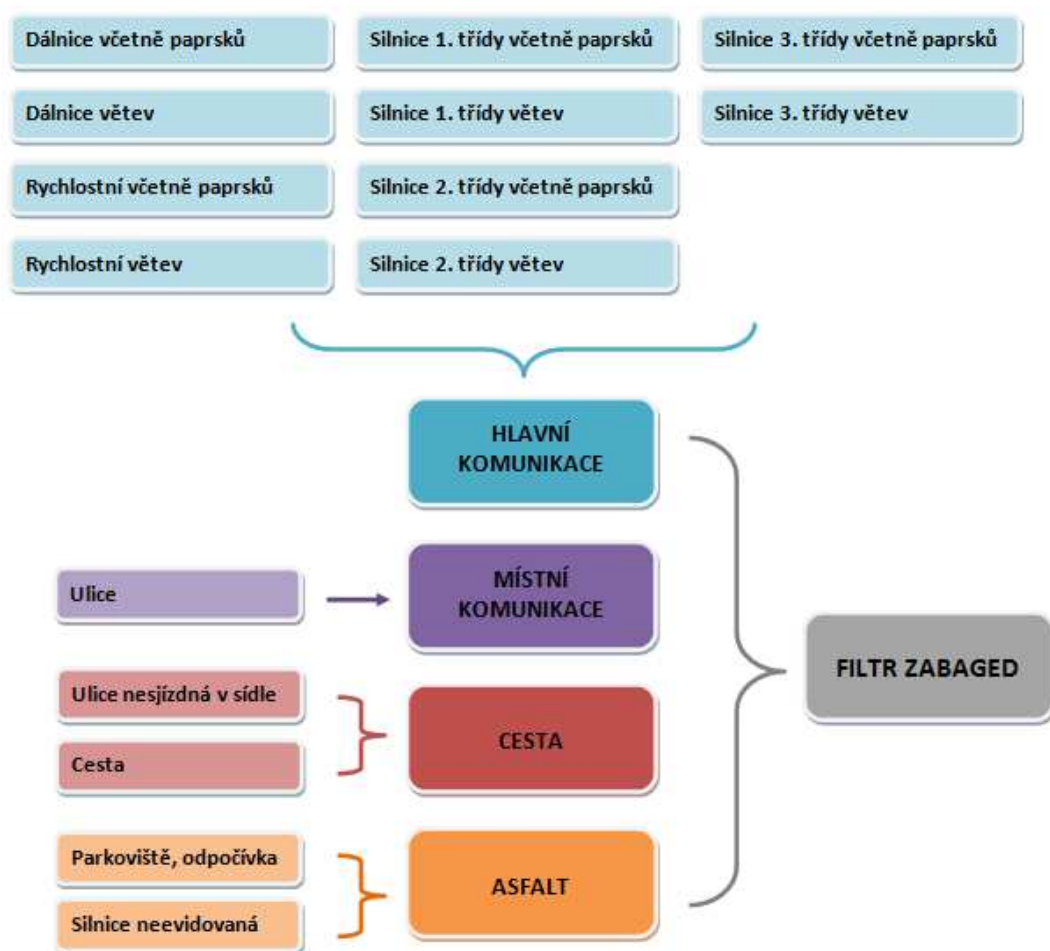
Obr. 21 – Ukázka detailu klasifikace obrazu nad průběhem plynovodu

Proces klasifikace obrazu byl jeden z časově nejnáročnějších analýz v rámci celé úlohy. Průměrně tento proces trval pro území jednoho ORP 50 minut (41 minut je směrodatná odchylka ze souboru 177 ORP zpracovávaných na výkonných PC). Nejdéle proces klasifikace obrazu trval 6 hodin 50 minut, kdy muselo být zpracováno 189 mapových listů ortofota. Druhý nejdélejší čas byl 2 hodiny 35 minut pro 105 mapových listů.

Je důležité poznamenat, že provádět klasifikaci obrazu pro rozlišení 0,25 m/pixel v tak malém měřítku jako je rozsah téměř celého území státu, navíc s různou barevností vstupních dat, je úkol specifický a není běžnou součástí úloh DPZ.

Filtrování klasifikace obrazu

Protože samotný výsledek z klasifikace obrazu by nebyl dostatečně přesný (např. koruny stromů zejména ve městech často zakrývají komunikace, kde jsou tato místa chybně určena jako vegetace) a protože bylo potřeba do analýzy zavést i atribut rozdělení komunikací na hlavní a místní, upravené vrstvy vybraných komunikací ze ZABAGED tvořily tzv. filtr pro vrstvu z klasifikace obrazu. Filtr vznikl sjednocením daných vrstev bufferů komunikací do jednotných tříd. Konkrétní vrstvy bufferů, které byly postupně sjednocovány, jsou naznačeny na obr. 22. Po procesu filtrace tato vrstva již obsahovala nejen atributy z klasifikace obrazu, ale i z filtru.



Obr. 22 – Filtr vytvořený z vrstev komunikací ZABAGED

Vznik výsledku a úprava na části obcí

V atributové tabulce výstupu se v jednom poli objevil atribut příslušející k zařazení do nejpodrobnější klasifikace povrchu terénu, tzn. zařazení do 10-ti klasifikačních tříd. Dále byl do tabulky ještě doplněn atribut přiřazení k územní jednotce kraje, okresu a ORP. Po tomto

kroku došlo k průniku této vrstvy s liniovým zákresem plynovodu, zvláště pro hlavní řad, přípojky a vysokotlak. Vznikly tak 3 vrstvy oklasifikovaného plynovodu pro celé ORP.

Klasifikační kategorie typu A – Základní členění povrchu			
HLAVNÍ ŘAD a PŘÍPOJKY			
Druh	Třída		Vzniká součtem tříd z klasifikace C:
Zpevněný	1		Asfalt, Střecha - taška, Střecha - rovná, Hlavní komunikace, Místní komunikace, Cesta
Nezpevněný	2		Les, Travní porost, Holá půda, Stín
VYSOKOTLAK			
Druh	Třída		Vzniká součtem tříd z klasifikace C:
Zpevněný	1		Asfalt, Střecha - taška, Střecha - rovná, Hlavní komunikace, Místní komunikace
Nezpevněný	2		Les, Travní porost, Holá půda, Stín, Cesta
Klasifikační kategorie typu B – Podrobnější členění povrchu podle RWE			
HLAVNÍ ŘAD a PŘÍPOJKY			
Druh	Třída		Vzniká součtem tříd z klasifikace C:
Asfalt	1		Asfalt, Střecha - taška, Střecha - rovná, Cesta
Hlavní komunikace	2		Hlavní komunikace
Místní komunikace	3		Místní komunikace
Nevázaný	4		Les, Travní porost, Holá půda
Neznámý	5		Stín
VYSOKOTLAK			
Druh	Třída		Vzniká součtem tříd z klasifikace C:
Asfalt	1		Asfalt, Střecha - taška, Střecha - rovná
Hlavní komunikace	2		Hlavní komunikace
Místní komunikace	3		Místní komunikace
Nevázaný	4		Les, Travní porost, Holá půda, Cesta
Neznámý	5		Stín
Klasifikační kategorie typu C – Podrobnější členění povrchu podle VUT			
HLAVNÍ ŘAD, PŘÍPOJKY a VYSOKOTLAK			
Druh	Třída		
Les	1		
Travní porost	2		
Holá půda	3		
Asfalt	4		
Střecha - taška	5		
Střecha - rovná	6		
Stín	7		
Hlavní komunikace	8		
Místní komunikace	9		
Cesta	10		

Tab. 7 – Výstup klasifikace druhů povrchů nad průběhem plynového vedení pro klasifikační kategorie A, B a C

Vrstvám musela být dále upravena atributová tabulka tak, aby obsahovala atributy nejen nejpodrobnější klasifikace typu C, ale i atributy stručnější klasifikace B a A. Tyto třídy vznikaly výběrem a spojováním z nejpodrobnější klasifikace. Jaké konkrétní třídy byly spojovány, je naznačeno v tabulce 7. Zde je vidět, že třída *Cesta* byla v případě hlavního řadu a přípojek zahrnuta do zpevněného povrchu, naopak u vysokotlaku byla zahrnuta do povrchu nezpevněného.

Souběžně proběhla reklasifikace vrstvy vysokotlaku z následujícího důvodu. Jak již bylo uvedeno několikrát výše, výsledek z prosté klasifikace obrazu nebyl dostatečně kvalitní, proto musela proběhnout filtrace. Protože se k filtraci využily vrstvy komunikací ZABAGED, zpřesnil se výsledek zejména v intravilánech obcí. Vysokotlak je však z velké části umístěn v oblasti extravilánů obcí. Díky této úvaze bylo rozhodnuto, že výsledky z klasifikace obrazu pro třídy *Asfalt*, *Střecha-taška* a *Střecha-rovná* budou zahrnuty do třídy *Holá půda*. Třída *Asfalt* byla tak u vysokotlaku vytvořena pouze z vrstvy filtru komunikací. Podobně proběhla reklasifikace u vrstvy přípojek, kdy byla třída *Les* zahrnuta do třídy *Travní porost*.

OID	Trída C	Popis C	FREQUENCY	SUM Shape
0	2	Travní_porost	1157	2341,54576256
1	3	Holá_půda	414	2240,27623721
2	4	Asfalt	380	537,176350649
3	5	Střecha_taška	160	130,434127866
4	6	Střecha_rovná	106	134,013216101
5	7	Stín	150	207,990308797
6	8	Hlavní_komunikace	23	547,159974066
7	9	Místní_komunikace	60	2289,26853083
8	10	Cesta	3	289,255836276

Obr. 23 – Ukázka DBF tabulky pro část obce u hlavního řadu

Upraveným vrstvám oklasifikovaného plynovodu chyběl v této fázi poslední atribut, a to příslušnost k dané části obce. Proto bylo potřeba podle masek hranic částí obcí tento výsledek rozřezat a doplnit potřebný atribut. Následně se počítala statistika pro jednotlivé části obce do tabulek DBF. Statistika byla počítána pouze pro nejpodrobnější klasifikační kategorii typu C. Ukázka takové DBF tabulky pro hlavní řad u jedné z částí obce je na obr. 23. Ve sloupečku *SUM_Shape* je součet délek oklasifikovaného plynu podle dané klasifikační třídy. Součet délek je v jednotkách metrů.

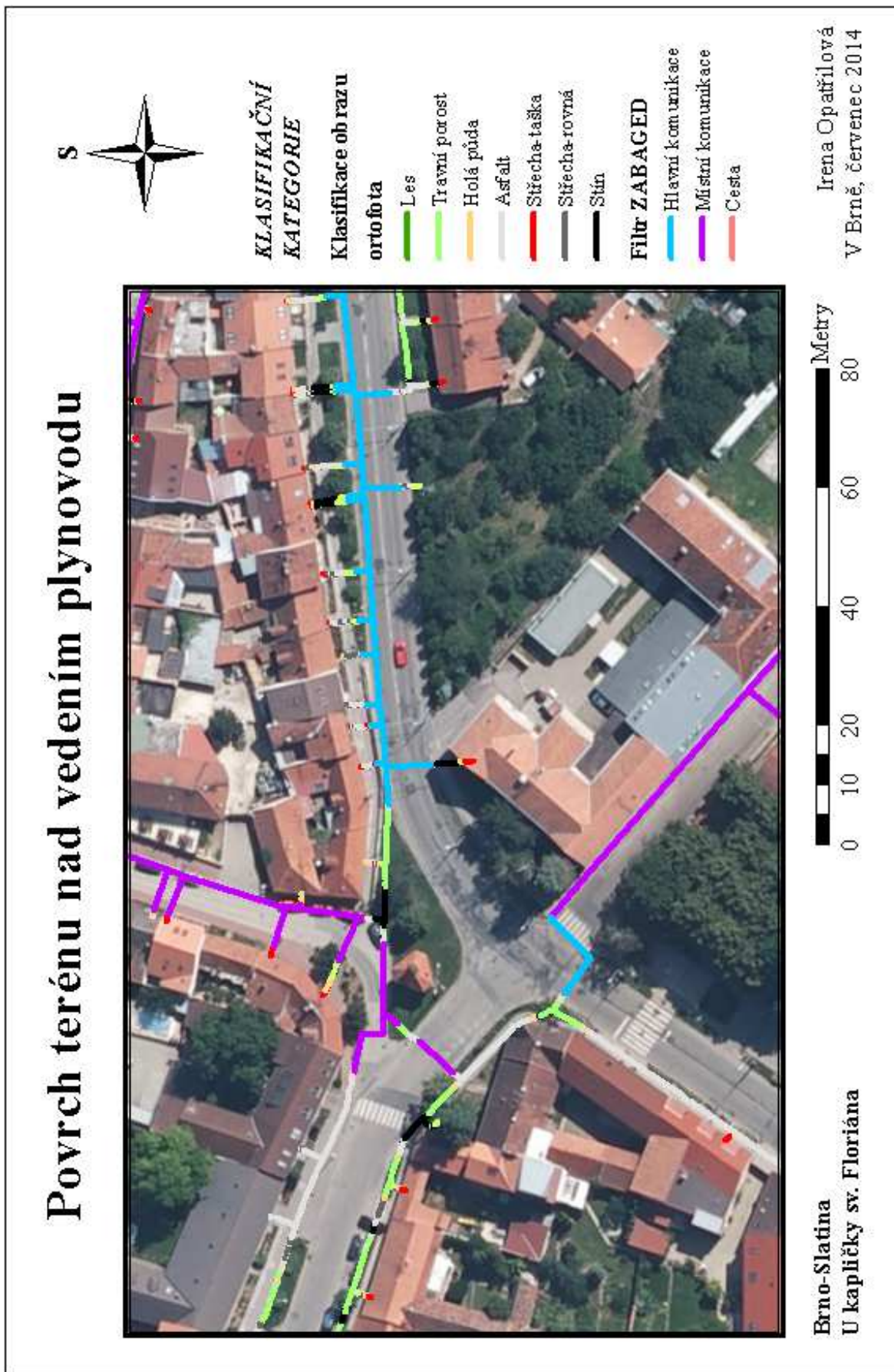
Na závěr se vytvořil grafický výsledek ve formátu SHP s oklasifikovaným průběhem trasy plynovodu pro celé ORP. Ukázka grafického výstupu v SHP na podkladě ortofoto včetně legendy je na obr. 24.

5.4 *Kontrola dat*

Kontrola dat byla přítomna ve všech fázích úlohy. Při přípravě dat se jednalo o vizuální kontrolu správně naplněné šablony, vytvořených bufferů komunikací vůči skutečným šířkám, či správně vytvořených mozaik pro stejné barevnosti ortofoto. Nejdůležitější část kontroly však byla po hlavním zpracování. Ta se skládala ze dvou fází, a to z kontrolních součtů délek plynu v DBF tabulkách a vizuální kontroly grafického výstupu v ArcMap.

Při hlavní analýze se v závěru procesu nevytvářely DBF tabulky pouze pro části obcí, ale byly počítány délkové statistiky také pro výstupy oklasifikovaného plynu pro celé ORP před rozřezáním na části obcí a po znovu sjednocení, dále pak statistiky pro celkové součty mezi všemi klasifikačními kategoriemi A, B a C a neoklasifikovaným plynem za ORP. Samotná kontrola probíhala jednoduše, vizuálně se shlédly hodnoty v tabulkách, které se musely navzájem rovnat.

Povrch terénu nad vedením plynovodu



Obr. 24 – Grafický výstup ve formátu SHP oklasifikovaného záznamu trasy plynovodu

Nejpracnější část kontroly byla vizuální kontrola oklasifikovaného plynovodu nad podkladem ortofoto v prostředí ArcGIS. Pokud byla nalezena chyba ve výstupu hlavního řadu, která vznikala zejména z důvodu nestejně barevnosti ortofota a ne vždy kvalitní citlivosti klasifikačního klíče nad některými oblastmi, bylo potřeba provést v oblastech ruční editaci výstupu (např. místo asfaltu byla ve městech půda, protože jsou obě dvě klasifikační třídy spektrálně blízké). Po editaci došlo k automatizovanému přegenerování výsledku a letmé vizuální kontrole. Výstupy vysokotlaku a přípojek nebylo zapotřebí po zpracování dat ručně editovat.

5.5 Distribuce výsledků a kompletace výstupu

Pokud byl výstup za ORP zkontrolován a byl korektní, bylo potřeba z něj vyexportovat výsledky do adresáře s výstupy. Automatizovaně se zkopírovaly DBF tabulky jednotlivých částí obcí s informacemi o délkách plynu pro klasifikační třídu C (viz obr. 23). Dále se zkopírovaly grafické výstupy v SHP a všechny textové soubory s protokolem o průběhu procesů u jednotlivých skriptů v Pythonu.

Plynařská společnost požadovala grafické výstupy ve formátu SHP po jednotlivých krajích, odděleně pro hlavní řad, přípojky a vysokotlak. Proto muselo dojít ke sloučení všech SHP za ORP daného typu plynovodu do jednoho shapefile za kraj. Vzniklo tak 36 SHP souborů o celkové velikosti 8,8 GB. Výsledky ze všech DBF tabulek pro části obcí, kterých bylo celkem 39249 (počet typů plynovodu x počet částí obcí = 3 x 13083 = 39249), byly poté automatizovaně přepsány do DBF šablony. Šablona byla upravena tak, aby bylo snadné následně ručně zkopírovat hodnoty do výsledného souboru XLS. Tabulka s daty má 13084 řádků (včetně řádku se záhlavím) a je upravena pomocí kontingenční tabulky tak, aby jí mohl uživatel lehce procházet a filtrovat hodnoty podle různých územních celků od krajů až po jednotlivé části obcí.

Součástí závěrečné zprávy projektu [87] jsou i další výstupy, jako jsou grafy zastoupení zpevněného a nezpevněného povrchu nad plynovodem pro jednotlivá ORP po krajích a dále souhrnné grafy výsledků za kraje pro celé zpracované území.

5.6 Stanovení chybovosti výstupu analýzy

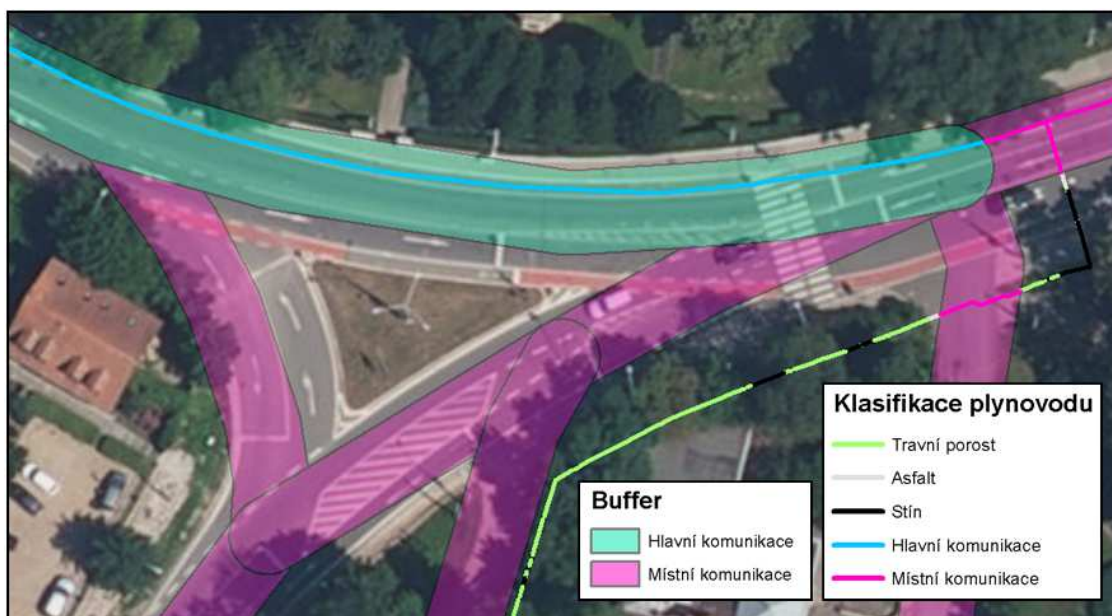
Metoda pro vyhodnocení chybovosti klasifikace je popsána v podkapitole 4.6. Cílem projektu bylo klasifikovat povrch nad liniovým průběhem vedení plynovodu s maximální chybovostí do 5 %. Hlavní metodou řešení byla řízená klasifikace rastrového obrazu metodou Maximum Likelihood v prostředí ArcGIS, která vykazovala průměrnou úspěšnost cca 72 %, což bylo pro potřeby projektu nedostačující. Předběžný průzkum ukázal, že největší chybovost rastrové klasifikace obrazu je v místě komunikací, kde okolní vegetace nebo budovy vrhají stín na zemský povrch, který se tak stává neklasifikovatelný a tím je zkreslen výsledek klasifikace. Proto byl výsledek automatizované klasifikace zpřesněn cca o 25 % následnou reklasifikací použitím filtru vektorové datové sady komunikací z databáze ZABAGED. Výsledek byl dále zpřesněn přetříděním klasifikovaných prvků v extravilánu a

intravilánu pomocí vzájemných logických podmínek klasifikovaných skupin implementovaných do účelově vytvořené aplikace v jazyku Python (úprava atributové tabulky výstupu). Poté proběhla vizuální kontrola výsledků v mapovém okně ArcGIS, ruční editace chybných úseků u hlavního řadu a přegenerování výsledků. Na závěr bylo provedeno stanovení chybovosti klasifikace.

5.6.1 Analýza příčin chybovosti výsledku datové analýzy

Mezi příčiny chybovosti výsledku patří následující:

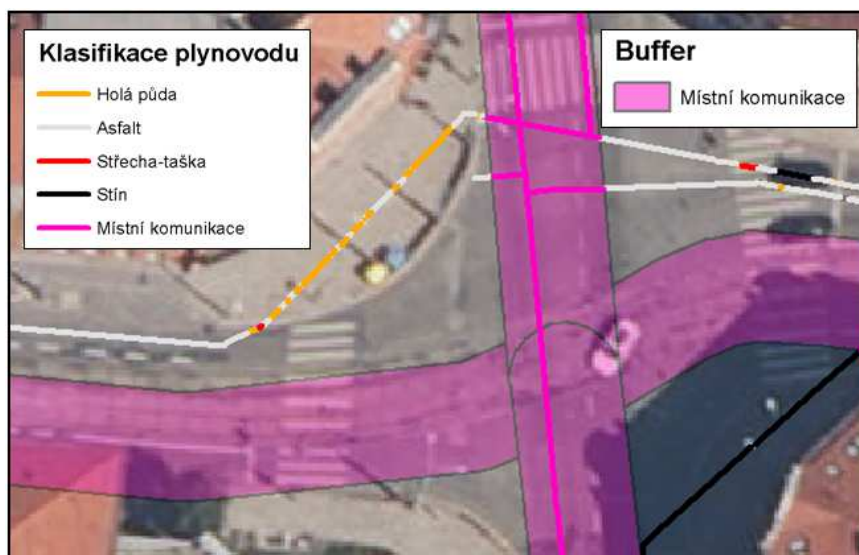
1. *různá kvalita barevnosti ortofota* (nejednotnost barev ortofota) – na obr. 15 jsou vyznačeny oblasti různé (nejednotné) barevnosti ortofota na části území ČR. Důsledkem toho nebylo možné použít jednu společnou trénovací množinu k tvorbě univerzálního klasifikačního klíče. Proto bylo nutné rozdělit ortofoto na území ČR do 20ti oblastí podobnosti barevnosti (obr. 13), v nich pak byly vytvořeny trénovací množiny a odvozeny klasifikační klíče. Vyhodnocení chybovosti v závislosti na barevnosti ortofota je v podkapitole 5.6.4.
2. *variabilita šířek komunikací stejného typu* – tento problém se objevil např. u změny šířek v intravilánu vůči extravilánu, u rozšíření komunikací apod. Pro každý typ komunikace byla určována přímým měřením obrazu její průměrná šířka v rámci řešené územní datové jednotky, která pak definovala jednotný šířkový buffer. Přesto se zde vyskytla místa, kde buffery šířek komunikací nepokrývaly jejich celou šířku, naopak někde zasahovaly do vegetace (půdy), viz obr. 25. Na obr. 26 je rovněž patrný problém variability šířky komunikací, které neodpovídá jednotná šířka bufferu.



Obr. 25 – Ukázka bufferů pro hlavní (modrá) a místní komunikace (fialová) + zakres průběhu plynovodu

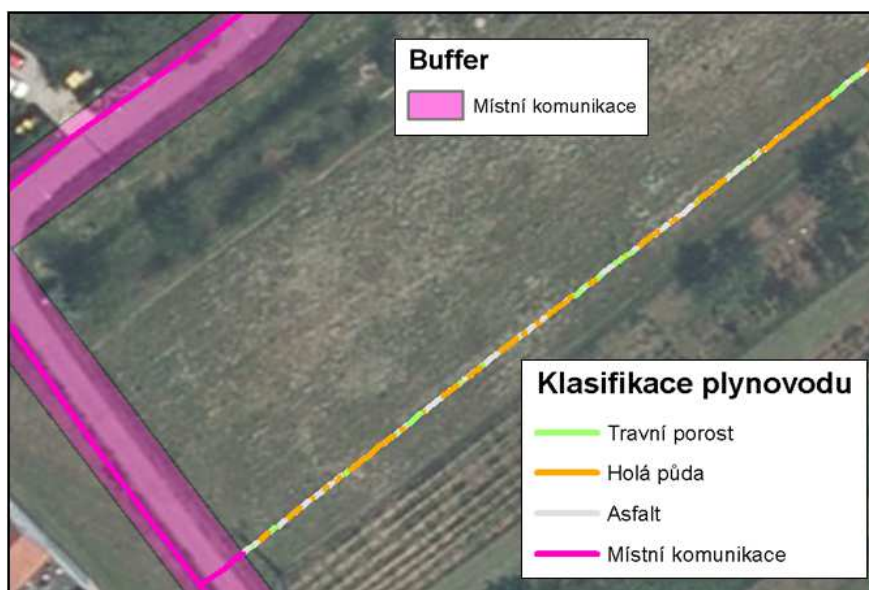
3. *variabilita barevnosti téhož povrchu* – tento problém se objevoval při klasifikaci obrazu u spektrálně blízkých tříd. Největším problémem byla záměna mezi klasifikačními třídami typu půda a asfalt, protože se zde vyskytla chyba v základním

členění povrchu na zpevněný a nezpevněný. Mezi nejrizikovější oblasti tak patřily dlažby chodníků a náměstí v intravilánech obcí, anebo v extravilánech rozsáhlé zemědělské plochy s plodinami světlých barev připomínající asphalt. Na obr. 26 je ukázka chybně klasifikovaného povrchu v intravilánu, kde ve zpevněném povrchu dlažby (šedá barva) je klasifikován povrch půdy jako nezpevněný (oranžová barva). Opačná situace je na obr. 27, kde je ukázka chybně klasifikovaného povrchu v extravilánu – v nezpevněném povrchu pole (zelená nebo oranžová barva) je klasifikován zpevněný povrch typu asphalt (šedá barva).



Obr. 26 – Ukázka zákresu plynu nad zpevněnou plochou

4. *výskyt zastíněných prostor v důsledku vegetace nebo budov* – tyto potíže jsou typické zejména pro oblasti intravilánu. Výsledkem klasifikace obrazu jsou zde pak neidentifikovatelné místa z hlediska typu povrchu terénu. Ukázky těchto míst jsou na obr. 25 a 26.



Obr. 27 – Zákres plynového vedení v extravilánu

5. *nesoulad průběhu komunikací v ortofoto a v ZABAGED* – problém vzniká neaktuálností průběhu os komunikací datové sady ZABAGED v některých úsecích komunikací s průběhem v ortofoto (obr. 25 a 26).

5.6.2 Vyhodnocení zpřesnění výsledků klasifikace obrazu

Vyhodnocení úspěšnosti klasifikace obrazu a jejího zpřesnění bylo provedeno pro tyto výstupy klasifikace (viz podkapitola 4.5):

1. výstup z prosté klasifikace obrazu v ArcGIS
2. výstup z prosté klasifikace obrazu + doplňující filtrace datovou sadou ZABAGED
3. výstup prosté klasifikace obrazu + filtrace datovou sadou ZABAGED + ruční editace v ArcGIS.

Pro každé ORP byly vypočítány rozdíly mezi těmito skupinami výstupů u plynovodu typu hlavní řad. Statistické vyhodnocení výsledků je v následujících tabulkách. Sledovány byly ukazatele klasifikovaných povrchů zpevněný, nezpevněný a stín. Referenčním ukazatelem byly hodnoty nejpřesnější (finální) klasifikace typu 3 (klasifikace obrazu + filtrace přes ZABAGED + ruční editace). U dílčího typu výstupu 1 (tab. 8) a výstupu 2 (tab. 9) byla vyhodnocena procentuální hodnota chybovosti od referenční hodnoty vzhledem k referenčnímu typu výstupu 3.

Z tab. 8 je vidět efektivnost zavedení filtru ZABAGED, kterým se výsledky výstupu 2 zlepšily v průměru o 25 % pro zpevněný povrch a o 21 % pro nezpevněný povrch proti výsledkům výstupu 1 (bez filtrace, pouze klasifikace obrazu). Ruční editací bylo z výstupu 2 odstraněno už jen cca 1 % chyb. Stín je nerozlišitelná kategorie z hlediska určení povrchu terénu. V tabulkách jsou uvedeny standardní odchylky vyhodnocené ze souboru 188 území ORP a dále jsou uvedeny i maximální hodnoty chybovosti v rámci datové analýzy u konkrétních ORP.

Chybovost výstupu 1 (klasifikace obrazu) [%]			
Klasifikační kategorie	Zpevněný povrch	Nezpevněný povrch	Stín
Průměr	25,4	21,0	4,5
Standardní odchylka	9,3	8,7	2,6
Maximální odchylka	48,5	45,3	15,9

Tab. 8 – Chybovost výstupu klasifikace obrazu

Chybovost výstupu 2 (klasifikace obrazu + filtrace) [%]			
Klasifikační kategorie	Zpevněný povrch	Nezpevněný povrch	Stín
Průměr	1,1	0,7	0,4
Standardní odchylka	2,4	2,3	0,5
Maximální odchylka	10,8	10,5	3,4

Tab. 9 – Chybovost výstupu klasifikace obrazu + filtrace

5.6.3 Vyhodnocení absolutní chybovosti analýzy

Tabulka 10 uvádí absolutní chybovost analýzy dat určenou statistickým vyhodnocením dle vztahu (32) z výběrového souboru pokrývajících 21 % řešeného území (tj. 39 vybraných ORP). Vzorky ORP byly vybírány účelově pro podchycení výše uvedených příčin chybovosti, zejména různobarevnosti ortofoto. Chybné úseky špatně oklasifikovaného plynovodu byly proměřovány ručně v prostředí ArcGIS mezi kategoriemi povrchu zpevněný, nezpevněný a stín.

Typ plynovodu	Absolutní chybovost analýzy [%]		
	Hlavní řad	Přípojky	Vysokotlak
Průměr	1,9	1,5	0,5
Standardní odchylka	1,4	1,1	0,4
Maximální odchylka	5,4	4,6	2,6
Minimální odchylka	1,4	1,1	0,4

Tab. 10 – Absolutní chybovost analýzy

Průměrná chybovost datové analýzy v dílčích kategoriích nepřekročila 2 %. Očekávaně největší chybovost vykazuje plynovod v kategorii hlavního řadu, která zasahuje do intravilánu i extravilánu. Maximální chybovost vzorku nepřekročila 5,4 %.

5.6.4 Vyhodnocení chybovosti v závislosti na barevnosti ortofoto

Jak bylo uvedeno výše, vzorky ORP pro určení absolutní chybovosti analýzy byly vybírány účelově. Výběr byl podmíněn tomu, aby vzorky reprezentovaly jednu z těchto zvolených kategorií:

- krajské město
- ORP, kde byly tvořeny trénovací množiny a spočítán klasifikační klíč pro danou oblast barevnosti ortofoto
- ORP v rámci jedné oblasti barevnosti, které bylo nejvzdálenější od místa, kde byl vytvořen klasifikační klíč
- ORP, které pokrývá více oblastí barevnosti ortofoto.

Statistické výsledky analýzy chybovosti v závislosti na těchto čtyřech kategoriích vzorků byly zpracovány pro hlavní řad a jsou uvedeny v tabulce 11, konkrétně pro 3 skupiny chybovosti. Změny po ruční editaci reprezentují chybovost výstupu 2, tj. zpřesnění výsledku automatizované analýzy dat manuálním zásahem (viz tab. 9). Absolutní chybovost analýzy vyjadřuje chyby celkové analýzy dat (automatizace + ruční editace) a tyto hodnoty byly určeny ručním odměřováním chybných úseků oklasifikovaného plynovodu (viz tab. 10). Chybovost automatizace zpracování je získána součtem změn po ruční editaci a absolutní chybovosti v daném ORP a vyjadřuje tak přesnost řešení úlohy bez zásahu ruční kontroly a úpravy výsledků.

Z tabulky vyplývají následující závěry. Průměrné výsledky v rámci jedné skupiny chybovosti jsou v rozmezí do 1,6 %, tzn., že významná závislost přesnosti analýzy dat na

charakteru ORP se nepotvrdila. Chybovost automatizace zpracování dat měla očekávanou největší hodnotu v případě kategorie ORP, které bylo nejvzdálenější od vytvořeného klasifikačního klíče, tj. 4,6 %. Pro vybrané vzorky ORP byla průměrná vzdálenost od místa tvorby klíče 60 km, nejvzdálenější ORP bylo ve vzdálenosti 90 km. Naopak nejlepší přesnost automatizace analýzy dosáhla kategorie krajských měst, kde zřejmě hrála velkou roli filtrace pomocí komunikací ZABAGED.

		Typ ORP			
		Krajské město (10 ORP)	Tvorba klíče (10 ORP)	Nejvzdálenější od klíče (10 ORP)	Více klíčů/oblastí barevnosti (9 ORP)
Změny po ruční editaci [%]	Průměr	2,2	1,7	2,7	1,5
	Standardní odchyška	1,9	2,1	2,0	2,8
	Maximální odchyška	6,0	6,6	7,0	9,4
Absolutní chybovost analýzy [%]	Průměr	1,0	2,2	1,9	2,6
	Standardní odchyška	0,8	1,3	1,3	1,4
	Maximální odchyška	2,9	5,4	5,2	5,0
Chybovost automatizace zpracování [%]	Průměr	3,2	4,0	4,6	4,1
	Standardní odchyška	1,9	2,1	2,3	2,7
	Maximální odchyška	6,3	7,4	8,9	10,8

Tab. 11 – Chybovost v závislosti na barevnosti ortofoto

5.7 Automatizace zpracování pomocí skriptů v jazyce Python

Jednotlivé procesy úlohy zpracování dat včetně datové přípravy bylo snahou co nejvíce zautomatizovat, a tím maximálně zefektivnit celou práci. Proto bylo vytvořeno celkem 7 skriptů v jazyce Python, které byly využívány v jednotlivých fázích úlohy. Přehled jednotlivých skriptů pro automatizaci úlohy je znázorněn na obrázku 28.

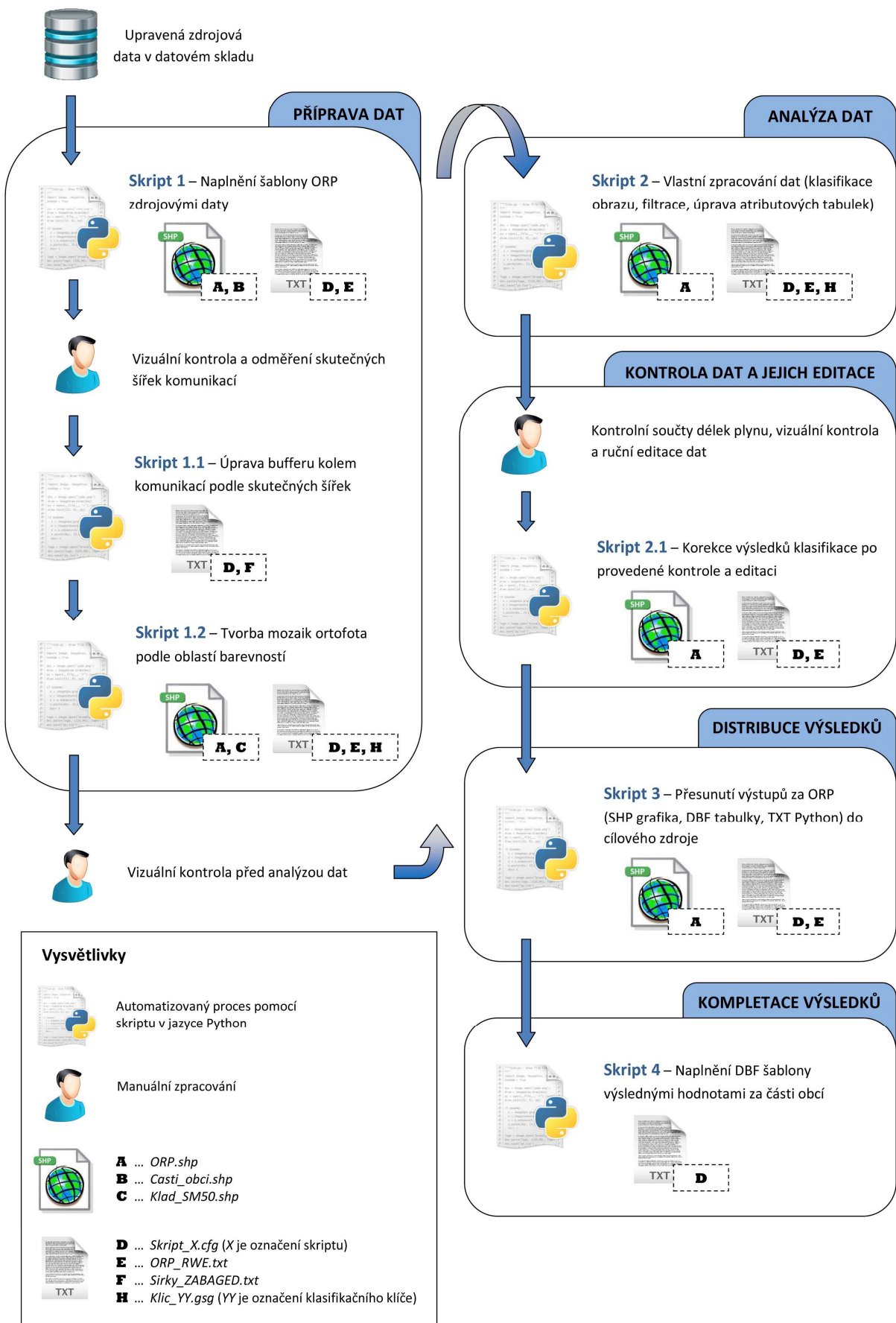
5.7.1 Popis koncepce automatizace procesů

Z hlediska charakteristiky základních vstupních dat, tj. ortofoto s rozlišením 25 cm/pixel na poměrně rozsáhlém území, vede tento proces z pohledu automatizace k tzv. masivním výpočtům. Proto bylo nutné celý proces vhodně rozčlenit, aby zpracování probíhalo efektivně a pod plnou kontrolou.

Proces datové analýzy byl územně rozdělen podle ORP s dalším vnitřním členěním na části obcí (CO). Zpracování probíhalo převážně v plně automatizovaných fázích prokládaných vizuální kontrolou. Automatizovaným nástrojem byl programový komplex ArcGIS. Z hlediska zrychlení se využívalo modulů vytvořených ve skriptovacím jazyce Python s podporou knihoven ESRI pro prostorové operace (geo-processing).

Zpracování dat probíhalo v těchto základních fázích:

1. Automatizované naplnění šablony jednoho projektu ORP upravenými zdrojovými daty – *Skript_1.py*
2. Vizuální kontrola vytvořených projektů v prostředí ArcGIS
3. Úprava šířek komunikací vrstvy ZABAGED – *Skript_1_1.py*



Obr. 28 – Schéma automatizace procesů úlohy s přehledem jednotlivých skriptů

4. Tvorba mozaikových rastrů ortofota podle klasifikačních klíčů – *Skript_1_2.py*
5. Vizuální kontrola před spuštěním analýzy v prostředí ArcGIS
6. Vlastní analýza dat, tj. klasifikace ortofota, filtrace podle komunikací ZABAGED, uložení výsledků do tabulek databáze apod. – *Skript_2.py*. Tento skript je jádrem zpracování celé analýzy.
7. Vizuální kontrola po provedené analýze v prostředí ArcGIS a případná ruční editace výsledku
8. Korekce výsledků klasifikace po vizuální kontrole a ruční editaci – *Skript_2_1.py*
9. Kopie výsledných souborů (SHP, DBF, TXT) z jednotlivých projektů ORP do adresářové struktury výsledků projektů – *Skript_3.py*
10. Zpracování výsledků do souhrnné tabulky DBF v členění podle částí obcí – *Skript_4.py*.

Všeobecné technické informace:

- a) Každý skript se implicitně nalézal v adresáři a spolu se skriptem tam musel být přítomen i jeho příslušný konfigurační soubor stejného názvu s příponou CFG, např. pro skript 1 to byl soubor *Skript_1.cfg*. Pokud tento konfigurační soubor chyběl, skript neproběhl.
- b) Průběh výpočtu včetně časových hodnot byl v důležitých bodech kódu zapisován na obrazovku a paralelně do textového souboru, který měl název *XXXX_Skript_Y.txt*. Zde *XXXX* je kód ORP a *Y* je číslo skriptu. Tento soubor (protokol o výpočtu) byl ukládán do adresáře projektu, kde byl MXD soubor (projekt GIS) pro dané ORP. Soubor se tedy generoval samostatně pro každé zpracovávané ORP.
- c) Jednotlivé ORP bylo možné ve všech skriptech zpracovávat dávkově. Posloupnost ORP bylo možné zapsat do textového souboru *ORP_RWE.txt* – viz dále.
- d) Pro skripty byly nutné řídicí soubory, přičemž jejich potřeba pro daný konkrétní skript byla individuální – viz níže popis jednotlivých skriptů.
- e) Vstupní zdrojová data pro projekty byla uložena v samostatných adresářích, popř. v geodatabázi, odkud se kopírovala do šablony projektů. Vstupní data jsou představena na obr. 12.

5.7.2 Přehled řídicích souborů

Mezi řídicí soubory pro fungování procesů skriptů patřily následující:

- *ORP* (formát SHP) – Číselník ORP, jehož rozsah je 189 řádků tabulky.
- *Casti_obci* (formát SHP) – Číselník částí obcí, jehož rozsah je 13084 řádků tabulky.
- *ORP_RWE* (formát TXT) – Tento textový soubor obsahuje kódy ORP, které mají být zpracovány. Téměř všechny skripty kontrolují validitu zapsaných kódů.
- *Sirky_ZABAGED* (formát TXT) – Textový soubor obsahující šířky komunikací ZABAGED pro jednotlivá ORP (zjištěné v kroku ad. 2 po 1. vizuální kontrole).

5.7.3 Popis jednotlivých skriptů

Následuje stručný popis jednotlivých skriptů prokládaný manuálními pracemi podle hierarchické struktury z podkapitoly 5.7.1. Popis skriptů je členěn do dvou částí:

- a) procesy skriptu
- b) potřebné řídicí soubory pro správný průběh skriptu.

Ad. 1. Skript 1 – Naplnění šablony ORP vstupními daty

a) Procesy skriptu (vývojový diagram je na obr. 29):

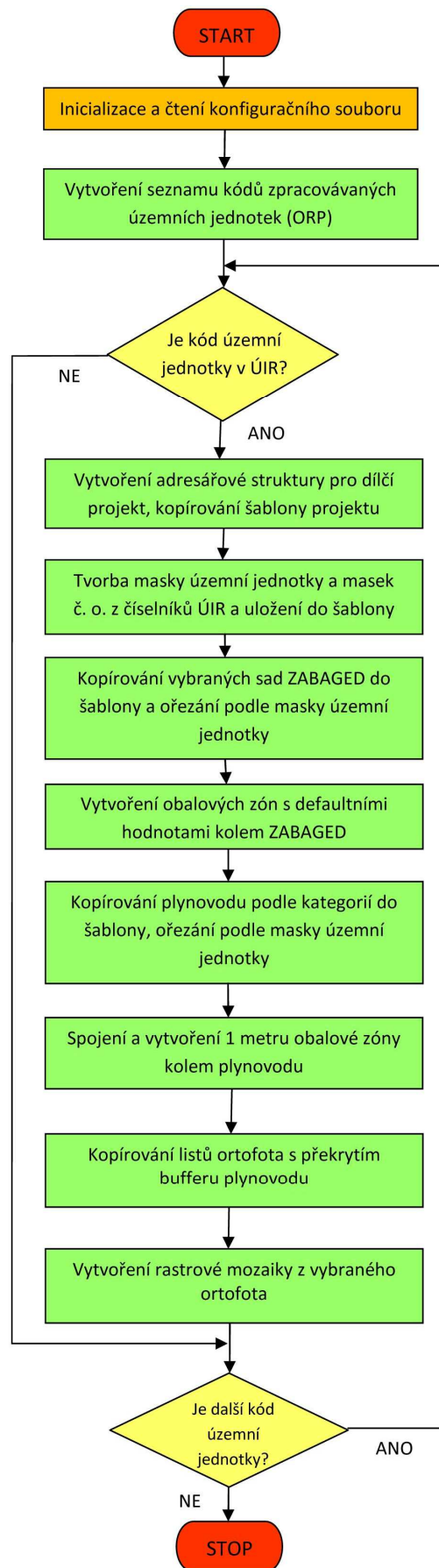
- Vytvoří se hierarchická struktura adresářů na zvoleném externím úložišti v tomto členění: RWE_Analyza – Kraje – Okresy – Obce s rozšířenou působností
- Do posledního podadresáře (tj. adresář ORP) se zkopíruje šablona prázdného projektu a přejmenuje se soubor projektu MXD podle ORP, např. *1317_Brno.mxd*. Název adresáře ORP i soubor MXD mají tento tvar: *kód_název* (bez diakritiky).
- Vytvoří se hranice masky pro ORP a hranice masek pro jednotlivé části obcí
- Do takto vzniklé struktury dojde ke zkopírování a zpracování těchto datových souborů:
 - Upravené vrstvy komunikací z databáze ZABAGED se zkopírují a oříznou podle hranice ORP. Kolem komunikací se vytvoří obalová zóna (buffer) o šířce podle implicitního nastavení, které je součástí skriptu.
 - Zákresy s průběhy plynovodu (hlavní řad, přípojky, vysokotlak) se zkopírují a oříznou podle hranice ORP. Spojí se tři typy plynovodu do jednoho a kolem něj se vytvoří obalová zóna o šířce 1 m.
 - Mapové listy ortofoto s rozlišením 25 cm/pixel, které pokrývají oblasti se zgeneralizovaným průběhem plynovodu, včetně obalové zóny 1 m kolem plynovodu v daném ORP, se zkopírují do šablony.
 - Z vybraných listů ortofota se vytvoří prvotní mozaika rastru, která slouží jako podkladová vrstva pro kontrolu v další fázi. Mozaika je uložena do geodatabáze s názvem *Mozaika_Ortofoto*.

b) Potřebné řídicí soubory:

- *Skript_1.cfg* – konfigurační textový soubor se všemi potřebnými cestami (zdroje dat, cílové umístění projektů, pracovní adresář, cesta k listům ortofota apod.)
- *ORP_RWE.txt* – textový soubor s kódy ORP, které je potřeba zpracovat
- *ORP.shp* a jeho přidružené soubory – upravený shapefile s údaji o ORP (číselník ÚIR)
- *Casti_obci.shp* a jeho přidružené soubory – upravený shapefile s údaji o částech obcí (číselník ÚIR).

Ad 2. Kontrola č. 1 – Vizuální pro ORP prostřednictvím MXD obce

Kontrolou se zjišťuje, zda byla všechna vstupní data pro dané ORP zkopírována. Dále je zkoumáno nad podkladem mozaiky z ortofota, zda se shodují implicitně nastavené šířky komunikací ZABAGED se skutečným stavem. Pokud je zjištěn nesoulad, ručně se odměří skutečné hodnoty šířek v prostředí ArcGIS a zapíše se do textového souboru. Následně se spustí skript 1.1.



Obr. 29 – Algoritmus předzpracování vstupních dat projektu

Ad 3. Skript 1.1 – Přegenerování šířek komunikací

a) Procesy skriptu:

- Kontroluje se, zda pro zadaná ORP existují příslušné projekty vytvořené skriptem 1
- Smažou se původní obalové zóny kolem komunikací ZABAGED s implicitními hodnotami šířek
- Vytvoří se nové buffery podle odměřených skutečných šířek komunikací.

b) Potřebné řídicí soubory:

- *Skript_1_1.cfg*
- *Sirky_ZABAGED.txt* – textový soubor s odměřenými hodnotami bufferů pro jednotlivé třídy komunikací pro dané ORP.

Ad 4. Skript 1.2 – Tvorba mozaik ortofota podle oblastí barevnosti (počtu klíčů)

a) Procesy skriptu:

- V případě jediného klasifikačního klíče dojde k přejmenování vrstvy *Mozaika_Ortofota* v geodatabázi projektu na *MozaikaXX*, kde *XX* je číslo příslušného klíče. Tento klíč je pak zkopírován ze vstupních dat do podadresáře *01_Zdrojova_data*.
- Pokud pro dané ORP existuje několik klíčů, vytvoří se tolik adresářů, kolik je klíčů a do každého tohoto adresáře se zkopírují příslušné listy ortofota. Z jednotlivých ortofot v adresářích se vytvoří mozaiky a pojmenují se *MozaikaXX*, kde *XX* je číslo klíče. Mozaiky se uloží do geodatabáze *02_GDB.gdb*. Dále se zkopírují všechny potřebné klasifikační klíče ze vstupních dat do podadresáře *01_Zdrojova_data*. Nakonec se vymaže původní adresář s ortofoty.

b) Potřebné řídicí soubory:

- *Skript_1_2.cfg*
- *ORP_RWE.txt*
- *ORP.shp* a jeho přidružené soubory
- *Klad_SM50.shp* a jeho přidružené soubory – Tento shapefile tvoří číselník kladu listů v měřítku 1: 50 000 a obsahuje informaci o přiřazení klasifikačních klíčů k jednotlivým mapovým listům. Jeden mapový list SM50 má přiřazen jediný klasifikační klíč.

Ad 5. Kontrola č. 2 – Vizuální pro ORP prostřednictvím MXD obce

Kontroluje se přegenerování bufferů komunikací a stav mozaikových vrstev pro vybraná ORP před spuštěním analýzy v prostředí ArcGIS.

Ad 6. Skript 2 – Vlastní analýza dat

a) Procesy skriptu:

- Vytvoří se filtr sjednocením vrstev ZABAGED

- Dojde ke klasifikaci obrazu podle klíče/klíčů v rámci masky tvořené obalovou zónou 1 metru kolem plynovodu. Pokud je klasifikačních klíčů více než jeden, spojí se výsledné vrstvy klasifikace do jediného rastru s názvem *Klasifikace*.
 - Reklasifikace formou filtrace sjednocenými vrstvami ZABAGED
 - Doplní se atribut příslušnosti k územní jednotce kraje, okresu a ORP
 - Dojde k průniku klasifikované vrstvy se zákresem plynovodu
 - Výsledek se zpracuje do podrobnosti části obce, tzn., ořízne se klasifikovaný plynovod podle hranice části obce, doplní se atribut příslušnosti k územní jednotce části obce a spočítají se statistiky délek plynovodu do tabulkové podoby formátu DBF pro všechny typy plynovodů v členění na dohodnuté klasifikační kategorie druhů povrchů (A, B, C)
 - Spojí se výsledky za části obcí a vygeneruje se výsledek za celé ORP ve formátu SHP
 - Spočítají se závěrečné statistiky a kontrolní součty.
- b) Potřebné řídicí soubory:
- *Skript_2.cfg*
 - *ORP_RWE.txt*
 - *ORP.shp* a jeho přidružené soubory.

Ad 7. Kontrola č. 3 – Vizuální pro ORP prostřednictvím MXD obce, kontrolní délky plynovodu

Kontroluje se stav po klasifikaci, zda výsledek odpovídá skutečnosti v místech se zákresem plynovodu nad podkladem ortofoto (především u hlavního řadu). Zjištěné nedostatky jsou u hlavního řadu ručně opraveny a výsledky korigovány následujícím skriptem 2.1. Dále revizí projdou kontrolní součty délek plynovodů.

Ad 8. Skript 2.1 – Úprava výstupu po ruční editaci

- a) Procesy skriptu:
- Smažou se původní výsledky oklasifikovaného plynovodu pro hlavní řad, tzn. vrstvy plynovodu pro části obcí, výsledný SHP soubor a příslušné DBF tabulky
 - Ručně upravená vrstva průniku plynovodu s filtrovanou vrstvou klasifikace se znovu zpracuje do podrobnosti části obcí
 - Spojí se výsledky za části obcí a vygeneruje se výsledek za celé ORP ve formátu SHP
 - Spočítají se závěrečné statistiky a kontrolní součty.
- b) Potřebné řídicí soubory:
- *Skript_2_1.cfg*
 - *ORP_RWE.txt*
 - *ORP.shp* a jeho přidružené soubory.

Ad 9. Skript 3 – Distribuce výsledků

- a) Procesy skriptu:
- Zkopírují se soubory s výsledky zpracování ORP projektů do předem dohodnutých výstupních adresářů:

- Soubory SHP z adresáře projektu *03_Vystup_ORP* do zadaného adresáře
- Soubory DBF z adresáře projektu *04_Vystup_CO* do zadaného adresáře
- Soubory TXT z kořenového adresáře projektu ORP (kde je soubor MXD) do zadaného adresáře.

b) Potřebné řídicí soubory:

- *Skript_3.cfg*
- *ORP_RWE.txt*
- *ORP.shp* a jeho přidružené soubory.

Ad 10. Skript 4 – Kompletace výsledků

a) Procesy skriptu:

- Zkopírují se hodnoty délek plynovodu pro klasifikační kategorii typu C z jednotlivých DBF tabulek částí obcí do výsledného DBF souboru *Vysledky.dbf*. Tento soubor má charakter šablony ve formě databázové tabulky. Tabulka obsahuje sloupce s kódy částí obcí, ORP, okresů, krajů a dále pak sloupce s výsledky statistik pro hlavní řad, přípojky a vysokotlak pro všechny kombinace s požadovanými druhy povrchů (kategorie A, B, C).
- Z tohoto naplněného souboru se následně výsledky ručně zkopírovaly do sešitu Excel a spolu s upravenými grafickými výsledky ve formátu SHP byly předány zadavateli jako hlavní výstupy zpracování úlohy.

b) Potřebné řídicí soubory:

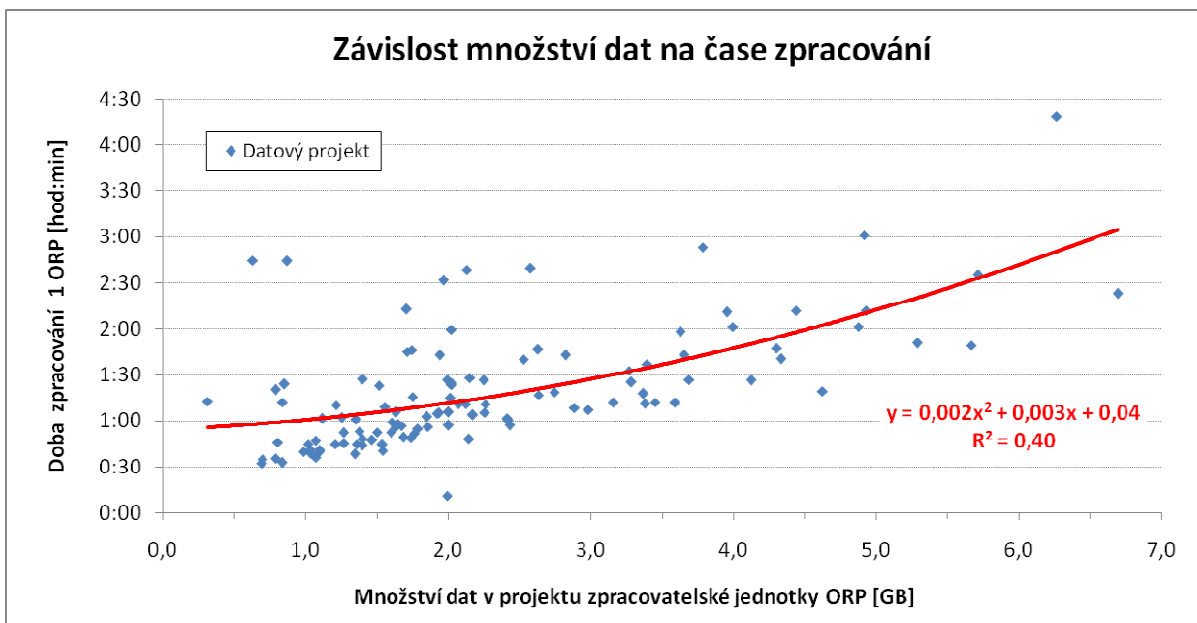
- *Skript_4.cfg*.

5.8 Optimalizace přípravy dat a poměr automatizace k ruční práci

5.8.1 Vyhodnocení optimalizace přípravy dat

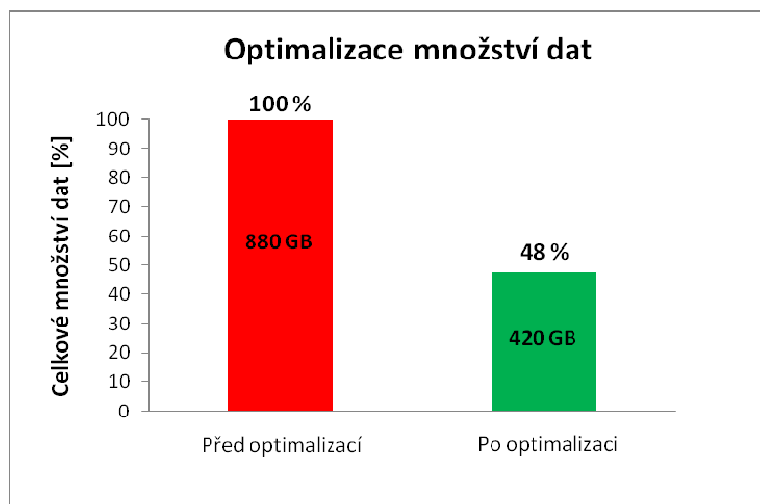
Optimalizovanou územní datovou jednotkou bylo území ORP o průměrné rozloze 368 km². Ze zdrojových dat byly automatizovaně vygenerovány pomocí účelové softwarové aplikace dílčí datové projekty pro území ORP. Automatizované vygenerování dílčích datových projektů představovalo nezanedbatelný časový nárok a probíhalo na počítačových strojích typu PC, jejichž hardwarová specifikace je v podkapitole 5.1.3. Na obr. 30 je znázorněna závislost množství dat v GB a času potřebného na vygenerování dílčích datových projektů skriptem 1. Maximální doba zpracování byla 4 hodiny 18 minut, projekt této jednotky měl 6,3 GB. Průměrná doba zpracování jednotky ORP skriptem 1 byla 1 hod 19 min, přičemž průměrný objem dat v jednotce činil 2,2 GB. Celkový objem dat pro celé zpracovávané území byl po datové přípravě 420 GB.

Z obr. 30 je patrná polynomiální závislost druhého stupně mezi množstvím dat a strojčasem potřebným na vygenerování dílčích datových projektů. Spolehlivost regrese vyjádřená koeficientem determinace R^2 je 0,4. Z charakteru hodnot grafu lze odvodit závislost času režie $T_{rež}$ na objemu zpracovávaných dat, která je vyjádřena vzorcem (28).



Obr. 30 – Závislost množství dat a času potřebného na vygenerování dílčích datových projektů

Strojčas je ovlivněn i členitostí průběhu plynovodu, což souvisí s množstvím grafických prvků. Úsporu času lze dosáhnout zmenšením množství dat. Na obr. 31 je znázorněn výsledek optimalizace množství dat tematickou selekcí podle věcného uspořádání podle průběhu vedení plynovodu. Celkové množství dat pokrývající územní jednotku bylo redukováno na data optimálně pokrývající pouze průběh vedení plynovodu. Po této optimalizaci byl snížen objem dat z celkového množství (100 %) na množství 48 %, tedy prakticky o polovinu, což znamená i cca poloviční úsporu strojčasu potřebného na přípravu dat

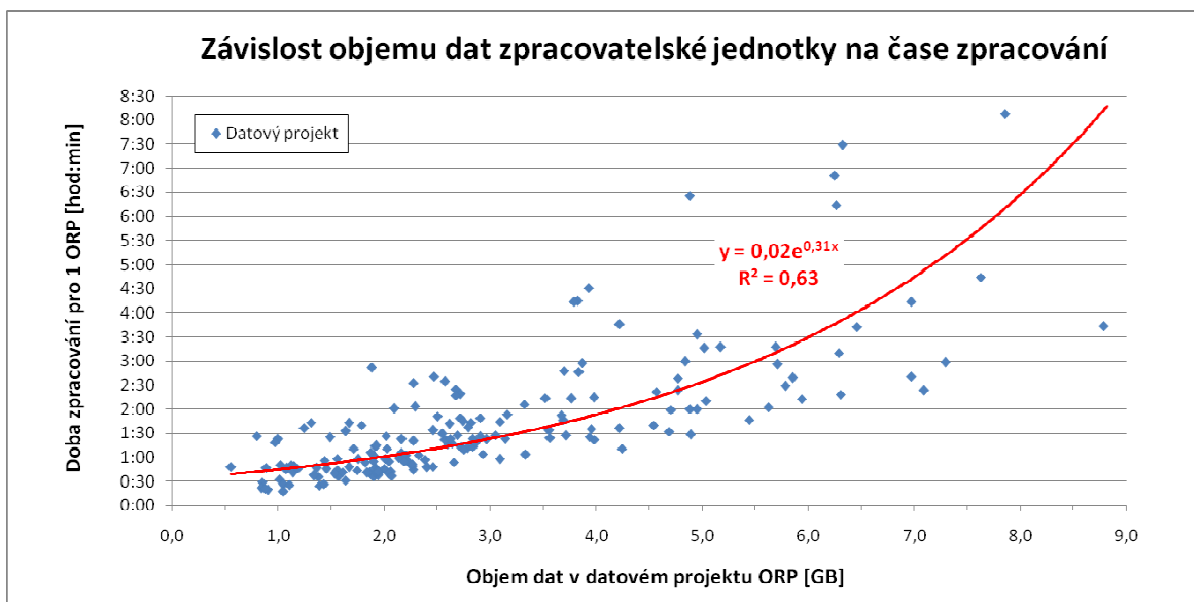


Obr. 31 – Výsledek datové optimalizace u fáze přípravy dat

5.8.2 Vyhodnocení poměru automatizace k ruční práci

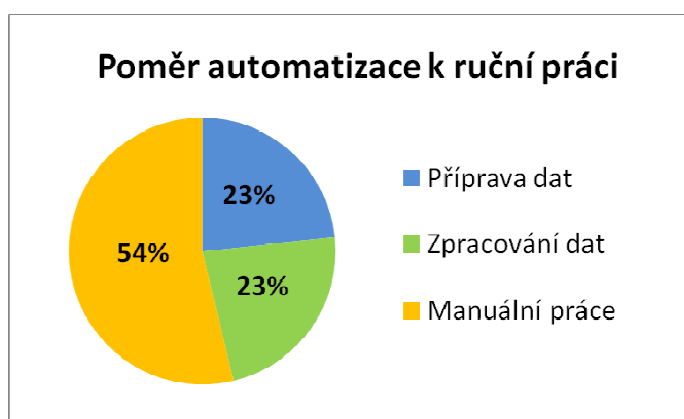
Na obr. 32 je znázorněna závislost objemu dat dílčí zpracovatelské jednotky ORP na čase zpracování skriptem 2, tzn. na čase zpracování hlavní analýzy. Závislost má exponenciální

charakter a vyjadřuje vztah (29), tzn. čas doby processingu T_p . Maximální doba zpracování byla téměř 16,5 hodiny, zpracovaný projekt měl 8,8 GB dat. Druhá nejdelší doba zpracování je již znázorněna v grafu. Jedná se o ORP, které bylo zpracovááno přes 8 hodin s výslednými 7,9 GB dat v projektu. Průměrná doba zpracování 1 ORP skriptem 2 byla 1 hod 47 min, průměrná velikost zpracovaného projektu ORP byla 3 GB. Celkový objem dat pro celé zpracovávané území byl pro hlavní analýze 560 GB.



Obr. 32 – Závislost času processingu T_p na objemu zpracovávaných dat

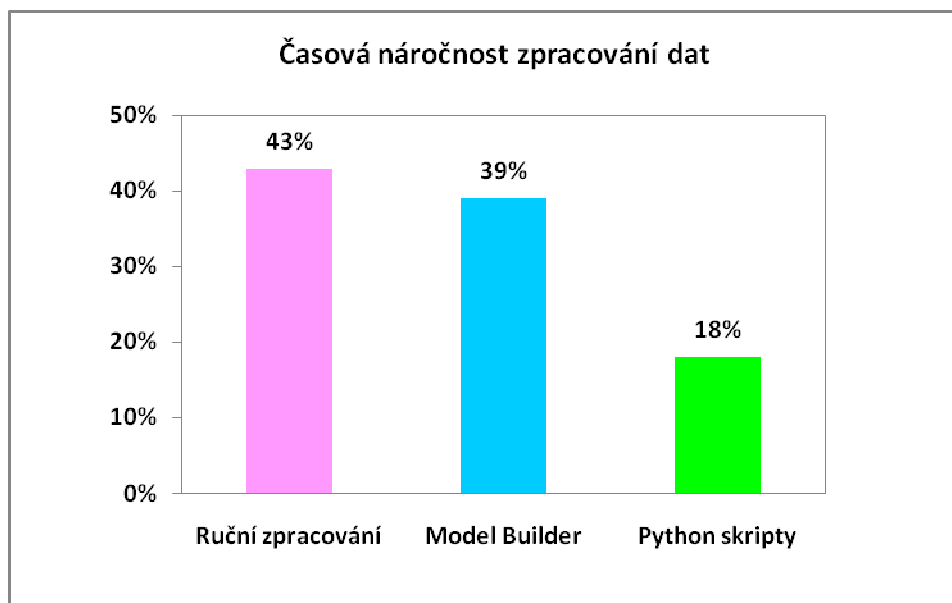
Na obr. 33 je znázorněn procentuální podíl dílčích automatizovaných procesů datové přípravy (T_{Ia}) a zpracování dat (T_{Oa}) a procesu manuální práce ($T_{Im} + T_{Om}$) na celku pro řešené území o celkové rozloze 68312 km², viz vztah (8). Manuální práce trvaly na 1 datové jednotce ORP průměrně 4,5 hodiny, přičemž tyto práce zahrnovaly zejména ruční odměřování skutečných šířek komunikací ve fázi přípravy dat, dále kontrolu a editaci dat po hlavním zpracování.



Obr. 33 – Podíly dílčích automatizovaných procesů a lidské práce na celku

Přestože podíl lidské práce je časově poměrově nadpoloviční, vysoká automatizace přípravy dat a zpracování dat formou Python skriptů vysoce zefektivňuje celý proces vyhodnocení. Při porovnání časové náročnosti výhradně ručního zpracování dat, zpracování

prostřednictvím nástroje Model Builder a zpracování prostřednictvím skriptu Python, je procentuální poměr časové náročnosti znázorněn na obr. 34.



Obr. 34 – Podíly časové náročnosti různých typů zpracování dat

Ručním zpracováním úlohy se myslí postupné manuální zadávání parametrů do nástrojů z ArcToolbox v prostředí ArcMap. Tato forma je zcela bez automatizace a je časově nejnáročnější. Podobně časově náročné je zpracování s využitím aplikace Model Builder v prostředí ArcMap, které může být poloautomatizované. Jednotlivé procesy pomocí Model Builderu trvaly přibližně o čtvrtinu déle, než když běžely při samostatném ručním spuštění. V porovnání s ručním zpracováním je zde však ta výhoda, že parametry jsou již nastaveny a nemusí se manuálně zadávat. Tím se šetří čas. Lidská práce je potřeba pouze u spuštění jednotlivých modelů. Modelů muselo být vytvořeno více, protože zpracování hlavní analýzy bylo velmi složité a vytvořit tak jeden jednodušší kompletní model bylo nemožné. Nejrychlejší, a tím nejefektivnější, je zpracování dat pomocí Python skriptů, které je plně automatizované.

5.9 Úskalí použité technologické linky

Při řešení datové analýzy se ukázala tato technologická úskalí:

- *topologické chyby ve vstupních datech* – chyby se objevily jak v datech zákresu plynovodu v DGN, tak v datech ZABAGED u plošných vrstev. Před zpracování dat je potřeba opravit tyto topologické chyby.
- *nutnost vytvoření číselníku ÚIR* – zdrojová data jednotlivých správních jednotek nebyla ve všech případech tabulkově provázána, navíc některá data obsahovala duplicitní řádky. Tyto chyby je nutné před zpracováním opravit.
- *různá barevnost ortofota* – nutnost ruční tvorby klasifikačních klíčů pro stejné, příp. podobné oblasti barevnosti

- *nepřesnost průběhu komunikací v ZABAGED, popř. neaktuálnost dat – odlišný průběh os dálnic a rychlostních komunikací od obrazu ortofota, nevhodný zákres průběhu os komunikací v rozšíření na víceproudové úseky*
- *nutnost manuální parametrizace šířek komunikací nad ortofoto a ZABAGED*
- *členitost vedení plynovodu vede k vyšším nárokům na čas a počet operací – vyžaduje použití výkonnějšího hardware s nutností použití speciálních instrukcí na rozčlenění na dílčí části (nastavení limitů geometrických elementů, ze kterých se skládá, zvýšení počtu geometrických segmentů)*
- *technologie datové analýzy je postavena na řešení v prostředí ESRI ArcGIS*
- *technologie vyžaduje použití výkonné výpočetní techniky*
- *nutnost síťového řešení sdílení dat*
- *relativně vysoký podíl kvalifikované manuální práce – podíl kvalifikované manuální práce s vlivem lidského faktoru na celém projektu vyžaduje časově cca 54 %, automatizovaný proces činí 46 %.*

5.10 Zkušenosti z řešení

V této podkapitole je uveden výčet zkušeností a poznatků, které nebyly doposud zmíněny v disertační práci a ke kterým se dospělo zejména v průběhu řešení pilotního projektu. Zpracování dat bylo v této pilotní fázi řešeno na notebooku značky Toshiba s parametry Intel Core i3, 4 GB RAM, Windows 7 64-bit.

Konverze mezi vektorovými a rastrovými daty

Vektorový zákres plynovodu nebylo možné převést na rastrový soubor do adresáře funkcí *Polyline to Raster* pro velikost rastru menší než 1 metr (např. pro výsledný pixel 0,2 m). Bylo zjištěno, že tento problém lze odstranit tím, že se rastr musí vytvořit do geodatabáze, a ne jako soubor do adresáře.

Pro území velikosti 5 x 4 km se nepodařilo převést rastrový výsledek z klasifikace obrazu pro 20 cm/pixel do SHP souboru typu polygon funkcí *Raster To Polygon*. Přitom pro výstup z klasifikace obrazu pro 50 cm/pixel tento proces konverze trval 4 minuty. Řešení problému bylo takové, že se klasifikace obrazu nebude provádět nad celými listy ortofota, ale pouze nad územím bufferu kolem plynovodu.

Konverze DGN souborů zákresu plynovodu do datasetu geodatabáze při fázi přípravy dat proběhla pomocí funkce *Feature Class To Feature Class*. Při zhlédnutí v ArcMap byly nalezeny chyby z této konverze, konkrétně se u vrstev se zákresem hlavního řádu a vysokotlaku objevily dlouhé i několika kilometrové přímé linie, které byly napojeny na úseky plynovodu. Těchto chyb se objevilo kolem 70-ti a byly ručně odstraněny.

Klasifikace obrazu a její zpřesnění

Výsledek klasifikace obrazu, tedy i kvalita výsledku řešení úlohy, je silně závislý na kvalitě mapových listů ortofota, tj. na barevnosti a rozlišení. Co se týká barevnosti, bylo by

ideální ke každému ortofotu vytvořit vlastní klasifikační klíč. To je ovšem vzhledem k množství dat nereálné. Díky barevnosti nelze spolehlivě zařadit do příslušných tříd některé terénní povrchy. Dále nelze spolehlivě vyhodnotit z ortofoto povrch v obraze stínu (budov, stromů, apod.). Vzhledem k nedostatečnému rozlišení ortofota (25 cm/pixel) nelze identifikovat třídu povrchu typu dlažba (na žádost společnosti RWE měla být původně tato třída povrchu součástí analýzy dat).

Čím je vyšší rozlišení ortofota, tím se sice zvyšuje kvalita analýzy dat, ale i objem zpracovávaných dat a časová náročnost úlohy. Např. klasifikace obrazu pro ORP Brno nad rastrovou mozaikou z ortofota s rozlišením 50 cm/pixel trvala 40 minut (zpracováno 0,9 GB dat), nad mozaikou z ortofoto 25 cm/pixel proces trval 2 hodiny (zpracováno 3 GB dat). Časová náročnost s polovičním rozlišením tak vzrostla o trojnásobek doby trvání.

Pro vybrané lokality byly vytvořeny optimální klasifikační klíče zvláště pro ortofoto 20, 25 a 50 cm/pixel. Experimenty s výsledky klasifikace obrazu nad stejnými lokalitami potvrdily, že není možné vytvořit univerzální klasifikační klíč pro všechna rozlišení ortofota.

Jak již bylo několikrát uvedeno výše, zpřesnění klasifikace obrazu bylo realizováno za pomoci liniových vektorových vrstev ze ZABAGED. Pro řešení analýzy se uvažovalo i o použití plošných vrstev z této databáze, avšak některé vybrané plošné vrstvy měly chybnou topologii, a proto bylo vyloučeno použití prostorových operací typu oříznutí, průnik apod. Z tohoto důvodu nebylo dále zkoumáno využití plošných vrstev ZABAGED pro zpřesnění výsledků analýzy.

Vedení trasy plynovodu přes šikmé zobrazení vysokých budov v ortofotu nelze spolehlivě vyčistit z důvodu neexistence jiných spolehlivých dat. Hledání řešení tohoto problému se jevilo jako neefektivní, vzhledem k zanedbatelné chybovosti výsledku klasifikace v těchto místech.

Výsledek analýzy a jeho přesnost

Analýza byla původně testována pro dvě varianty řešení, a to konkrétně pro vektorové a rastrové řešení. Délky úseků jednotlivých povrchů při rastrovém řešení byly v průměru o 20 % delší ve srovnání s vektorovým řešením. Tento problém vznikal díky systematické chybě způsobené aproximací při výpočtu euklidovské vzdálenosti v rastru. Protože vektorové řešení bylo přesnější, co se týká výpočtu délek, a nevyžadovalo tolik konverzních procedur, bylo vybráno pro další postup. Avšak toto řešení je časově náročnější. Např. průnik (*Intersect*) vektorové vrstvy plynovodu s vektorovým výsledkem klasifikace obrazu (buffer 16 m kolem plynovodu) pro území 5 x 4 km trval cca 1 hodinu pro řešení s ortofotem 25 cm/pixel a 4 minuty pro řešení s ortofotem 50 cm/pixel. Odpovídající funkce průniku v rastrovém řešení (*Combine*), tj. kombinace rastru výsledku klasifikace obrazu v bufferu s rastrovým zákresem plynovodu, trvala pro řešení s ortofotem 25 cm/pixel 34 sekund, pro řešení s ortofotem 50 cm/pixel 9 sekund. Zde je vidět, že rastrové řešení by bylo výrazně rychlejší, avšak se systematickou chybou ve výsledku a že vektorové řešení silně závisí na rozlišení ortofota.

Postupným testováním bylo zjištěno, že přesnost výsledku analýzy je ovlivněna nastavením hodnoty tzv. *XY tolerance*. Tato hodnota stanovuje minimální vzdálenost mezi

souřadnicemi, které v prostředí GIS reprezentují identický bod. Tolerance se nastavuje v geodatabázích při tvorbě nových feature class nebo celých datových sad, u tvorby topologie a dále je jí potřeba nastavit u některých prostorových analýz, jako je např. oříznutí, sjednocení, ztotožnění či průnik. Defaultně je tolerance nastavena na hodnotu 1 mm, tato hodnota však byla pro analýzu nedostatečná. Testováním byla stanovena hodnota tolerance pro všechna možná nastavení při zpracování dat na hodnotu 0,0001 mm.

Interpretace výsledků klasifikace na rozhraní různých povrchů je obtížná. Výsledek závisí na polohové přesnosti podkladů a spolehlivě lze zjistit jen terénním šetřením.

Automatizace procesů

Bylo vyzkoušeno, zda by mohlo běžet na jednom počítači více procesů paralelně, čímž by se mohla výrazně zkrátit doba zpracování dat pro celé analyzované území. Proto byl na počítači spuštěn ve stejnou dobu dvakrát skript 2 pro totožné území, které bylo uloženo na různých místech na disku. Doba zpracování tím nebyla nikterak ovlivněna, obě analýzy trvaly stejně dlouho, jako by byly zpracovávány samostatně. Problémem však bylo, že u jednoho procesu z nevysvětlitelného důvodu nedošlo k vytvoření jedné vrstvy bufferu kolem komunikací. Tato vrstva byla prázdná, a tak neovlivnila filtraci a konečné výsledky. Proto bylo rozhodnuto, že paralelní spuštění skriptů na jednom počítači nebude využíváno, z důvodu nespolehlivosti zpracování dat.

Paralelní zpracování je obecně složité a v případě mnohojádrového procesoru se řeší vláknovým systémem. Vlákno je dílčí ucelený proces, který se chová jako jedna entita. Obecně rozklad procesů do samostatných vláken je velmi obtížně algoritmizovatelná úloha, která se řeší heuristickými metodami. Ukazuje se, že operační systém Windows 7 má prostředky pro řešení tohoto problému, ale při tak dlouhých procesech v řádech několika hodin zřejmě nelze stoprocentně zajistit spolehlivost. Zvláště v případě, kdy jsou spuštěny naprosto shodné procesy, které vyžadují stejné prostředky.

6 ZÁVĚR

Hlavním cílem disertační práce bylo navrhnout a ověřit metodiku řešení masivních úloh v GIS. Ověření metodiky bylo realizováno na výzkumném projektu pro společnost RWE, jehož podstatou byla rozsáhlá datová analýza nad geodaty na území ČR, vyjma dvou krajů. Předmětem projektu bylo navrhnout a ověřit technologii pro klasifikaci povrchů terénu nad plynovodním vedením.

Ze současného stavu řešené problematiky vyplynulo, že mezi nejúčinnější metodu pro řešení složitých úloh v GIS patří paralelismus. Jeho podstatou je rozdělení úlohy do dílčích procesů, které musí být nezávislé a lze je zpracovávat souběžně na několika prostředcích (počítačích). V rámci disertační práce byl vyvinut algoritmus pro optimalizaci úlohy přidělování procesů jednotlivým prostředkům. V práci bylo dokázáno, že v GIS lze tyto dílčí procesy identifikovat snadněji než v jiných obecných systémech, a to na základě územního členění.

Při řešení složitých úloh hraje velmi důležitou roli fáze přípravy dat a její optimalizace. V disertační práci je prezentována jedna z metod optimalizace předzpracování vstupních množin datových sad v rámci rozsáhlých projektů v GIS, která je založená na hierarchickém rozkladu modelovaného území do ekvivalentních tříd. V případě výzkumného projektu se osvědčilo rozdělení zpracovávaného území na datové jednotky o velikosti ORP (obce s rozšířenou působností). [89]

Za základní metodu výzkumného projektu byla zvolena řízená klasifikace obrazu ortofota metodou Maximum Likelihood v systému ArcGIS. Protože výsledky této metody byly nedostatečné, muselo dojít ke zpřesnění výsledku pomocí filtrace vybranými datovými sadami ZABAGED. Výsledky prokázaly, že k efektivnímu dosažení požadované kvality klasifikace pro daný účel stačí použití jedné vhodné doplňkové datové sady pro filtraci mezivýsledků, která zpřesní výsledky automatizovaného procesu klasifikace. [90]

V rámci disertační práce byla navržena metoda, na základě které byla stanovena absolutní chybovost datové analýzy. Hodnoty chybovosti byly určeny statistickým vyhodnocením na výběrovém souboru pokrývající 21 % řešeného území. Podkladem pro vyhodnocení sloužilo ortofoto. Průměrná chybovost datové analýzy v dílčích kategoriích nepřekročila 2 %. Tato hodnota je limitní hranicí informační výtěžnosti z datových zdrojů a získat výsledky analýzy dat s chybovostí pod 2 % je v tomto případě nereálné. Hlavními důvody je nerozlišitelnost správného výsledku klasifikace kvůli atributové a polohové přesnosti podkladů a dále vliv lidského faktoru. Zlepšení výsledků by bylo možné jen terénním šetřením, což bylo k rozsahu projektu nereálné. [91]

Snahou bylo jednotlivé fáze zpracování úlohy co nejvíce automatizovat, a tím zefektivnit celou práci. Vzniklo tak 7 řídicích skriptů v jazyce Python s podporou knihoven ESRI. Výsledky zpracování dat potvrdily skutečnost, že efektivnost datové analýzy na velkém území závisí na míře (stupni) automatizace [89]. V rámci experimentálních výsledků práce byla empiricky stanovena funkční hodnota časové náročnosti přípravy dat a vlastního zpracování [92].

Realizovaná technologie s možností řešení úloh v GIS s velkým objemem dat vedla k uživatelsky přijatelnému řešení. Výsledky zpracovávané analýzy lze charakterizovat těmito základními vlastnostmi:

- technologie umožňuje z rastrového obrazu extrahovat objekty a identifikuje jejich atributy, jejichž hodnoty dále zpracovává. Tím dochází k vytěžení informace z rastrového obrazu, která je dále zpracovatelná.
- datová analýza vykazuje velmi vysokou vypovídací schopnost s nízkou chybovostí
- technologie datové analýzy umožňuje opakovatelnost bez vlivu lidského faktoru
- technologie se vyznačuje rychlostí získání výsledků datové analýzy
- efektivnost reprocessingu se odvíjí od aktualizace klíčových zdrojových sad, např. ortofota nebo od většího množství změn v území nad hodnotu absolutní chybovosti datové analýzy, která činí 2 %
- technologie datové analýzy má do budoucna kvalitativní potenciál dalšího zpřesnění datové analýzy s možnostmi budoucího využití nových, kvalitativně vyšších, zdrojových datových sad.

Výsledky disertace mohou být obecně zhodnoceny podle přínosů pro teorii a praxi:

Za přínos pro teorii je možné považovat:

- koncepce řešení složitých úloh v GIS, využití paralelismu metodou hierarchického rozkladu modelovaného území do tříd na územním a věcném (tematickém) principu
- návrh algoritmu pro optimalizaci úloh s pevným termínem řešení. Algoritmus pracuje na principu přidělování dílčích úloh, na které lze rozdělit hlavní úlohu jednotlivým prostředků, které má uživatel k dispozici.
- návrh hodnocení chybovosti výsledků klasifikace na základě metriky, odvozené ze struktury klasifikačního stromu.

Přínosem pro praxi jsou tyto poznatky:

- ověření technologie zpracování rozsáhlých projektů v GIS na konkrétní úloze
- metoda zpřesnění výsledků klasifikace obrazu použitím relevantních datových sad
- zjištění úskalí a kritických míst v navržené technologické lince způsobených kvalitou vstupních dat, spolehlivostí některých funkcí prostorových analýz v ArcGIS a lidským faktorem
- stanovení reálné limitní hranice kvality výsledků klasifikačního procesu (2 %).

Technologie popsaná v disertační práci má obecný charakter a může být využita pro klasifikaci povrchu nad liniovými inženýrskými sítěmi, jako jsou např. vodovody, produktovody, kanalizace, elektrické energetické rozvody, sdělovací vedení apod. [92]

SOUVISEJÍCÍ PROJEKTY

Pilotní projekt byl řešen v rámci:

- standardního projektu FAST-S-13-2069 Specifického výzkumu pro rok 2013, s názvem *Dolování geo-prostorových dat z disponibilních standardních datových zdrojů prostřednictvím GIS*
- smluvního výzkumu HS123570212101 s názvem *Klasifikace údajů o uložení plynárenských zařízení pod určitými typy povrchů terénu.*

Problematikou tematické oblasti disertační práce se zabývá:

- standardní projekt FAST-S-14-2298 Specifického výzkumu pro rok 2014, s názvem *Řešení masivních úloh v GIS*
- smluvní výzkum HS12357021212200 s názvem *Analýza povrchů nad plynovody RWE na území ČR.*

Autorka disertační práce se na těchto projektech podílela v roli spoluřešitele. Dále byla v rámci doktorského studia zapojena do dalších projektů:

- juniorský projekt FAST-J-12-23 Specifického výzkumu pro rok 2012, s názvem *Návrh rozhraní člověk-stroj pro přístroje k určování polohy pomocí družicových metod* (role hlavní řešitel)
- juniorský projekt FAST-J-13-2108 Specifického výzkumu pro rok 2013, s názvem *Teorie rozhodování a její aplikace v GIS* (role spoluřešitel)
- juniorský projekt FAST-J-11-48/1447 Specifického výzkumu pro rok 2011, s názvem *Verifikace možností mobilního sběru dat v GIS a jeho využití v archeologii* (role spoluřešitel).

SOUVISEJÍCÍ PRODUKTY

Ověřená technologie s názvem *Klasifikace údajů o uložení inženýrských sítí a zařízení pod určitými typy povrchů terénu* z roku 2013 je zapsána v databázi RIV s označením RIV/00216305:26110/13:PR27262.

Ověřená technologie s názvem *Technologie zpracování velkého množství geografických dat*, která souvisí s výše uvedeným smluvním výzkumem *Analýza povrchů nad plynovody RWE na území ČR*, je připravená k podání do databáze RIV.

PŘEHLED POUŽITÝCH ZKRATEK

CO	části obcí
CPU	procesor počítače (angl. Central Processing Unit)
ČR	Česká republika
ČÚZK	Český úřad zeměměřický a katastrální
DGN	vektorový soubor programu MicroStation
DPZ	dálkový průzkum Země (angl. Remote Sensing)
FC	třída prvků (angl. feature class)
GIS	geografický informační systém (angl. Geographic Information System)
GO	geografický objekt
GPU	grafický procesor (angl. Graphics Processing Unit)
HW	hardware
LCTS	angl. Land Cover Time Series
LiDAR	Light Detection And Ranging
ML	Maximum Likelihood
ORP	obec s rozšířenou působností
PCA	Metoda hlavních komponent (angl. Principal Component Analysis)
RGB	spektra barevných pásem červená-zelená-modrá (angl. red-green-blue)
SAR	radar se syntetickou aperturou (angl. Synthetic Aperture Radar)
SHP	souborový formát, tzv. shapefile, firmy ESRI
SM50	Státní mapa 1: 50 000
SVM	angl. Support Vector Machine
SW	software
ÚIR	územní identifikační registr
VGHMÚř	Vojenský geografický a hydrometeorologický úřad
WPS	webová zpracovatelská služba (angl. Web Processing Service)
ZABAGED	Základní báze geografických dat

SEZNAM OBRÁZKŮ

Obr. 1 – Graf doby zpracování procesů prostředky.....	24
Obr. 2 – Sloučení závislých procesů a prostředků do složených uzlů.....	25
Obr. 3 – Paralelní zpracování procesů na prostředcích	26
Obr. 4 – Algoritmus optimalizace přiřazení procesů prostředkům	28
Obr. 5 – Klasifikační strom	31
Obr. 6 – Schéma předzpracování vstupních dat projektu.....	33
Obr. 7 – Harmonogram polo-automatických procesů ve velkém GIS projektu.....	36
Obr. 8 – Blokové schéma klasifikace povrchů.....	37
Obr. 9 – Postupné zpřesňování výsledků klasifikace obrazu ortofota.....	37
Obr. 10 – Schéma zapojení a parametry strojů v počítačové síti	43
Obr. 11 – Fáze pracovního postupu řešení úlohy	44
Obr. 12 – Struktura uložení upravených zdrojových dat.....	45
Obr. 13 – Oblasti barevnosti ortofota podle kladu mapových listů SM50.....	47
Obr. 14 – Ukázka vyhodnocení rozptylogramů pro vybrané trénovací množiny	48
Obr. 15 – Rozmístění trénovacích množin v oblastech barevnosti ortofota.....	49
Obr. 16 – Zjednodušené schéma řešení přípravy dat pro analýzu.....	50
Obr. 17 – Stromová struktura adresářů Plzeňského kraje včetně detailní struktury projektu ORP Rokycany po fázi naplnění šablony zdrojovými daty.....	52
Obr. 18 – Stromová struktura projektu ORP Rokycany po fázi hlavní analýzy dat.....	52
Obr. 19 – Počet potřebných klasifikačních klíčů pro jednotlivá ORP.....	53
Obr. 20 – Schéma řešení analýzy na základě klasifikace obrazu ortofoto a filtrace vrstvami ze ZABAGED	54
Obr. 21 – Ukázka detailu klasifikace obrazu nad průběhem plynovodu.....	55
Obr. 22 – Filtr vytvořený z vrstev komunikací ZABAGED	56
Obr. 23 – Ukázka DBF tabulky pro část obce u hlavního řadu.....	58
Obr. 24 – Grafický výstup ve formátu SHP oklasifikovaného zákresu trasy plynovodu.....	59
Obr. 25 – Ukázka bufferů pro hlavní (modrá) a místní komunikace (fialová) + zákres průběhu plynovodu	61
Obr. 26 – Ukázka zákresu plynu nad zpevněnou plochou	62
Obr. 27 – Zákres plynového vedení v extravišanu.....	62
Obr. 28 – Schéma automatizace procesů úlohy s přehledem jednotlivých skriptů.....	66
Obr. 29 – Algoritmus předzpracování vstupních dat projektu	69
Obr. 30 – Závislost množství dat a času potřebného na vygenerování dílčích datových projektů.....	73
Obr. 31 – Výsledek datové optimalizace u fáze přípravy dat.....	73
Obr. 32 – Závislost času processingu T_p na objemu zpracovávaných dat.....	74
Obr. 33 – Podíly dílčích automatizovaných procesů a lidské práce na celku	74
Obr. 34 – Podíly časové náročnosti různých typů zpracování dat.....	75

SEZNAM TABULEK

Tab. 1 – Hodnocení metod řešení.....	17
Tab. 2 – Mapa procesů	24
Tab. 3 – Vliv faktorů na dobu řešení úloh s enormním objemem dat	34
Tab. 4 – Určení dob zpracování jednotlivých procesů v závislosti na počtu prvků (ilustrativní část).....	42
Tab. 5 – Topologická kontrola zákresu plynového vedení.....	46
Tab. 6 – Klasifikační třídy pro klasifikaci obrazu ortofota	48
Tab. 7 – Výstup klasifikace druhů povrchů nad průběhem plynového vedení pro klasifikační kategorie A, B a C	57
Tab. 8 – Chybovost výstupu klasifikace obrazu.....	63
Tab. 9 – Chybovost výstupu klasifikace obrazu + filtrace	63
Tab. 10 – Absolutní chybovost analýzy	64
Tab. 11 – Chybovost v závislosti na barevnosti ortofoto	65

LITERATURA

- [1] JUANY, L.; TANG, G.; LIU, X.; SONG, X.; YANG, J.; LIU, K. Parallel contributing area calculation with granularity control on massive grid terrain datasets. *Computers & Geosciences*, 2013, vol. 60, pp. 70-80.
- [2] WEISS, B.; BAILEY, M. Massive Parallel Computing to Accelerate Genome-Matching. *GPU Computing Gems*, 2011, pp. 173-184.
- [3] FORSLUND, D., W.; HANSEN, C.; JUNKER, P., JOHN, W., St.; TENBRINK, S., BRETON, J. High-speed networks, visualization, and massive parallelism in the Advanced Computing Laboratory. *Computing Systems in Engineering*, 1992, vol. 3, no. 1-4, pp. 521-524.
- [4] BARTONĚK, D.; OPATŘILOVÁ, I. Design of Man-Machine Interface for Mobile Mapping. *Advanced Science Letters*, 2014, vol. 20, no. 2, pp. 501-504.
- [5] WANG, L.; TAO, J.; RANJAN, R.; MARTEN, H.; STREIT, A.; CHEN, J.; CHEN, D. G-Hadoop: MapReduce across distributed data centers for data-intensive computing. *Future Generation Computer Systems*, 2013, vol. 29, no. 3, pp. 739-750.
- [6] WEN, J.; MA, Y.; LIU, P.; SUN, S. Distributed multipliers in MWM for analyzing job arrival processes in massive HPC workload datasets. *Future Generation Computer Systems*, 2014, vol. 37, pp. 335-344.
- [7] ZHANG, W.; ZHAO, M.; TU, Z.; LOU, S. Research and Implement of Distributed Nodes Collaboration-Based Management and Publishing Technologies for Massive Geospatial Information. *Proceedings of the 2011 International Conference on Informatics, Cybernetics, and Computer Engineering (ICCE2011)*, 2012, vol. 112, pp. 391-400.
- [8] WU, L.; et all. Key Techniques of Distributed Geospatial Information Operations. *18th International Conference on Geoinformatics*, 2010.
- [9] WU, H.; PAN, M.; YAO, L.; et all. A partition-based serial algorithm for generating viewshed on massive DEMs. *International Journal of Geographical Information Science*, 2010, vol. 21, no. 9, pp. 955-964.
- [10] MIN, J., K.; PARK, H., H.; CHUNG, C., W. Multi-way spatial join selectivity for the ring join graph. *Information and Software Technology*, 2005, vol. 47, no. 12, pp. 785-795.
- [11] YANG, D.; FENG, Y.; JUAN, Y.; HAN, X.; WANG, J.; LI, J. Ad-hoc aggregate query processing algorithms based on bit-store for query intensive applications in cloud computing. *Future Generation Computer Systems*, 2013, vol. 29, no. 7, pp. 1725-1735.
- [12] SAKELLARI, G.; LOUKAS, G. A survey of mathematical models, simulation approaches and testbeds used for research in cloud computing. *Simulation Modelling Practice and Theory*, 2013, vol. 39, pp. 92-103.
- [13] RICHTER, R.; DÖLLNER, J. Concepts and techniques for integration, analysis and visualization of massive 3D point clouds. *Computers, Environment and Urban Systems*, 2013, vol. 45, pp. 114-124.
- [14] WU, J.; et all. The Improvement on Terrain LOD Algorithm Using Quadtree. *International Symposium on Intelligent Information Systems and Applications*, Proceedings, 2009, pp. 91-94.

- [15] PARK, H., H.; MIN, J., K.; CHUNG, C.; W., CHÁNY, T., G. Multi-way R-tree joins using indirect predicates. *Information and Software Technology*, 2004, vol. 46, no. 11, pp. 739-751.
- [16] CHETYRKIN, K., G.; KÚHN, J. K.; STURM, C. Recent progress in computing four-loop massive correlators. *Nuclear Physics B, Proceedings Supplements*, 2007, vol. 164, pp. 203-206.
- [17] CHENG, J.; WANG, W.; GAO, R. Massive RDF Data Complicated Query Optimization Based on MapReduce. *Physics Procedia*, 2012, vol. 25, pp. 1414-1419.
- [18] ARGE, L.; CHASE, J.; S., HALPIN, P.; TOMA, L. Efficient Flow Computation on Massive Grid Terrain Datasets. *GeoInformatica*, 2003, vol. 7, no. 4, pp. 283-313.
- [19] JUANY, B. A Short Note on Data-Intensive Geospatial Computing. *Information Fusion and Geographic Information Systems*, 2011, pp. 13-17.
- [20] LANIG, S.; SCHILLING, A.; STOLLBERG, B.; et all. Towards standards-based processing of Digital Elevation Models for Grid Computing through Web Processing Service (WPS). *Computational Science and Its Applications – ICCSA 2008*, 2008, pp. 191-203.
- [21] LUO, Y., W.; WANG, X., L.; XU, Z., Q. An agent approach to spatial information grid architecture design. *Computing and Informatics*, 2005, vol. 24, no. 2, pp. 201-222.
- [22] BUREŠ, J.; BARTONĚK, D. Application of GIS in Civil Engineering, *4th International Conference on Cartography & GIS*, 2012, ISSN 1314-0604.
- [23] KIDNER, D., B.; RALLINGS, P., J.; WARE, J., A. Parallel Processing for Terrain Analysis in GIS: Visibility as a Case Study. *GeoInformatica*, 1997, vol. 1, no. 2, pp. 183-207.
- [24] CHEN, Z., T. Hierarchy of knowledge in GIS. *Science in China Series E: Technological Sciences*, 2008, vol. 51, no. 1, pp. 6-12.
- [25] ARGE, L. External-memory algorithms with applications in GIS. *Algorithmic Foundations of Geographic Information Systems*, 1996, pp. 213-254.
- [26] ARVANITIS, L., G; RAMACHANDRAN, B; BRACKETT, D., P. Multiresource inventories incorporating GIS, GPS and database management systems: a conceptual model. *International Conference on Integrated Tools for Natural Resources Inventories in the 21st Century Location*, 2000, vol. 28, pp. 89-100.
- [27] NITA, I.; COSTACHIOIU T.; LAZARESCU V. Speed-Up of GIS Processing Using Multicore Architectures. *Computational Science and Its Applications – ICCSA 2011*, 2011, vol. 6783, pp. 293-302
- [28] GUAN, X.; LIESMARS, H., W.; LIESMARS, L., L. A Parallel Framework for Processing Massive Spatial Data with a Split-and-Merge Paradigm. *Transactions in GIS*, 2012, vol. 16, no. 6, pp. 829-843.
- [29] DE ROO, A., P., J., BURROUGH, P., A. Using GIS for Processing Data from the Survey and Monitoring of Pollution. *Soil Monitoring*, 1993, pp. 329-344.
- [30] ZAIMOV, V.; ALEXIEV, T.; SIRMA, A., I. Intelligent decision support based on GIS: The EcoRisk Project. *Geographic Information Systems*, 1994, vol. 884, pp. 237-246.
- [31] SCHOTTEN, K; GOETGELUK, R; HILFERINK, M. Residential construction, land use and the environment. Simulations for the Netherlands using a GIS-based land use model. *Environmental Modeling & Assessment*, 2001, vol. 6, no. 2, pp. 133-143.

- [32] HU, J., J. Visualization of environmental data using GIS, CADD and graphic software: A tool box approach. *Annual Conference and Exposition Location*, 1996.
- [33] WANG, X.; SUN, Z. The Design of Water Resources and Hydropower Cloud GIS Platform Based on Big Data. *Geo-Informatics in Resource Management and Sustainable Ecosystem*, 2013, vol. 399, pp. 313-322.
- [34] CRIFASI, G; GRASSA, F; SCROFANI, M. Validity of Geographic Information Systems (GIS) applied to protected cultivations. *6th International Symposium on Protected Cultivation in Mild Winter Climate Location*, Acta Horticulturae, 2003, no. 614, pp. 41-46.
- [35] KINDLER, A.; BANZHAF, E. Databases, data organization and data processing. *Landscape Balance and Landscape Assessment*, 2001, pp. 47-86.
- [36] KVAMME, K., L. Recent Directions and Developments in Geographical Information Systems. *Journal of Archaeological Research*, 1999, vol. 7, no. 2, pp. 153-201.
- [37] LV, Z.; RÉHMAN, S. U.; CHEN, G. WebVRGIS: A P2P Network Engine for VR Data. *Neural Information Processing*, 2013, vol. 8226, pp. 503-510.
- [38] SAHIL, S.; MANOJ, A.; SEILER, R. Applicability and performance of some similarity metrics for automated image registration. *Conference on Multispectral, Hyperspectral, and Ultraspectral Remote Sensing Technology, Techniques, and Applications Location*, SPIE Proceedings, 2006, vol. 6405.
- [39] MEEHAN, W., B.; BROOK, R., G. GIS in Energy and Utilities. *Springer Handbook of Geographic Information*, 2012, pp. 545-556.
- [40] FREVILLE, A.; PLATEAUB, G. An efficient preprocessing procedure for the multidimensional 0-1 knapsack problem. *Discrete Applied Mathematics*, 1994, vol. 49, no. 1-3, pp. 189-212.
- [41] THIELE, M.; BADER, A.; LEHNER, W. Multi-objective scheduling for real-time data warehouses. *Computer Science - Research and Development*, 2009, vol. 24, no. 3, pp. 137-151.
- [42] ARGE, L.; VENGROFF, D., E.; VITTER, J., S. External-Memory Algorithms for Processing Line Segments in Geographic Information Systems. *Algorithmica*, 2007, vol. 47, pp. 1-25.
- [43] ALVAREZ, J., V., L.; LARRUCE, M., A.; FERNÁNDEZ-CARRIÓN QUERO, S.; JIMÉNEZ DEL VALLE, A. Optimizing the collection of used paper from small businesses through GIS techniques: The Legane's case (Madrid, Spain). *Waste Management*, 2008, vol. 28, no. 2, pp. 282-293.
- [44] POLAT, K. A novel data preprocessing method to estimate the air pollution (SO₂): neighbor-based feature scaling (NBFS). *Neural Computer and Applications*, 2011, vol. 21, no. 8, pp. 1987-1994.
- [45] ZHOU, D.; XU, J.-Ch.; RADKE, J.; MU, L. A spatial cluster method supported by GIS for urban rural classification. *Chinese Geographical Science*, 2004, vol. 14, no. 4, pp. 337-342.
- [46] LIU, J.; CHENA, Y. CQ.; MAISOGB, J.; M., LUTA, G. A new point containment test algorithm based on preprocessing and determining triangles. *Computer-Aided Design*, 2010, vol. 42, no. 12, pp. 1143-1150.
- [47] LI X.-w.; QI Y.-f. A Data Preprocessing Algorithm for Classification Model Based On

- Rough Sets. *Physics Procedia*, 2012, vol. 25, pp. 2025-2029.
- [48] CANNATARO, M.; GUZZI, M.; MAZZA, P., H.; TRADIGO, T.; VELTRI, G., P. Using ontologies for preprocessing and mining spectra data on the Grid. *Future Generation Computer Systems*, 2007, vol. 23, no. 1, pp. 55-60.
- [49] McCUE, C. Operationally Relevant Preprocessing. *Data Mining and Predictive Analysis*, 2007, pp. 93-115.
- [50] LI, L.; XU, Z., X. A Preprocessing Program for Hydrologic Model—A Case Study in the Wei River Basin. *Procedia Environmental Sciences*, 2012, vol. 13, pp. 766-777.
- [51] BARTONĚK, D.; BUREŠ, J. Expert GIS for Locating of Archaeological Sites, *4th Inter-national Conference on Cartography & GIS*, 2012, ISSN 1314-0604.
- [52] VIRDIS, S., G., P. An object-based image analysis approach for aquaculture ponds precise mapping and monitoring: a case study of Tam Giang-Cau Hai Lagoon, Vietnam. *Environmental Monitoring and Assessment*, 2014, vol. 186, no. 1, pp. 117-133.
- [53] SCHWARTZKOPF, W., C.; BOVIK, A.; EVANS, B. L. Maximum-likelihood techniques for joint segmentation-classification of multispectral chromosome images. *Medical Imaging*, IEEE Transactions, 2005, vol. 24, pp. 1593-1610.
- [54] DONG, G. J.; ZHANG, Y. S.; FAN, Y., H. Remote Sensing Image Classification Algorithm Based on Rough Set Theory. *Fuzzy Information and Engineering*, 2007, vol. 40, pp. 846-851.
- [55] LEI, T., C.; WAN, S., CHOU, T., Y. The comparison of PCA and discrete rough set for feature extraction of remote sensing image classification – A case study on rice classification, Taiwan. *Computational Geosciences*, 2008, vol. 12, no. 1, pp. 1-14.
- [56] SADER, S., A.; AHL, D.; LIOU, W., S. Accuracy of Landsat-TM and GIS Rule-Based Methods for Forest Wetland Classification in Maine. *Remote Sensing of Environment*, 1995, vol. 53, no. 3, pp. 133-144.
- [57] WANG, X.; LI, Z.; GAO, S. Parallel Remote Sensing Image Processing: Taking Image Classification as an Example. *Computational Intelligence and Intelligent Systems*, 2012, pp. 159-169.
- [58] ALMEIDA, J., P.; MORLEY, G., J.; DOWMAN, I., J. A graph-based algorithm to define urban topology from unstructured geospatial data. *International Journal of Geographical Information Science*, 2014, vol. 27, no. 8, pp. 1514-1529.
- [59] GAO, Y.; LI, Y. Improving Gaussian Process Classification with Outlier Detection, with Applications in Image Classification. *Computer Vision – ACCV 2010*, 2011, vol. 6495, pp. 153-164.
- [60] SEO, K., K. An Ant Colony Optimization Algorithm Based Image Classification Method for Content-Based Image Retrieval in Cloud Computing Environment. *Computer Applications for Web, Human Computer Interaction, Signal and Image Processing, and Pattern Recognition*, 2012, vol. 342, 2012, pp. 110-117.
- [61] WAHID, N. A Novel Image Classification Algorithm Using Swarm-Based Technique for Image Database. *Ubiquitous Computing and Multimedia Applications*, 2011, vol. 151, pp. 460-470.
- [62] SHIRAKAWA, S.; NAKAYAMA, S.; NAGAO, T. Genetic Image Network for Image Classification. *Applications of Evolutionary Computing*, 2009, vol. 5484, pp. 395-404.

- [63] CHOI, B.; CHUNG, B.; RYOU, J. Adult Image Detection Using Bayesian Decision Rule Weighted by SVM Probability. *Fourth International Conference on Computer Sciences and Convergence Information Technology*, 2009, pp. 659–662.
- [64] CHAPELLE, O.; HAFFNER, P.; VAPNIK, V. N. Support vector machines for histogram-based image classification. *Neural Networks, IEEE Transactions*, 1999, vol. 10, no. 5, pp. 1055-1064.
- [65] DEVASENA, C., L.; HEMALATHA, M. Hybrid Image Classification Technique to Detect Abnormal Parts in MRI Images. *Computational Intelligence and Information Technology*, 2011, vol. 250, pp. 200-208.
- [66] WHITESIDE, T., G.; MAIER, S. W.; BOGGS, G. S. Area-based and location-based validation of classified image objects. *International Journal of Applied Earth Observation and Geoinformation*, 2014, vol. 28, pp. 117-130.
- [67] ZHANG, Y.; ZHANG, H.; LIN, H. Improving the impervious surface estimation with combined use of optical and SAR remote sensing images. *Remote Sensing of Environment*, 2014, pp. 155-167.
- [68] POULIOT, D.; LATIFOVIC, R.; ZABCIC, N.; GUINDON, L.; OLTHOF, I. Development and assessment of a 250 m spatial resolution MODIS annual land cover time series (2000-2011) for the forest region of Canada derived from change-based updating. *Remote Sensing of Environment*, 2014, vol. 140, pp. 731-743.
- [69] BARALDI, A.; BOSCHETTI, L.; HUMBER, M., L. Probability Sampling Protocol for Thematic and Spatial Quality Assessment of Classification Maps Generated From Spaceborne/Airborne Very High Resolution Images 2, *IEEE Transactions on Geoscience and Remote Sensing*, 2014, vol. 52, no. 1, pp. 701-760.
- [70] IAMNES, J., S.; CONGALTON, R., G.; LUNETTA, R., S. Analyst variation associated with land cover image classification of Landsat ETM plus data for the assessment of coarse spatial resolution regional/global land cover products. *GIScience & Remote Sensing*, 2013, vol. 50, no. 6, pp. 604-622.
- [71] PUARTAS, L.; BRENNING, O.; MEZA, A., J. Balancing misclassification errors of land cover classification maps using support vector machines and Landsat imagery in the Maipo river basin (Central Chile, 1975-2010). *Remote Sensing of Environment*, 2013, vol. 137, pp. 112-123.
- [72] OKUJENI, A.; VAN DER LINDEN, S.; TITS, L.; SOMERS, B.; HOSTERT, P. Support vector regression and synthetically mixed training data for quantifying urban land cover. *Remote Sensing of Environment*, 2013, vol. 137, pp. 184-197.
- [73] GRAY, J.; SONG, C. Consistent classification of image time series with automatic adaptive signature generalization. *Remote Sensing of Environment*, 2013, vol. 134, pp. 333-341.
- [74] SILVA, A., Q.; PARADELLA, W., R.; FREITAS, C., C.; OLIVEIRA, C., G. Evaluation of Digital Classification of Polarimetric SAR Data for Iron-Mineralized Laterites Mapping in the Amazon Region. *Remote Sensing*, 2013, vol. 5, no. 6, pp. 3101-3122.
- [75] STEHMAN, S., V. Estimating area from an accuracy assessment error matrix. *Remote Sensing of Environment*, 2013, vol. 132, pp. 202-211.

- [76] MAKARAU, A.; PALUBINSKAS, G.; REINARTZ, P. Alphabet-Based Multisensory Data Fusion and Classification Using Factor Graphs. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2013, vol. 6, no. 2, pp. 969-990.
- [77] JOHNSON, D., M. An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. *Remote Sensing of Environment*, 2014, vol. 141, pp. 116-128.
- [78] MEMARIAN, H.; BALASUNDRAM, S., K., KHOSLA, R. Comparison between pixel- and object-based image classification of a tropical landscape using Systeme Pour l'Observation de la Terre-5 imagery. *Journal of Applied Remote Sensing*, 2013, vol. 7.
- [79] WU, D.; SUN, D.-W. Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review - Part II: Applications. *Innovative Food Science & Emerging Technologies*, 2013, vol. 19, pp. 15-28.
- [80] FINK, M.; MEDELNIK, J.; STROBEL, K.; HIRSCHFELDER, U.; HOFMANN, E. Metric precision via soft-tissue landmarks in three-dimensional structured-light scans of human faces. *Journal of Orofacial Orthopedics-Fortschritte der Kieferorthopädie*, 2014, vol. 75, no. 2, pp. 133-142.
- [81] MOUELHI, A.; SAYADI, M.; FNAIECH, F.; MRAD, K.; BEN R., KHALED. P. A new automatic image analysis method for assessing estrogen receptors' status in breast tissue specimens. *Computers in Biology and Medicine*, 2013, vol. 43, no. 12, pp. 2263-2277.
- [82] ZOHMANN, M.; PENNERSTORFER, J.; NOPP-MAYR, U. Modelling habitat suitability for alpine rock ptarmigan (*Lagopus muta helvetica*) combining object-based classification of IKONOS imagery and Habitat Suitability Index modelling. *Ecological Modelling*, 2013, vol. 254, pp. 22-32.
- [83] ZABALA, A.; CEA, C.; PONS, X. Segmentation and thematic classification of color orthophotos over non-compressed and JPEG 2000 compressed images. *International Journal of Applied Earth Observation and Geoinformation*, 2012, vol. 15, pp. 92-104.
- [84] SEARCOÍD, M, O. Metric Spaces, *Springer Undergraduate Mathematics Series*, 2007, 304 p., ISBN 1-84628-369-8.
- [85] PAWLAK, Z. Rough sets. *International Journal of Parallel Programming*, 1982, vol. 11, no. 5, pp. 341-356.
- [86] BUREŠ, J.; BARTONĚK, D.; OPATŘILOVÁ, I. *Klasifikace údajů o uložení plynárenských zařízení pod určitými typy povrchů terénu*, 2013, Závěrečná zpráva o řešení pilotního projektu č. 9413000195 (HS123570212101), FAST VUT v Brně, 32 str.
- [87] BUREŠ, J.; BARTONĚK, D.; OPATŘILOVÁ, I. *Analýza povrchů nad plynovody RWE na území ČR*, 2014, Závěrečná zpráva o řešení projektu AdMaS ED2.1.00/03.0097 (HS1235702121200), FAST VUT v Brně, 30 str.
- [88] BRÁZDIL, K. *Ortofoto České republiky*, Technická zpráva k ortofotografickému zobrazení území ČR, 2012, Zeměměřický úřad a VGHMÚř, 16 str.
- [89] BARTONĚK, D.; BUREŠ, J.; OPATŘILOVÁ, I. The Solution of Massive Tasks in GIS Exemplified by Determining Terrain Surface Types above Gas Pipelines in the Czech Republic. *Thematic Cartography for the Society*, Springer, 2014, pp. 95-104

- [90] BARTONĚK, D.; BUREŠ, J.; OPATŘILOVÁ, I. Possibilities of Improvement of Image Classification via GIS Tools. *International Conference on Cartography and GIS*, 2014, pp. 96-102
- [91] BARTONĚK, D.; BUREŠ, J.; OPATŘILOVÁ, I.; VITULA, A. Method of error assessment in image classification. *GeoConference on Informatics, Geoinformatics and Remote Sensing*, 2014, vol. 3, pp. 745-752
- [92] BARTONĚK, D.; BUREŠ, J.; OPATŘILOVÁ, I. Technology of processing of enormous amounts of geographical data. *GeoConference on Informatics, Geoinformatics and Remote Sensing*, 2014, vol. 3, pp. 917-924

Elektronické zdroje

- [93] http://en.wikipedia.org/wiki/Out-of-core_algorithm, 31. 7. 2014
- [94] <http://cs.wikipedia.org/wiki/R-strom>, 31. 7. 2014
- [95] <http://noel.feld.cvut.cz/vyu/scs/prezentace2000/transputery/>, 31. 7. 2014

PŘEHLED PUBLIKOVANÝCH PRACÍ STUDENTA

Publikace v impaktovaných časopisech

*BARTONĚK, D.; BUREŠ, J.; OPATŘILOVÁ, I. Optimization of Pre-Processing of Extensive Projects in Geographic Information Systems. *Advanced Science Letters*, 2014, vol. 20, no. 10/11/12, pp. 2026-2029, ISSN 1936-6612, DOI: 10.1166/asl.2014.5664.

BARTONĚK, D.; OPATŘILOVÁ, I. Design of Man–Machine Interface for Mobile Mapping. *Advanced Science Letters*, 2014, vol. 20, no. 2, pp. 501-504, ISSN 1936-6612, DOI: 10.1166/asl.2014.5340.

BARTONĚK, D.; OPATŘILOVÁ, I. The use of GIS technology for planning of GNSS measurement. *Advances in Intelligent Systems*, WIT Press, UK, 2014, ISSN 1743-3517, DOI: 10.2495/Intelsys130251, vydáno i jako kapitola v knize s ISBN 978-1-84564-869-5.

Kapitola v knize

*BARTONĚK, D.; BUREŠ, J.; OPATŘILOVÁ, I. The Solution of Massive Tasks in GIS Exemplified by Determining Terrain Surface Types above Gas Pipelines in the Czech Republic. *Thematic Cartography for the Society*, Springer, 2014, pp. 95-104, ISBN 978-3-319-08180-9, DOI: 10.1007/978-3-319-08180-9_8.

Konferenční články vedené v databázi Conference Proceedings Citation Index

*BARTONĚK, D.; BUREŠ, J.; OPATŘILOVÁ, I. Technology of processing of enormous amounts of geographical data. *Proceedings of 14th GeoConference on Informatics, Geoinformatics and Remote Sensing, Albena 19. 6 – 25. 6. 2014*, SGEM2014, vol. 3, pp. 917-924, ISBN 978-619-7105-12-4, DOI: 10.5593/SGEM2014/B23/S11.116.

*BARTONĚK, D.; BUREŠ, J.; OPATŘILOVÁ, I.; VITULA, A. Method of error assessment in image classification. *Proceedings of 14th GeoConference on Informatics, Geoinformatics and Remote Sensing, Albena 19. 6 – 25. 6. 2014*, SGEM2014, vol. 3, pp. 745-752, ISBN 978-619-7105-12-4, DOI: 10.5593/SGEM2014/B23/S11.095.

BARTONĚK, D.; OPATŘILOVÁ, I. Design of Human-machine Interface. *Proceedings of 7th International Conference on Materials Structure & Micromechanics of Fracture (MSMF-7), Brno 1. – 3. 7. 2013*, VUTIUM, 2013, 5 p., ISBN 978-80-214-4739-4.

Publikace v neimpaktovaných časopisech ve světových databázích

*BARTONĚK, D.; BUREŠ, J.; OPATŘILOVÁ, I. Possibilities of Improvement of Image Classification via GIS Tools. *Proceedings of 5th International Conference on Cartography and GIS, Riviera, Bulgaria 15. – 21. 6. 2014*, pp. 96-102, ISSN 1314-0604.

BARTONĚK, D.; OPATŘILOVÁ, I. Design of Human-machine Interface for GNSS Device. *Journal of Advances in Information Technology*, 2014, vol. 5, no. 1, pp. 15-20, ISSN 1798-2340, DOI: 10.4304/jait.5.1.15-20.

BARTONĚK, D.; OPATŘILOVÁ, I. Design of Human – Machine Interface for GNSS Devices. *Proceedings of 3rd World Conference on Information Technology (WCIT-2012), Barcelona 14. – 16. 11. 2012*, AWERProcedia Information Technology & Computer Science, 2013, vol. 3, pp. 771-776, ISSN 2147-5105.

Články ve sbornících mezinárodních konferencí

BARTONĚK, D.; OPATŘILOVÁ, I. The Use of GIS Technology for Planning of GNSS Measurement. *Proceedings of 26th International Cartographic Conference, Dresden 25. – 30. 8. 2013*, International Cartographic Association, 2013, 13 p., ISBN 978-1-907075-06-3.

BARTONĚK, D.; OPATŘILOVÁ, I.; DERMEKOVÁ, S. Database design of the historical maps. *Proceedings of 7th International Workshop Digital Approaches to Cartographic Heritage, Barcelona 19. – 20. 4. 2012*, International Cartographic Association, 2012, 15 p., ISBN 960-7999-18-5.

*OPATŘILOVÁ, I.; BARTONĚK, D. Classification of types of terrain surface above route of gas line. *Acta Moraviae*, 2013, vol. 5, pp. 58-65, ISSN 1803-7607.

*BARTONĚK, D.; BUREŠ, J.; OPATŘILOVÁ, I. Technology Line for Analysis of Enormous Amounts of Geographical Data. *Sborník Mezinárodní konference Geodézia, kartografia a geografické informačné systémy 2014, Tatranské Matliare 17. – 19. 9. 2014*, Ústav geodézie, kartografie a geografických informačných systémov, Technická univerzita v Košiciach, 2014, 8 s. ISBN 978-80-553-1781- 6.

BARTONĚK, D.; OPATŘILOVÁ, I. Návrh uživatelského rozhraní pro zařízení pro určování polohy pomocí družicových metod. *Sborník Mezinárodní konference Geodézia, kartografia a geografické informačné systémy 2012, Tatranská Lomnica 24. – 25. 10. 2012*, Ústav geodézie, kartografie a geografických informačných systémov, Technická univerzita v Košiciach, 2012, 8 s., ISBN 978-80-553-1173-9.

OPATŘILOVÁ, I. GIS as Support of the State Administration. *Sborník konference XVI Międzynarodowa Konferencja Studenckich Kół Naukowych i XXVIII Sejmik SKN, Wrocław 12. – 13. 5. 2011*, Wydawnictwo Uniwersytetu Przyrodniczego we Wrocławiu, 2011, 11 s., ISBN 978-83-7717-046-5.

OPATŘILOVÁ, I.; BARTONĚK, D.; DERMEKOVÁ, S. The usage of GIS from standpoint of the environment. *Proceedings of European Federation for Information Technology in Agriculture, Praha 11. – 14. 7. 2011*, Czech Centre for Science and Society, 2011, pp. 499-508, ISBN 978-80-904830-3-3.

BARTONĚK, D.; DERMEKOVÁ, S.; OPATŘILOVÁ, I. Historical Maps in GIS. *Proceedings of 16th International Conference on Cultural Heritage and New Technologies, Vienna 14. – 16. 11. 2011*, Museen der Stadt Wien – Stadtarchäologie, 2011, pp. 721-731, ISBN 978-3-200-02740-4.

BARTONĚK, D.; DERMEKOVÁ, S.; OPATŘILOVÁ, I. The Prediction Model in Archaeology. *Proceedings of 16th International Conference on Cultural Heritage and New Technologies, Vienna 14. – 16. 11. 2011*, Museen der Stadt Wien – Stadtarchäologie, 2011, pp. 339-347, ISBN 978-3-200-02740-4.

BARTONĚK, D.; OPATŘILOVÁ, I. Use of GIS Technology for Planning of GNSS Measurement. *Proceedings of XI. International Conference of Soft Computing, 25. 1. 2013*, European Polytechnic Institute, Ltd. Kunovice, 2013, pp. 9-15, ISBN 978-80-7314-291-9.

OPATŘILOVÁ, I.; BARTONĚK, D. Testing of Topcon GRS-1 instrument using RTK method. *Proceedings of X. International Conference of Soft Computing, Kunovice 20. 1. 2012*, European Polytechnic Institute, Ltd. Kunovice, 2012, pp. 81-88, ISBN 978-80-7314-279-7.

BARTONĚK, D.; OPATŘILOVÁ, I.; DERMEKOVÁ, S. Historical Maps in GIS. *Proceedings of X. International Conference of Soft Computing, Kunovice 20. 1. 2012*, European Polytechnic Institute, Ltd. Kunovice, 2012, pp. 17-26, ISBN 978-80-7314-279-7.

OPATŘILOVÁ, I.; BARTONĚK, D. GIS in Municipality Administration. *Proceedings of IX. International Conference of Soft Computing, Hodonín 21. 1. 2011*, European Polytechnic Institute, Ltd. Kunovice, 2011, pp. 67-73, ISBN 978-80-7314-221-6.

Články ve sbornících tuzemských konferencí

OPATŘILOVÁ, I.; BARTONĚK, D. GIS for Small Municipalities. *Sborník Symposia GIS Ostrava, Ostrava 23. – 26. 1. 2011*, Vysoká škola báňská - Technická univerzita Ostrava, 2011, 8 s., ISBN 978-80-248-2366-9.

OPATŘILOVÁ, I. Využití technologie GIS pro plánování GNSS měření. *Sborník 15. odborné konference doktorského studia Juniorstav, Brno 7. 2. 2013*, Vysoké učení technické v Brně, Fakulta stavební, 2013, 9 s., ISBN 978-80-214-4669-4.

OPATŘILOVÁ, I. Testování přístroje Topcon GRS-1 a některých služeb poskytovaných sítěmi permanentních referenčních stanic. *Sborník 14. odborné konference doktorského studia Juniorstav, Brno 26. 1. 2012*, Vysoké učení technické v Brně, Fakulta stavební, 2012, 18 s., ISBN 978-80-214-4393-8.

OPATŘILOVÁ, I. GIS malých obcí. *Sborník 13. odborné konference doktorského studia Juniorstav, Brno 4. 2. 2011*, Vysoké učení technické v Brně, Fakulta stavební, 2011, 8 s., ISBN 978-80-214-4232-0.

VOLAŘÍK, T.; DERMEKOVÁ, S.; OPATŘILOVÁ, I. Výuka matematiky a fyziky pro obor geodézie a kartografie očima studentů. *Sborník referátů workshopu Matematicko-fyzikální vědy ve výuce geodézie a kartografie, Brno 17. – 18. 2. 2011*, Ústav geodézie, Fakulta stavební, Vysoké učení technické v Brně, 2011, s. 97-102, ISBN 978-80-86433-53-0.

OPATŘILOVÁ, I. GIS Městské části Brno-Jundrov. *Sborník 6. studentské konference Student GIS Projekt, zámek Kozel 29. 9. 2010*, ARCDATA Praha, s.r.o., 2010, s. 147-152, ISBN 978-80-904450-3-1.

Skripta

BARTONĚK, D.; OPATŘILOVÁ, I. *HE15 Územní informační systémy: Modul #3 Návody do cvičení*. Studijní opory pro studijní programy s prezenční formou studia, Vysoké učení technické v Brně, Fakulta stavební, 2013, 76 s.

BARTONĚK, D.; OPATŘILOVÁ, I. *GE09 Počítačová grafika I: M2 Technické prostředky, zpracování obrazu a základy 3D grafiky*. Studijní opory pro studijní programy s prezenční formou studia, Vysoké učení technické v Brně, Fakulta stavební, 2013, 143 s.

Výzkumné zprávy projektu

*BUREŠ, J.; BARTONĚK, D.; OPATŘILOVÁ, I. *Klasifikace údajů o uložení plynárenských zařízení pod určitými typy povrchů terénu*, 2013, Závěrečná zpráva o řešení pilotního projektu č. 9413000195 (HS123570212101), FAST VUT v Brně, 32 str.

*BUREŠ, J.; BARTONĚK, D.; OPATŘILOVÁ, I. *Analýza povrchů nad plynovody RWE na území ČR*, 2014, Závěrečná zpráva o řešení projektu AdMaS ED2.1.00/03.0097 (HS1235702121200), FAST VUT v Brně, 30 str.

Publikace přijaté na mezinárodních konferencích, které nebyly dosud publikovány

*BARTONĚK, D.; OPATŘILOVÁ, I. Optimization of Processing of Enormous Amounts of Geographical Data. *The 2nd Global Conference on Computer Science, Software, Networks and Engineering, Kuşadası, Turkey* 6. – 8. 11. 2014, 11 p.

*BARTONĚK, D.; BUREŠ, J.; OPATŘILOVÁ, I. Enhancement of Image Classification via GIS. *The 2nd International Conferences on Computer Graphics, Visualization, Computer Vision, and Game Technology, Bandung, Indonesia* 29. – 30. 10. 2014, 6 p.

*BARTONĚK, D.; BUREŠ, J.; OPATŘILOVÁ, I. Workflow for Analysis of Enormous Amounts of Geographical Data. *The 2nd International Conference on Advances in Intelligent Systems in Bioinformatics, Chem-Informatics, Business Intelligence, Social Media and Cybernetics, Jakarta, Indonesia* 27. – 28. 9. 2014, 8 p.

BARTONĚK, D.; OPATŘILOVÁ, I. The Management of GNSS Measurement via GIS. *The 2nd International Conference on Advances Technology in Telecommunication, Broadcasting, and Satellite, Bali, Indonesia* 11. – 12. 10. 2014, 5 p.

Poznámka: Publikace označené symbolem * souvisí s oblastí disertační práce.