

VĚDECKÉ SPISY VYSOKÉHO UČENÍ TECHNICKÉHO V BRNĚ

*Edice Habilitační a inaugurační spisy, sv. 608*

*ISSN 1213-418X*

**Dalibor Bartoněk**

**MOŽNOSTI ŘEŠENÍ  
VELKÝCH PROJEKTŮ GIS  
S VYUŽITÍM BĚŽNÝCH TECHNICKÝCH  
A PROGRAMOVÝCH PROSTŘEDKŮ**

**VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ**

**Fakulta stavební**

**Ústav geodézie**

**doc. Ing. Dalibor Bartoněk, CSc.**

**MOŽNOSTI ŘEŠENÍ VELKÝCH PROJEKTŮ  
GIS S VYUŽITÍM BĚŽNÝCH TECHNICKÝCH  
A PROGRAMOVÝCH PROSTŘEDKŮ**

**POSSIBILITIES OF SOLVING OF BIG GIS PROJECTS  
WITH ACCESSIBLE HARDWARE AND SOFTWARE SUPPORT**

**TEZE PŘEDNÁŠKY  
K PROFESORSKÉMU JMENOVACÍMU ŘÍZENÍ  
V OBORU  
GEODÉZIE A KARTOGRAFIE**



**BRNO 2018**

## **KLÍČOVÁ SLOVA**

Geografický informační systém (GIS), big data, paralelní výpočet, klasifikace, optimalizace, filtrace, mračno bodů.

## **KEYWORDS**

Geographic information system (GIS), Big Data, parallel computing, classification, optimization, filtration, point cloud.

© Dalibor Bartoněk, 2018

ISBN 978-80-214-5665-5

ISSN 1213-418X

## Obsah

1	ÚVOD .....	6
1.1	Motivace.....	6
1.2	Současný stav problematiky.....	6
2	METODY ŘEŠENÍ.....	11
3	ŘEŠENÉ PROJEKTY .....	12
3.1	Klasifikace povrchu nad plynovody v ČR .....	12
3.1.1	<i>Formulace úlohy</i> .....	12
3.1.2	<i>Teoretická východiska</i> .....	13
3.1.3	<i>Zvolená metoda řešení</i> .....	15
3.1.4	<i>Experimentální výsledky</i> .....	16
3.1.5	<i>Závěr</i> .....	18
3.2	Zpracování mračna bodů z laserového skenování .....	19
3.2.1	<i>Charakteristika úlohy a metoda řešení</i> .....	19
3.2.2	<i>Experimentální výsledky</i> .....	20
3.3	Hledání optimální cesty při měření metodou RTK .....	22
3.3.1	<i>Charakteristika problému</i> .....	22
3.3.2	<i>Formulace úlohy</i> .....	22
3.3.3	<i>Teoretický model a metoda řešení</i> .....	22
	Datový model .....	23
	Transformace datového modelu do tabulek relační databáze .....	23
3.3.4	<i>Využití v praxi</i> .....	27
4	DISKUSE A ZÁVĚR.....	28
5	KONCEPCE DALŠÍ ČINNOSTI .....	29
5.1	VĚDECKÁ ČINNOST .....	29
5.2	PEDAGOGICKÁ ČINNOST .....	30

## **PŘEDSTAVENÍ AUTORA**

Jméno a příjmení: Dalibor Bartoněk

Datum a místo narození: 10. 6. 1952 v Brně

Adresa do zaměstnání:

Ústav geodézie, Fakulta stavební, Vysoké učení technické v Brně,  
Veveří 331/95, 602 00 Brno

Telefon: 541147204

e-mail: bartonek.d@fce.vutbr.cz



### Vzdělání:

- Ing. – 1976, Fakulta elektrotechnická, VUT v Brně, obor elektronické počítače,
- CSc. – 1993, Fakulta elektrotechnická, VUT v Brně, obor výpočetní technika, téma disertační práce: „Návrh technického simulátoru SFDDL/SCL“,
- doc. – 2005, Fakulta stavební, VUT v Brně, obor Geodézie a kartografie, téma habilitační práce: „Vybrané aplikace z oblasti GIS“.

### Přehled o praxi:

- 1976 – 1986 samostatný inženýr výpočetního střediska Ústav výpočetní techniky Ostrava, výpočetní středisko Brno,
- 1986 – 1991 ved. výzkumný a vývojový pracovník Výzkumný ústav měřicí techniky Brno,
- 1991 – 2009 podnikání v oboru „Poskytování software“,
- 1992 – 1994 samostatný výzkumný a vývojový pracovník Výzkumný ústav rozvoje oblastí a měst Ostrava, pracoviště Brno,
- 1994/2006 - dosud - odborný asistent/docent Ústav geodézie Fakulty stavební VUT v Brně.
- 2002 - dosud - externí výuka - Evropský polytechnický institut (EPI), Kunovice.

### Odborné zaměření:

Teoretická informatika, MicroStation, počítačová grafika, databáze, geografické informační systémy (GIS), programování v Delphi, Python.

### Pedagogická činnost - výuka na VUT v Brně:

- Garant předmětů

Interaktivní grafické systémy, Počítačová grafika, Databáze I, Geografické informační systémy I, Geografické informační systémy II., Diplomový seminář, ArcInfo, Geomedia.

- Vedení přednášek a cvičení

Interaktivní grafické systémy, Základy programování, Počítačová grafika, Databáze I, Geografické informační systémy I, Geografické informační systémy II, Diplomový seminář, Geomedia.

- Individuální vedení studentů

17 obhájených bakalářských prací, 40 obhájených diplomových prací, 2 obhájené disertační práce, dosud školitel sedmi doktorandů.

Člen komise pro obhajoby bakalářských a diplomových prací, člen oborové rady doktorského studijního programu v oboru geodézie a kartografie.

### Pedagogická činnost - výuka na Evropském polytechnickém institutu v Kunovicích:

- Přednášky i cvičení pro obor elektronické počítače:

Teoretické základy informatiky, Programovací techniky, Databázové systémy, Informační systémy, Datamining.

Člen komise pro obhajoby bakalářských prací v oboru elektronické počítače a ekonomická informatika. Člen rady pro informační systém EPI.

### Seznam řešených projektů:

1. Zavedení automatizace inženýrských prací v bývalém koncernu TESLA Měřicí a laboratorní přístroje (MLP) v době působení ve Výzkumném ústavu měřicí techniky v letech 1986 – 1991:
  - Koncepce systému automatizace inženýrských prací v koncernu TESLA MLP.
  - Automatizace návrhu plošných spojů (Metra Blansko, Tesla Valašské Meziříčí).
  - Automatizace konstrukčních prací (Tesla Brno, Metra Blansko).
2. Řešení úkolu č. 1/31/FI/1992 Regionální informační systém. Zadavatel Ministerstvo hospodářství ČR. Dílčí úkoly: metainformační subsystém a databáze.
3. Evidence a vyhodnocení protikorozní ochrany potrubí pro bývalé ČPP (české plynárenské podniky) – spoluřešitel. Software - dvě instalace byly realizovány ve Francii.
4. Evidence a vyhodnocení protikorozní ochrany úseků a sledů pro Transgas Praha (software - systém GASSERV) – spoluřešitel. (Zdrojové moduly vytvořeny v Borland Delphi).
5. Evidence a vyhodnocení protikorozní ochrany úseků a sledů pro MERO ČR a. s. Kralupy nad Vltavou (včetně produktovou Ingolstadt – Kralupy). Software - systém MGASSERV.
6. Automatické sledování a vyhodnocení vad potrubí pro ČPP a ČEPRO a. s - software. (Pearsonova metoda měření vad a detekce vad inteligentním ježkem) – spoluřešitel.
7. Úkol FRVŠ 1680 „Vybudování moderního vzdělávacího pracoviště Geodézie a kartografie na Ústavu geodézie FAST VUT v Brně“ - spoluřešitel.
8. Spoluřešitel projektu OP VK CZ.1.07/2.200/15.0144. Inovace studijních programů Geodézie a kartografie.
9. Spoluřešitel projektu ADVASOFT - Pokročilé softwarové nástroje ve stavebním inženýrství.
10. Odpovědný řešitel projektu: Klasifikace údajů o uložení plynárenských zařízení pod určitými typy povrchů terénu. Řešeno pro RWE GasNet, s.r.o., v rámci AdMaS Projekt registrován jako Ověřená technologie RIV: RIV/00216305:26110/13.

### Členství v profesních organizacích a jiné aktivity

1. Člen vědeckého výboru International Conference on Soft Computing Applied in Computer and Economic Environments – ICSC, Evropský polytechnický institut Kunovice (2004).
2. Člen vědeckého výboru mezinárodní konference 2th, 3th and 4th Czech–Polish Symposium Brno–Wrocław “ACTUAL PROBLEMS OF GEODESY, CARTOGRAPHY AND PHOTOGRAMMETRY 2009, 2011, 2014, 2017.
3. Člen vědeckého výboru mezinárodní konference VII. and VIIIth Scientific and Professional International Conference Geodesy, cartography & geographic information systems 2012, 2014.
4. Stálý oponent projektů grantové agentury APVV Bratislava, Slovensko.
5. Člen International Society for Development and Sustainability (ISDS), Membership ID: M170431, Japan.
6. Členství v redakční radě časopisu Journal of Geodesy and Geomatics Engineering
7. Členství v redakční radě časopisu Journal of Civil Engineering and Architecture Research
8. Členství v redakční radě časopisu Journal of Computer, Science Research Association
9. Členství v redakční radě časopisu Journal of Architectural Research and Development,

### Publikační a další činnost:

15 x publikace v časopisech SCI, 11 x kapitola v knize, 19 x příspěvky na mezinárodních vědeckých konferencích, 21 x publikace v odborném časopisu (zahraničním), 44 x posudků zahraničních článků nebo zahraničních projektů, 30 x řešené projekty, 19 x skripta, 52 x ostatní publikace (publikace v tuzemských časopisech, abstrakty atd.), 4 x vyžádaná přednáška na mezinárodních konferencích.

# 1 ÚVOD

## 1.1 MOTIVACE

V současné době jsme svědky mohutného rozvoje digitálních informačních technologií. Tento trend je možné pozorovat na všech úrovních společenského života – individuální, institucionální, na úrovni státní správy a samosprávy atd. Popisovaný proces má svoji statickou stránku – téměř ve všech oblastech probíhá rozsáhlá digitalizace analogových podkladů a dynamickou část – komunikace mezi účastníky využívá digitální kanály a komunikační pojítka. Situace zasáhla i do oblasti tzv. prostorových dat, což jsou data, která kromě atributové složky obsahují i lokalizační složku, která jednoznačně určuje polohu v závazném souřadnicovém systému (v ČR je to S-JTSK – Systém Jednotné Trigonometrické Síť Katastrální). V důsledku těchto změn je prudký nárůst digitálních dat, který je podporován i tím, že výrazně roste technologická vyspělost zařízení pro výrobu těchto dat. Tato zařízení obsahují řadu vestavěných funkcí, která zjednodušují obsluhu a zefektivňují jejich nasazení. Dnes už může téměř každý uživatel po nezbytném zaškolení využívat zařízení pro sběr dat přímo v terénu. Současně s nárůstem vyspělosti technologie zařízení pro pořízení prostorových dat však poněkud stagnují metody a potažmo i programové vybavení pro zpracování dat velkých objemů. Tento jev lze označit jako paradox problematiky dat velkých objemů, které se v zahraniční literatuře označují termínem „Big Data“, v Terminologickém slovníku zeměměřictví a katastru nemovitostí je navrhován překlad „velká data“ [61].

Dalším jevem je skutečnost, že z celkového objemu se využívá jen asi 2% dat [24]. Hlavním důvodem tohoto stavu je, že zpracování velkých dat vyžaduje v mnoha případech výpočetní prostředky vysokých výkonů se speciálním programovým vybavením. Takovou informační infrastrukturou ale disponuje jen velmi omezený počet uživatelů.

Poměr mezi množstvím zpracovaných dat a jejich celkovým objemem by se zlepšil ve prospěch využitých dat, kdyby velká data mohli zpracovávat i běžní uživatelé informačních technologií. Prakticky to znamená navrhnout takové metody a pracovní postupy, aby velká data bylo možné zpracovávat s podporou běžných technických a programových prostředků.

V geografických informačních systémech (GIS) se s výše popsány problémy setkáváme při řešení projektů s rozsáhlou územní vazbou nebo značnou sekundární složitostí, které v konečném důsledku vedou ke zpracování velkých dat. Sekundární složitostí se v tomto případě myslí případ, kdy objem primárních dat daného projektu je relativně malý, ale výsledek řešení závisí na kombinaci těchto vstupních dat. Příkladem může být např. hledání Hamiltonovské cesty v grafu, kde je zapotřebí prohledávat rozsáhlý stavový prostor. V experimentální části jsou navrženy 3 možné typické způsoby zpracování rozsáhlých projektů v oblasti GIS, které autor řešil.

V další kapitole jsou uvedeny nejpoužívanější metody pro zpracování velkých dat v obecné rovině. Návrh metod pro zpracování velkých prostorových dat obsahuje kapitola 2.

## 1.2 SOUČASNÝ STAV PROBLEMATIKY

Pro zpracování velkých dat se používají převážně tyto metody:

### 1. Optimální řízení zpracovatelského procesu (inteligentní management).

Jde např. předzpracování dat formou účelové filtrace s cílem redukce objemu dat nebo využití inteligentního rozhraní pro výběr relevantních dat z rozsáhlých databází.

V práci [2] je navržen systém pro optimalizaci procesů řízení podnikového informačního systému. Podstatou je přímý on-line uživatelský přístup k relevantním datům v rámci Business Intelligence daného podniku.

Autoři v [33] navrhli systém řízení silničního provozu, který má 3 role: správce, operátor a klient. Správce působí jako vedoucí oddělení městské dopravy. Rada operátorů je zodpovědná za získávání obrázků a fyzických umístění silničních aktivit pomocí kamery a chytrých telefonů

s vestavěnou GPS. Klient má důležitou roli při monitorování provozu a údržby silničního provozu. Systém je založen na Geo JavaScript Object Notation a příslušná webová služba na GeoJSON.

V příspěvku [34] je navržen pracovní postup pro 3D vizualizaci modelů pro simulaci počasí a klimatu. Principem je integrace heterogenních prvotních dat do 3D vizualizací, které lze zobrazit na stolním počítači nebo v prostředí interaktivní virtuální reality.

V práci [49] byla vyvinuta infrastruktura založená na cloudových technologiích pro přístup k datům a zpracování dat o pozorování Země s cílem řešit problémy s účinnou manipulací s daty, jimž čelí malé a střední podniky. Byl implementován Input Data Portal (IDP). IDP se skládá z konformního katalogu Open Geospatial Consortium (OGC), konsolidačního modulu pro konverzi formátů a rámce objednávek, který odpovídá požadavkům OGC.

Článek [52] popisuje výzkum zaměřený na usnadnění rozsáhlého vědeckého výpočtu v počítačových sítích. Řešení zahrnuje programovací metodu, režijní nárok na middleware založený na programovém rozhraní a data předpokládají migraci. Výsledky ukazují, že programové rozhraní usnadňuje spouštění složitých vědeckých aplikací a mechanismus migrace očekávání dat může zlepšit efektivitu této platformy.

## **2. Aplikace speciálních technických nebo programových prostředků (akcelerátorů).**

V práci [1] byl vyvinut speciální softwarový produkt pro správu geologických dat. Je používán k vytváření spolehlivých trojrozměrných geologických modelů, které vyžadují velké množství paměťových a výpočetních prostředků. Produkt pracuje na bázi speciálního algoritmu, který plně využívá akcelerace grafických karet v stolním počítači.

V publikaci [38] je navržen programový akcelerátor pro vyhodnocení snímků teplotních map velkých objemů v biologii. Vykreslování interaktivních teplotních map z velkých vstupních datových množin bylo výpočetně nedostupné jak na stolních počítačích, tak na webových prohlížečích. Program je koncipován jako vestavěný vysoce výkonný webový zásuvný modul, který umožňuje rychlé vykreslování interaktivních teplotních map datových souborů až  $10^5$  až  $10^7$  řádků během několika sekund. Aplikace je dostupná na odkazu: <http://shinyheatmap.com>.

Další akcelerátor pro řadící algoritmy v grafice byl představen v práci [57]. Podstatou je využití grafických koprocesorů v běžných počítačích pro předzpracování dat před exportem do cloudového systému.

Speciální klastry počítačů s grafickými akcelerátory systému Nvidia Tegra K1 (TK1) pro zpracování velkých objemů prostorových dat navrhuje autoři v [59]. Navrhovaný akcelerátor má být levný s příznivým poměrem výpočetního výkonu vzhledem k rychlosti komunikace mezi uzly.

Podobný akcelerátor jako v [59] je představen v publikaci [60]. Jeho hlavní funkcí je filtrovat velký objem dat a tím přispívat k jejich efektivnímu zpracování.

## **3. Síťové řešení (cloudové řešení, aplikace Hadoop, Apache Spark, MapReduce).**

Speciální konfigurace klastru Hadoop pro náročné zpracování velkých obrazových dat je popsána v práci [3]. Jde o vyhodnocování snímků s objemem řádově terabyte z elektronického mikroskopu.

Další cloudové řešení pro zpracování grafických dat velkých objemů je uvedeno v [20]. Navržené řešení využívá nový algoritmus pro bezztrátovou kompresi obrazu, který je implementován v novém grafickém formátu. Kompresi snižuje uložené objemy dat a snižuje rychlost přenosu dat, ale snížená velikost dat musí být vyvážena vzhledem k době potřebné pro dekompresi obrazu. V práci je popsán nový algoritmus v jazyku JavaScript, který optimalizuje tuto rovnováhu. Kombinace tohoto nového formátu ukládání obrázků založeného na cloudovém formátu a algoritmus komprese pomůže vyřešit některé problémy velkých obrazových dat na internetu.



Problematikou zpracování mračna bodů s využitím technologie Apache Spark se zabývá publikace [21]. Pro distribuci dat k jednotlivým uzlům sítě se využívá speciálně navržená mapovací funkce, která výrazně zvyšuje výkonnost celého systému.

Článek [30] popisuje zpracování velkého objemu prostorových dat pomocí Hadoop na němž je spuštěna aplikace MapReduce. Při vhodné konfiguraci je toto kloudové řešení o řád rychlejší ve srovnání s výpočty na běžných počítačích. Koncepce byla vyvinuta pro vyhledávání geografických dat na webu.

#### **4. Virtualizace technických nebo programových prostředků.**

Studie v [39] obsahuje návrh jak efektivně ukládat a zpracovávat velké objemy dat generované mobilními zařízeními a jejich službami. Principem je virtualizace úložišť stolních počítačů, které integrují svoje prostředky a poskytují tyto prostředky uživatelům. V systému je řešena dostupnost a spolehlivost navrhovaného řešení a dále snadné přidávání nebo odebírání jednotlivých stolních počítačů do/ze síťové infrastruktury. Při chybě v systému se aktivují alternativní servery.

V publikaci [56] je popsán hybridní plánovací algoritmus pro různé typy aplikací: dávkové úlohy a pracovní postupy. V inicializační fázi vytváří navrhovaný algoritmus hierarchické seskupení dostupných zdrojů do skupin. Úkoly se řeší ve dvou fázích: v první jsou úkoly přiřazeny skupinám zdrojů a ve druhé fázi je pro každou skupinu prostředků použit klasický plánovací algoritmus. Navrhovaný algoritmus je vhodný pro heterogenní distribuované zpracování, zejména pro moderní systémy HPC (High Performance Computing), ve kterých běží výpočty především s daty z multimediálních aplikací.

#### **5. Využití ontologie, sémantický Web.**

Pro zpracování velkého objemu dat ze sociálních sítí je možné využít ontologický přístup. Inherentní problém sémantického webu je poskytování společné slovní zásoby, která je přijatelná pro mnoho uživatelů, přičemž všichni mohou mít velké rozdíly v kulturním prostředí, vzdělání, roli, úkolech a zájmech. V článku [26] je navrženo řešení na bázi sémantické plochy (Semantic Desktop). Každý zdroj je identifikován jednotným identifikátorem zdrojů (URI) a všechna data jsou přístupná a dotazovatelná jako graf RDF (Resource Description Framework). Tímto způsobem mohou být uloženy zdroje z webu a obsah autora může být sdílen s ostatními. Ontologie umožňují uživateli vyjádřit osobní mentální modely a vytvářet sémantické pojítka propojující informace a systémy. Aplikace pracují a komunikují prostřednictvím ontologií a protokolů sémantického webu. Semantic Desktop je rozšířený doplněk k uživatelské paměti, který pomáhá uživatelům vytvářet, sdílet, sdílet a zpřístupňovat veškeré digitální informace, jako jsou dokumenty, multimedia a zprávy.

Práce [55] prezentuje system pro spolupráci dat z mobilních zařízení v lékařství na bázi sémantického webu. Jde o vlastní adaptivní přenosový přístup pro synchronizované zpracování na webu mezi různými terminály, jako jsou mobilní zařízení a stolní počítače. Systém poskytuje funkce, grafické prvky a další údaje pro potřeby lékařů.

Na bázi ontologie je založen systém pro zpracování dat z telekomunikačních sítí [58]. Byl vytvořen nový dotazovací systém speciálně navržený pro telekomunikační sítě, který integruje jak on-line předprocesní, tak i off-line analýzu pro datové přenosy v sociálních sítích. Aplikace navíc dokáže zrychlit zpracování dotazu vytvořením a analýzou stromu abstraktní syntaxe.

#### **6. Využití paralelismu a následné souběžné zpracování.**

Využití paralelismu pro zrychlení výpočtu je jednou z nejstarších a nejpoužívanějších metod. Největším problémem je v tomto případě přidělování úloh prostředkům (zpracovatelským jednotkám). V práci [27] je navržen přístup, kdy rozdělování dat na paralelní segmenty se provádí na bázi vzorů těchto vstupních dat.

V [31] je popsána paralelní programová platforma, která běží v koudovém řešení, na superpočítačích i na stolních počítačích. Využívá technologii C ++ a je zaměřena na analýzu dat v biologických systémech.

Práce [41] poukazuje na to, že bylo vyvinuto pouze několik strategií s cílem maximalizovat využití paralelních výpočetních prostředků na podporu interaktivní vizualizace. Autoři proto navrhli metodu predikce intenzity vizualizace závislou na pohledu, která může dynamicky předpovídat intenzitu vizualizace založenou na distribučních schématech časoprostorových dat. Předpovídané výsledky se používají k naplánování přidělení vizualizačních úkolů. Tato strategie je integrovaná s paralelním vizualizačním systémem nasazeným v kloudu grafických procesorových jednotek počítačů s jednotnou architekturou zařízení.

V publikaci [51] je představen nástroj pro analýzu velkých obrazových dat, softwarový rámec pro zpracování a analýzu heterogenních databází, jako jsou například mračna bodů, rastrová data, nebo data v klastrech nebo v individuálních počítačích (stolních či laboratorních). Využívá se paralelismus založený na rozdělení dat přičemž se vyloučí potřeba meziprocesové komunikace. Koncepce nemá centrální plánovací jednotku pro přidělování úloh procesorům, uživatelé přistupují ke sdíleným informačním zdrojům asynchronně.

### **7. Využití strojového učení a umělé inteligence (např. klasifikace vstupních dat).**

Strojového učení je další možností, jak zjednodušit zpracování dat velkých objemů. Využívá se především k účelovému výběru vstupních dat, čímž se jejich objem pro další zpracování zmenší. V článku [28] je navržen soubor generalizovaných lineárních modelů jako rychlá alternativní metoda pro řešení úloh z oblasti astronomie. Software je vyvinut v jazyku Python a je veřejně dostupný prostřednictvím interaktivní webové aplikace.

Autoři v práci [37] navrhli dynamickou distribuci úloh pro řešení na inteligentních mobilních zařízeních založených na lidském chování pro použití webových prostředí příští generace. Principem je předběžné zpracování dat s cílem minimalizovat dobu odezvy služby podle definice metadat a požadavků uživatelů na služby rozšířené reality. Prakticky jde o návrh dynamického plánu rozdělení úloh na základě výpočetní kapacity stolních počítačů, které tvoří kloudovou infrastrukturu.

V práci [48] je navržen akcelerátor pro úlohy strojového učení založený na heterogenní structure ve webovém prostředí. Aplikace využívá Hadoop a MapReduce. Tyto specializované heterogenní architektury však vyžadují rozsáhlé posouzení aspektů návrhu, aby bylo dosaženo optimální architektury z hlediska výkonu a nákladů. Článek analyzuje způsob vytížení prostředků a navrhuje optimální konfiguraci výpočetních uzlů vzhledem k charakteru řešené úlohy.

### **8. Nasazení superpočítačů (High Performance Computing).**

Zřejmě nejjednodušší způsob, jak zajistit rychlý výpočet, je využití speciálních superpočítačů, tj. počítačů s vysokým výkonem. Studie [29] poskytuje přehled této problematiky od počátků nasazení jednoho z prvních komerčních superpočítačů - Intel IPSC hypercube v polovině 80. let až do dnešní doby a s výhledem do budoucnosti.

### **9. Využití optimalizačních metod pro snížení náročnosti zpracování.**

Optimalizační metody lze využít buď pro účelový výběr vstupních dat a tím zmenšení jejich objemu nebo ve fázi zpracování, kdy volbou vhodného algoritmu můžeme zrychlit výpočet.

V pracích [23] a [36] je navržen speciální algoritmus pro zpracování dat z mobilních zařízení s rozšířenou realitou. V souvislosti s rostoucí kvalitou služeb (Quality of Service - QoS) vyžaduje zpracování dat z těchto zařízení vysokou výpočetní výkonnost. Principem algoritmu je omezení zpracování redundantních dat (zejména snímků s vysokým rozlišením), čímž se výrazně zrychlí výpočet.

Velkým problémem je proces vizualizace (zejména ve 3D) na webu, protože síla webového prohlížeče dosud nedosáhla úrovně stolních (desktopových) aplikací. V publikaci [42] jsou uvedeny metody a techniky používané k vizualizaci pomocí virtuálního glóbu ve webovém prohlížeči. K vizualizaci je použita optimalizační strategie ve WebGL kombinovaná s přístupem render-to-texture pro postupné vykreslování a vizualizaci dat.

Optimalizační algoritmus pro stínování 3D objektů je navržen v publikaci [54]. Podstatou je optimalizované schéma komprese, která zkoumá redundanci času a prostorovou redundanci časově měnících objemových dat. Komprimovaná data jsou pak přenášena do grafické paměti a přímo vykreslena grafickým koprocesorem, což významně snižuje přenos dat mezi hlavní a grafickou pamětí. Další optimalizační strategie je představena v práci [53]. Jde o aplikaci v lékařství založenou na technologii SparkR.

### **10. Využití n-dimenzionálního přístupu k dané úloze (transformace problému).**

V tomto případě jde o zjednodušení výpočtu problému, který je zadán v určité dimenzi. Metoda spočívá v tom, že vhodnou transformací převedeme problém do nižší dimenze, v této dimenzi úlohu vyřešíme a výsledek transformujeme do původní (vyšší) dimenze. Při tomto přístupu se však můžeme setkat s jevem, kterému se říká „prokletí dimenzionality“. Jev spočívá v tom, že při převodu do jiné dimenze se mohou změnit některé vlastnosti transformovaných prvků. Touto problematikou se zabývá publikace [54]. Velké objemy dat se podrobí hierarchickému funkčnímu rozkladu metodou kódování vlastní autokorelační struktury v Gaussových procesech. Tím se sníží dimenze datové struktury a následné zpracování pak lze provádět pomocí lineárních algoritmů.

### **11. Řetězené zpracování (pipeline).**

Tohoto přístupu se používá při řešení velkého množství úloh stejné nebo podobné struktury. Podstatou této metody je rozdělení úlohy na dílčí části, které mohou být různého charakteru a následné zpracování každé této části specializovanou funkční jednotkou. Funkční jednotky pak pracují paralelně, čímž dochází k časovému překrytu dílčích částí ve zpracovatelském řetězci a tím se celková doba řešení zrychlí. Funkční jednotky mohou být tvořeny technickým zařízením nebo programovými moduly.

Příkladem je publikace [32], která popisuje projekt pro zpracování velkých dat v astronomii. Autoři navrhli uživatelské rozhraní Trident, které nabízí komplexní webový a kloudový software pro mikroprocesory. Tento přístup umožňuje bezproblémové nasazení vysoce adaptabilních systémů Scalable Compute Archive (SCA) - včetně rozsáhlých vizualizačních a analytických funkcí - s minimálními požadavky na dodatečné kódování. Trident plynule zvyšuje nebo snižuje objem dat a výpočetní výkon a umožňuje, aby sady funkcí v rámci webového uživatelského rozhraní byly rychle přizpůsobeny požadavkům jednotlivých úloh. Řetězené zpracování probíhá v pěti fázích a pro každou z nich byly vytvořeny programové moduly s těmito funkcemi: 1. pracovní postup a integrace, 2. služba pro sledování pracovních toků, 3. interaktivní služba pro vizualizaci, 4. datová služba, 5. služba pro statistické výpočty a vykazování potřeb různých projektů. Prvky webového rozhraní jsou vytvořeny pomocí knihoven JavaScript NodeJS, AngularJS a HighCharts, zatímco backendové služby jsou zapsány v systémech NodeJS, PHP a Python.

### **12. Kombinace výše uvedených přístupů**

Kombinace metod, které zefektivňují zpracování dat velkých objemů tvoří zřejmě nejpočetnější početnou skupinu. Ze všech možných příkladů uveďme je výčtem ty nejtypičtější:

V publikaci [25] autoři vytvořili webovou aplikaci, která integruje - virtualizaci technických prostředků a klasifikaci v kloudovém řešení s cílem snížit počet vstu/výstupních (I/O) operací. Koncepce řešení v práci [40] je kombinací kloudového řešení s nasazením superpočítačů.

## ***Bezpečnost při zpracování dat velkých objemů***

Při všech výše uvedených metodách je zapotřebí zajistit bezpečnost zpracování z hlediska možného úniku dat a informací. Analýza tohoto zabezpečení z organizačního hlediska však vyžaduje silnou a nákladnou informační infrastrukturu. Proto je pro mnoho uživatelů z finančních důvodů prakticky nedostupná. Článek [47] představuje intuitivní a cenově dostupný přístup k analýze bezpečnosti dat pomocí umělé inteligence. Pro analýzu zabezpečení slouží kombinace programování a spouštění dávkových souborů. Simulace se provádí na skutečném datovém souboru s více než 10 miliony pozorování, které jsou shromažďovány z protokolů brány firewall systému Windows a umožňují uživateli stolního počítače nahlédnout do svých dat a získávat užitečné informace o jejich bezpečnosti.

## **2 METODY ŘEŠENÍ**

V předchozí kapitole byly představeny možnosti zpracování velkých dat v obecných případech, které autorovi poskytly inspiraci pro návrh metod, které je možné využít při řešení rozsáhlých projektů v GIS. Geografický informační systém je v současné době již velmi známý pojem, který je popsán v mnoha publikacích např. [43], proto se zde omezíme jen na nejdůležitější fakta, která mají úzký vztah k danému tématu.

Základem každého GIS je digitální databáze, jejímž základním prvkem je geografický objekt (GO), který je obrazem jevů nebo prvků reálného světa. Zobrazení transformující podmnožinu prvků reálného světa do digitální databáze musí splňovat homeomorfismus, což je jednojednoznačné zobrazení. Všechny GO musí obsahovat mimo jiné i územní vazbu, což znamená, že

- poloha každého GO je jednoznačně určena v platném referenčním systému
- existuje metrika, která splňuje tzv. trojúhelníkovou nerovnost.

GIS je tedy informační systém, jehož doménou je územně vázaná databáze tzv. geo-databáze. Jen pro zajímavost, spoluautor knihy [43] M. Goodchild povýšil GIS koncem 90. let z původně chápané technologie na vědu. Podle americké společnosti Daratech Inc., specializující se na průzkum trhu pracuje 80% všech existujících informačních systémů s územně vázanými daty, což což zdůrazňuje potřebu, řešit problematiku velkých dat v této oblasti.

Při řešení rozsáhlých projektů v GIS se můžeme setkat především s těmito základními charakteristikami úloh:

1. zpracování analýz na rozsáhlém území (např. území celé ČR), které předpokládá využití objemných datových sad a výsledky v každé lokalitě jsou na sobě relativně nezávislé,
2. úlohy s primárně značným objemem vstupních dat s vysokou redundancí nebo s daty, která jsou na sobě relativně závislá – příkladem může být např. pozemní nebo letecké skenování,
3. projekty s relativně malým objemem primárních dat, ale složitým algoritmem řešení, který je většinou založen na prohledávání rozsáhlého stavového prostoru, jenž je tvořen kombinací vstupních dat – viz již dříve zmíněná úloha hledání Hamiltonovské cesty v grafu.

Podle výše uvedených charakteristik můžeme pro řešení rozsáhlých projektů v GIS navrhnout tyto metody:

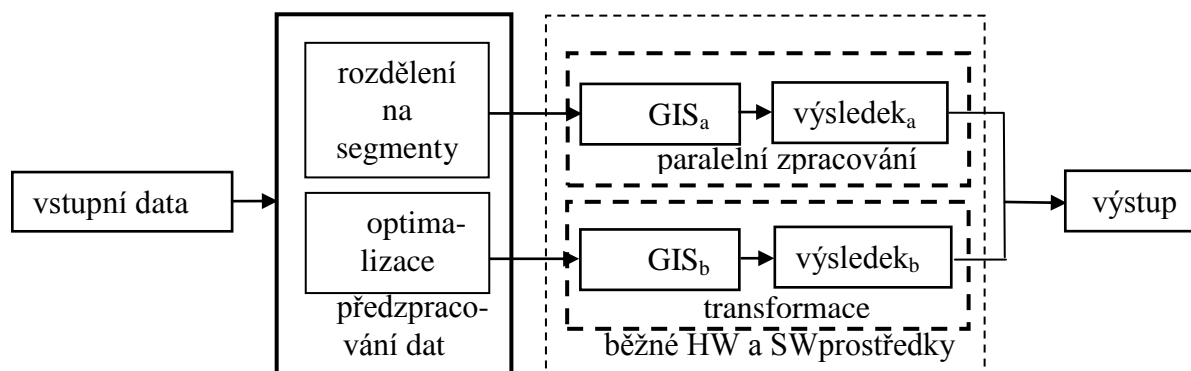
V případě ad 1), kdy doménou projektu je velké území, bude vhodné toto území

1. Rozdělení vstupních dat na segmenty (územně vázané systémy) a pro další zpracování využít paralelismus. Metoda byla využita v projektu „Klasifikace povrchu nad plynovody v ČR“ – viz kap. 3.1.

Není-li možné toto rozdělení a musí se pracovat s celým souborem vstupních dat (např. u síťové analýzy), pak

2. účelová redukce nebo zjednodušení (optimalizace) vstupních dat
  - a) filtrací dat (tato metoda byla použita v případě laserového skenování – viz kap. 3.2)
  - b) snížením dimenze problému tj. řešení v nižší dimenzi a poté transformace zpět do vyšší dimenze. Uvedená metoda byla aplikována při hledání optimální cety při měření metodou RTK (Real Time Kinematic) – viz kap. 3.3.

Situace je znázorněna na obr. 1. Obě metody vyžadují testovací úlohu (pilotní projekt), aby mohl být proveden kvalifikovaný odhad relevantních parametrů (doba řešení, kapacita prostředků a řešitelského týmu).



Obr. 1. Metody řešení velkých objemů dat na běžných HW a SW prostředcích

Teoretická východiska jednotlivých navrhovaných metod budou uvedena v úvodu každého demonstračního projektu v další kapitole.

### 3 ŘEŠENÉ PROJEKTY

#### 3.1 KLASIFIKACE POVRCHU NAD PLYNOVODY V ČR

##### 3.1.1 Formulace úlohy

Cílem projektu bylo:

a) zhotovení analýzy povrchů nad plynovody RWE na území ČR, tj. zpracování klasifikace údajů o uložení plynárenských zařízení pod určitými typy povrchů terénu za účelem stanovení reprodukčních hodnot plynárenských zařízení (plynovodů) a ocenění nákladů, které by bylo nutné vynaložit na vybudování nových sítí,

b) analýza povrchů měla být provedena ve třech variantách pro průběh plynovodu VTL, hl. řadu a přípojek:

b1) 2 třídy druhu povrchu: zpevněný a nezpevněný

b2) 5 tříd druhu povrchu: asfalt, hlavní komunikace, místní komunikace, nevázaný, neznámý

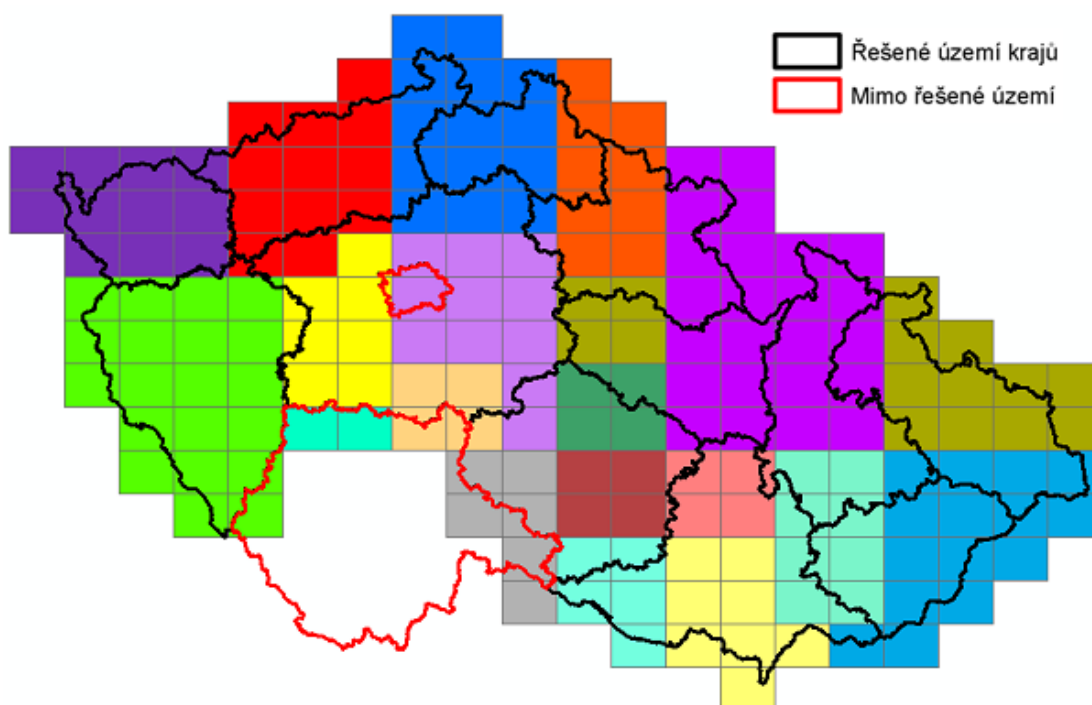
b3) 10 tříd druhu povrchu: les, travní porost, holá půda, asfalt, střecha – taška, střecha – rovná, stín, hlavní komunikace, místní komunikace, cesta,

c) výstupem je strukturovaná tabulka ve formátu XLSx s klasifikovanými údaji typů povrchů nad průběhem trasy plynovodu - soubor „VystupRWE.xlsx a grafický výstup strukturovaných klasifikovaných údajů – formát SHP pro VTL, hlavní řad místních sítí a přípojky, soubory „VT\*.\*\*\*“ – vysokotlak, „HR\*.\*\*\*“ – hlavní řad a „P\*.\*\*\*“ – přípojky,

d) rozsah analýzy území - celkem 12 krajů: Karlovarský kraj, Plzeňský kraj, Středočeský kraj, Ústecký kraj, Liberecký kraj, Královéhradecký kraj, Pardubický kraj, Kraj Vysočina,

Moravskoslezský kraj, Olomoucký kraj, Zlínský kraj a Jihomoravský kraj, tj. celkem pro území 188 obcí s rozšířenou působností (ORP) – viz obr.2.

V takto rozsáhlých projektech územní analýzy zpracovávaných v prostředí GIS se pracuje s velkým množstvím prostorových geodat. Využitelnými jsou veřejně dostupná data základního státního mapového díla georeferencovaného ortofota ČR a základní báze geografických dat ČR (ZABAGED). Tématická data jsou zpravidla privátní a jsou poskytována příslušným správcem technické infrastruktury z jeho disponibilních zdrojů v surovém stavu a pro potřeby konkrétního projektu. Pro účel datových analýz je třeba vždy data 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. Efektivní datové struktury se vytvářejí s využitím optimalizačních metod.



Obr. 2. Rozsah modelovaného území s vyznačením různé barevnosti ortofoto

### 3.1.2 Teoretická východiska

Nechť neprázdná množina  $U \neq \emptyset$  je univerzum diskurzu a  $X$  je 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  $U/R = \{X_1, X_2, \dots, X_n\}$  tak, že pro každé  $i, j$  platí:

- 1)  $X_i \subseteq U, X_i \neq \emptyset$ ; (všechny podmnožiny jsou neprázdné),
- 2)  $X_i \cap X_j = \emptyset$ ; (průnik všech podmnožin je prázdný),
- 3)  $\bigcup_{i=1,2,\dots,n} X_i = U$ . (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):  $X_i / S = \{X_{i1}, X_{i2}, \dots, X_{im}\}$  (1)

pro všechna  $i=1, 2, \dots, n$ .

Pak systém relací  $\{R, S\}$  spolu s množinami  $U, X$  představuje hierarchický rozklad množiny  $U$ . Relace  $S$  v našem případě 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 (GO), tj. prvky, které jsou hlavním předmětem zpracování v daném projektu GIS (v našem případě jde o vedení plynovodu). Relace  $S$  tak rozdělí areály do podtříd  $X_{ij}$  pro  $j = 1, 2, \dots, m$ , podle vztahu (1). 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 GO v území ve druhé hierarchické úrovni.

Nechť  $X$  je množina GO v reálném světě a množina  $Y$  obraz množiny  $X$  v digitální geodatabázi. Pak zobrazení  $\varphi: X \rightarrow Y$  musí mít tyto vlastnosti:

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

Pak zobrazení  $\varphi$  je homeomorfismus, který je základní vlastností topologických vztahů mezi GO. V našem případě jsou GO (množina  $X$ ) vstupní datové sady a množina  $Y$  pak obrazy těchto datových sad v datovém skladu projektu.

Nyní definujme dobu řešení celého projektu  $T_{tot}$ :

$$T_{tot} = T_{rez} + T_{pa} + T_{pm} \quad (2)$$

kde  $T_{rez}$  je doba režie,  $T_{pa}$  je čas automatizovaného a  $T_{pm}$  čas ručního zpracování.

Předpokládáme-li rozdělení celého projektu do  $n$  dílčích částí (rovnice 1), pak rovnice (2) má tvar:

$$T_{tot} = T_{rez} + \sum_{i=1}^n (t_{pai} + t_{pmi}) = \min \quad (3)$$

Pro efektivní zpracování je nutné, aby doba režie nebyla větší než součet dob automatizovaného a ručního zpracování, tj. poměr  $T_{pa} + T_{pm}$  a  $T_{rez}$  by měl být maximální (alespoň větší než 1):

$$\eta = \frac{\sum_{i=1}^n (t_{pai} + t_{pmi})}{T_{rez}} = \max \quad (4)$$

Rovnice (3) a (4) reprezentují účelové funkce, které mají být optimalizovány. Na základě tohoto koeficientu 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.

Dále definujeme koeficient územní podrobnosti:

$$\delta = \frac{P_{tot}}{P_{det}}, \quad (5)$$

kde  $P_{tot}$  je celková plocha dílčího území (rozklad podle relace ekvivalence  $R$ ) a  $P_{det}$  je plocha nejmenšího detailu, který je obsažen ve vstupních datech. Tento koeficient ovlivňuje výběr vhodného technického a programového vybavení co do výpočetní výkonnosti.

Nechť  $S$  je datová množina výsledků projektu,  $R$  je datová množina stejného typu jako  $S$  a atributy  $R$  mají vyšší kvalitu než atributy  $S$ . Pak můžeme zpřesnit výsledky přepsáním hodnot atributů množiny  $S$  atributy množiny  $R$ :

$$\text{Je-li } S \cap R \neq \emptyset \text{ pak } a_r \rightarrow a_s \quad (6)$$

kde  $\cap$  je prostorové překrytí vrstev,  $a_s$  je atribut  $s \in S$  a  $a_r$  je atribut  $r \in R$ .

### 3.1.3 Zvolená metoda řešení

Základním přístupem bylo rozdělení celého projektu do dílčích částí a využití paralelismu pro zrychlení zpracování. Pro úspěšné řešení projektu bylo nutné kvalifikovaně určit parametry v rovnicích (1) – (6) a to:

- optimální zpracovatelskou jednotku (dílčí projekt) (rovnice 1) z hlediska efektivity a spolehlivosti,
- koeficient  $\delta$ , který určuje datovou náročnost projektu (rovnice 2),
- hodnoty režijního času, doby automatizovaného a manuálního zpracování a účinnost (rovnice 3, 4, 5),
- volba referenční datové sady  $R$  pro vylepšení kvality výsledků (rovnice 6).
- volba vhodného softwarového vybavení
- volba vhodných dostupných technických prostředků (počet, konfigurace apod.).

Uvedené parametry byly určeny řešením pilotního projektu stejného typu jako hlavní projekt GIS na omezeném území. Parametry pro hlavní projekt pak byly určeny extrapolací na základě poměru rozsahů modelovaných území.

Navržená metodika řešení velkého projektu GIS je na obr. 3. Celý projekt je rozdělen na subprojekty podle územního principu. Bylo využito územně identifikačního registru (číselník územních jednotek ČR). Pro korektní automatizované zpracování je důležité, aby struktura územní identifikace splňovala homeomorfismus uvedený v předchozí kapitole. Data pro dílčí projekt jsou rozdělena do šablon a uložena do adresářové struktury. Názvy adresářů je vhodné označit jednoznačným identifikačním kódem v souladu s kódem příslušné územní jednotky. Pak následuje redukce objemu dat na základě charakteru hlavní úlohy (optimalizace podle tématu). Jde např. o vytvoření bufferu nebo zmenšení rozsahu modelované oblasti podle polohy významných prvků. Po úvodních procedurách následuje hlavní zpracování tj. aplikace územně analytických funkcí. Závěrečnou fází je vylepšení výsledků filtrací podle referenční datové sady (rovnice 6). Po skončení každé z uvedených dílčích fází se provádí manuální kontrola výsledků a případná oprava. Na závěr se všechny dílčí výsledky sestaví do společné formy výstupu.

Pro zajištění maximální míry automatizace celého postupu je nutné využít buď existující programové vybavení nebo si vytvořit vlastní aplikaci. V našem případě byla vytvořena sada skriptů v jazyku Python. Jde o tyto procedury:

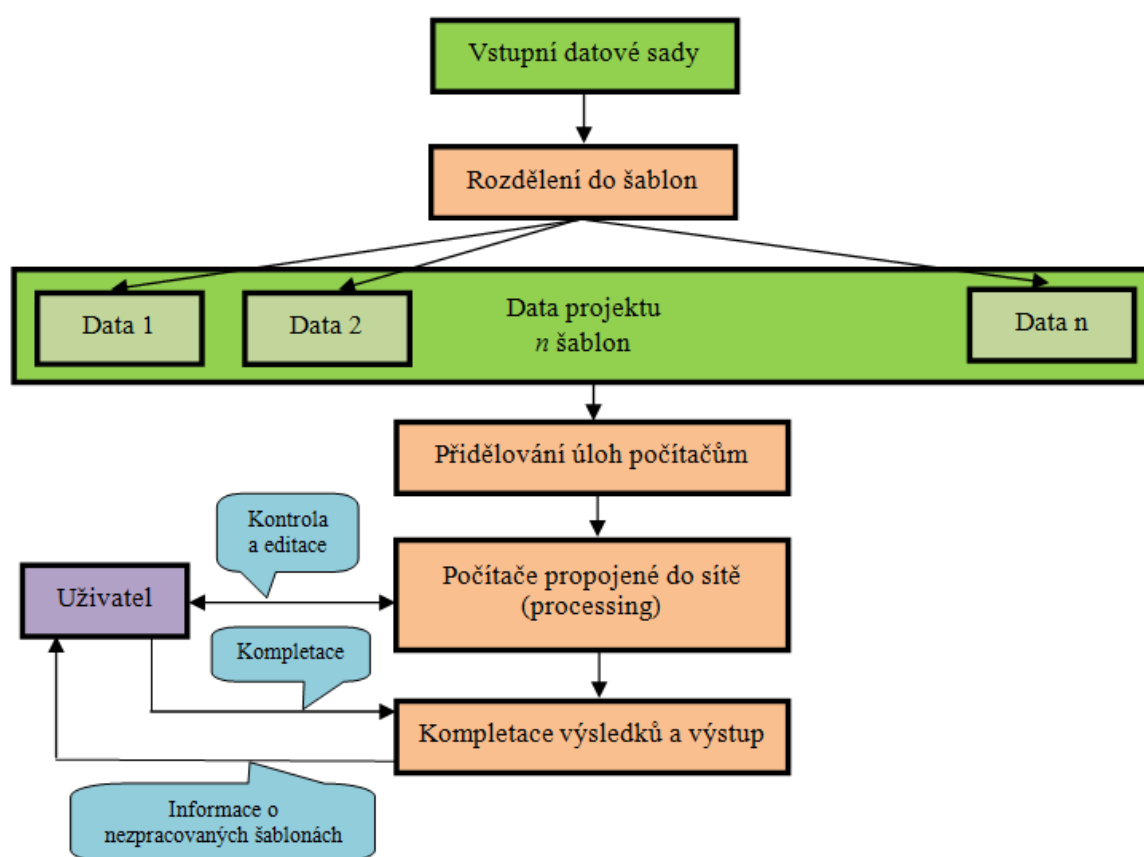
- Procedura pro rozdělení dat do dílčích šablon. Řídící datovou strukturou je územně identifikační registr ČR s vlastností homeomorfismu,
- Procedura pro redukci objemu dat (optimalizace) z hlediska hlavního tématu (tvorba bufferu či jiné omezení),



- Procedura pro předzpracování dat v dílčích šablonách (úprava dat, tvorba pomocných datových sad nebo struktur),
- Procedura pro hlavní (paralelní) zpracování v dílčích sub-projektech,
- Procedura pro vylepšení výsledků (redukce chybovosti) s využitím vhodné referenční vrstvy podle vztahu (6),
- Procedura pro sestavení výsledků do společné datové sady nebo databázové tabulky.

Nezbytným předpokladem k úspěchu je optimální skladba řešitelského týmu. V praxi se osvědčilo toto minimální složení 3 kategorií řešitelů:

1. 1 x manager, zajišťující řízení projektu jako celku,
2. 1x programátor GIS, který vytváří výše uvedené procedury pro automatizované zpracování,
3. 1 x pracovník pro ruční a vizuální kontrolu všech průběžných výsledků.



Obr. 3. Koncepte zpracování projektu

### 3.1.4 Experimentální výsledky

Navržená metoda byla testována v projektu klasifikace povrchů, pod kterými jsou uložena plynárenská zařízení v rámci České republiky. Autoři řešili tento projekt pro společnost GasNet, sro, která je součástí skupiny RWE v ČR. Vstupními daty byly datové sady ortofoto s rozlišením 25 cm/pixel, vrstvy silnic ZABAGED ČR (základní báze geografických dat) a vektorové soubory tras podzemních inženýrských sítí. Vzhledem k teritoriálnímu pokrytí ČR s rozlohou 64 350 km<sup>2</sup> šlo o zpracování velkých dat s celkovým objemem dat 500 GB.

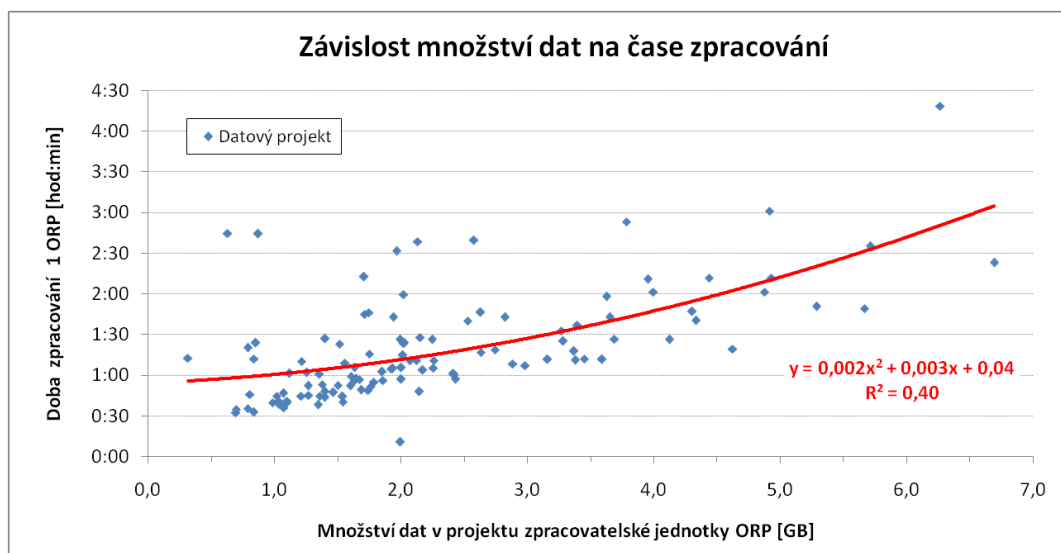
Pro kvalifikovaný odhad hodnot parametrů v rovnicích (1) – (6) byl řešen pilotní projekt na modelovaném území města Brna o ploše cca 250 km<sup>2</sup>. V tomto testovacím projektu byla stanovena tato strategie:

- jako optimální územní jednotka (segment) byla zvolena obec s rozšířenou působností (ORP) - viz rovnice (1), jejíž průměrná plocha činí 368 km<sup>2</sup>, průměrné množství dat pro 1 ORP je 2,5 GB. Toto rozhodnutí ovlivnily i vztahy (3) a (4).
- Optimální poměr  $\eta$  (rovnice 4) pro celkový počet dat  $x = 500$  GB se vypočítá z hodnot:  $n = 200$ ,  $x_i = 2,5$  GB.
- Optimalizace dat podle řešeného tématu byla provedena oříznutím vrstvy ortofoto s obalovou zónou o velikosti 1 m podél plynovodu. Hodnota byla konzultována s firmou RWE. Tím se objem dat snížil z původních 500 GB až na 230 GB.
- Po každém dílčím automatizovaném zpracování ORP byla provedena manuální kontrola dílčích výsledků.
- Doby režie, automatizovaného a manuálního zpracování byly určeny podle grafu na obr. 4 a po dosažení hodnot do vztahu (4) získáme vztah (7):

$$\eta = \frac{\sum_{i=1}^n 0,02e^{0,4x_i}}{\sum_{i=1}^n 0,002x_i^2 + 0,003x_i + 0,04} \quad (7)$$

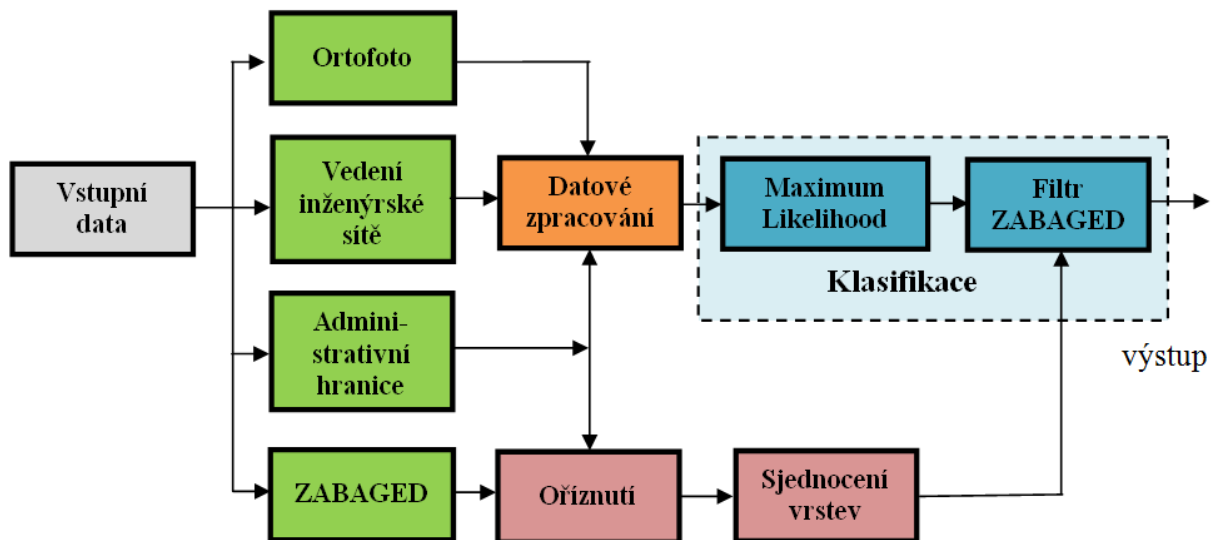
$$T_{rez} = 0,002x^2 + 0,003x + 0,04, \quad T_{pa} = 0,02e^{0,4x}$$

- Na základě vztahu (7) se odhadlo trvání celého projektu po dobu 4 měsíců.
- Poměr manuálního a automatického zpracování je 55%: 45%.
- Implementace projektu byla provedena na počítačích typu PC Integra 7025 (parametry: Intel Core i5, 3,8 GHz, 16 GB RAM, NVIDIA GTX650, 2 GB, HD SDD a VelociRaptor), které byly propojeny s počítačem síť 100 Mb/s. Hlavním kritériem pro toto rozhodnutí byla hodnota koeficientu  $\delta = 10^9$ . V testovacím projektu bylo zjištěno, že vyšší hodnota koeficientu  $\delta$  by znamenala použití výkonnějších počítačů.



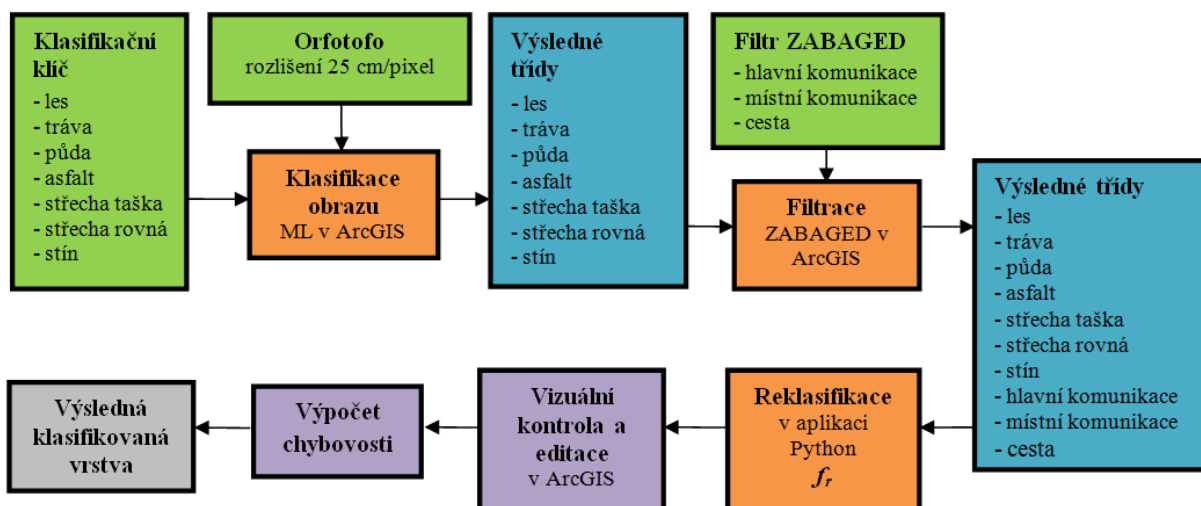
Obr. 4. Závislost doby režie  $T_{rez}$  na objemu zpracovávaných dat

- Analýza dat proběhla s podporou vlastních procedur vytvořených v jazyce Python s využitím knihoven ESRI v prostředí ArcGIS 10.0.
- Pro klasifikaci rastrového obrazu (ortofoto) byla zvolena metoda Maximum likelihood v ArcGIS 10.0. Tato klasifikace však vykazovala pouze průměrnou úspěšnost 70%, což bylo pro zadavatele nedostatečné. Blokové schéma klasifikace povrchů je na obr. 5.



Obr. 5. Blokové schéma klasifikace povrchů

• Proto byla vstupní data doplněna datovou sadou 9 vrstev komunikací, která je součástí ZABAGED ČR. Pomocí těchto referenčních vrstev se zpřesnil výsledek klasifikace (rovnice 6). Koncepce metody pro zpřesnění klasifikace ukazuje obr. 6.



Obr. 6. Postupné zpřesňování výsledků klasifikace obrazu ortofota

• Výsledky prokázaly vysokou účinnost technologie a nízkou chybovost klasifikace v rozmezí 2% - 3% v celé modelované oblasti (ČR).

• Dílčí výsledky byly postupně zapisovány do listu Excel, který byl pro řešitele dostupný na webu. Tím se zabránilo duplicitě zpracování jednotlivých ORP a zlepšilo se řízení celého projektu.

### 3.1.5 Závěr

• Byla navržena metodika pro řešení rozsáhlých projektů GIS a ověřena možnost jejich zpracování na běžných hardwarových a softwarových prostředcích. Metodika spočívá v daty řízených procedurách realizovaných ve skriptovacím jazyku Python s podporou knihoven ESRI.

Hlavní rysy navrhované metodiky jsou:

- optimální rozdělení projektu na dílčí podprojekty podle územního principu,
- paralelní zpracování dílčích projektů na vhodném hardware a software,
- optimalizace objemu dat podle tematiky,

- zajištění kvality výsledků manuální kontrolou výstupů jednotlivých dílčích projektů,
- zvýšení kvality výsledků (snížení chybovosti) pomocí referenční datové sady - v našem případě ZABAGED ČR.

Předpoklady pro úspěšné řešení projektu jsou:

- Řízení kvality projektu, koordinace fází dílčích projektů, interoperabilita a zaměnitelnost výzkumného týmu v rámci řešení jednotlivých zpracovatelských úkolů,
- Řešení modelového pilotního projektu stejného typu jako hlavní projekt v omezeném prostoru za účelem získání potřebných hodnot parametrů pro stanovení strategického postupu.

Technologie má obecný charakter a může být použita pro klasifikaci povrchu nad inženýrskými sítěmi jako vodovody, plynovody, produktovody, energetické sítě apod. Projekt byl popsán v publikacích [4] - [12].

## 3.2 ZPRACOVÁNÍ MRAČNA BODŮ Z LASEROVÉHO SKENOVÁNÍ

### 3.2.1 Charakteristika úlohy a metoda řešení

V tomto projektu byla řešena detekce hran z dat pořízených pozemním laserovým skenováním. Podstatou úlohy je vybrat množinu bodů, které polohově určují obrysy staveb v místě styku s terénem (půdorys). Svou podstatou patří tato úloha do zpracování dat velkých objemů, přičemž mračno bodů, získané skenováním obsahuje vysoký stupeň redundance.

Navržený postup vyhodnocení je založen na parametrické filtraci vstupního mračna nad objekty aktualizované digitální formy katastrální mapy s cílem výrazného zmenšení objemu dat s následnou detekcí hran půdorysů budov a jiných objektů.

Cílem je navrhnout vhodnou metodu pro automatizovanou aktualizaci vybraných prvků technických a katastrálních map. Z vybraných prvků jsou to budovy a stavby. Podstatou aktualizace je identifikovat rozdíly půdorysů budov nebo staveb mezi skutečným stavem a kresbou v mapě. Vstupními daty jsou: mračno bodů, které reprezentuje aktuální stav v terénu a mapa v digitální podobě pro danou lokalitu. V případě nesouladu mohou nastat 2 varianty:

1. skutečný stav vykazuje více objektů než kresba v mapě,
2. aktuální stav má méně objektů než digitální mapa.

V případě ad 1) můžeme proces detekce změn popsat matematickým modelem:

$$R = (C \rightarrow D) - (S \rightarrow B) \quad (8)$$

$$\text{kde } D = f_r(C) \quad (9)$$

$$B = f_b(S) \quad (10)$$

kde  $R$  je výsledná množina objektů, množina  $C$  reprezentuje mračno bodů z laserového skenování a  $S$  množinu objektů v digitální mapě. Množinu  $D$  získáme z mračna bodů předzpracováním funkcí  $f_r$  a analogicky pak množinu  $B$  vygenerujeme aplikací bufferu (funkce  $f_b$ ) na množinu  $S$ . Funkce  $f_r$  realizuje hranovou detekci (obrysy budov a staveb), parametrem funkce  $f_b$  (bufferu) je kritérium přesnosti dle ČSN 01 3410 (např. pro 3. třídu přesnosti mapy je standardní odchylka určení lomového bodu hranice 14 cm).

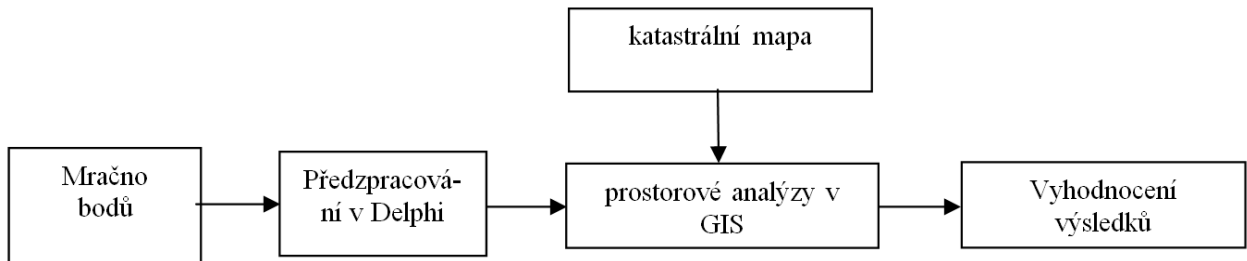
V případě ad 2) lze proces detekce změn popsat podobným matematickým modelem se stejnou symbolikou:

$$R = [(C \rightarrow D) \rightarrow B] - S \quad (11)$$

$$\text{kde } D = f_r(C) \quad (12)$$

$$B = f_b(D) \quad (13)$$

Je-li  $R = \{\emptyset\}$ , pak nejsou detekovány žádné změny, je-li  $R \neq \{\emptyset\}$ , pak existují rozdíly mezi současným stavem v katastrální mapě a reálným stavem v terénu. Blokové schéma celého procesu je na obr. 7.



Obr. 7. Koncepte detekce změn vybraných objektů v technických a katastrálních mapách

Na obr. 7 funkci  $f_b$  zajišťuje blok „prostorové analýzy GIS“, funkce  $f_r$  je realizována v bloku „předzpracování v Delphi“. Zatímco funkce bufferu  $f_b$  je všeobecně známá z prostorových analýz, nalezení vhodné funkce  $f_r$  vyžadovalo provést řadu experimentů. Hranová detekce na bázi centrální diskretní konvoluce se ukázala jako nevhodná pro svoji časovou náročnost. Pro příklad - aplikace konvolučního filtru na budovu o rozměru 15 x 10 m v programu ArcGIS trvala na desktop počítači běžných parametrů asi 8 minut strojového času, což je pro další postup zcela nevyhovující. Jako výhodnější se ukázalo využití agregace mračna bodů podle polohy v rovině ( $x, y$ ) s atributem četností bodů v ose  $z$  (výška objektu) a s filtrací:

$$\text{Select } x, y, \text{count}(z) \text{ from list group by } x, y \text{ having count}(z) > p \quad (14)$$

kde *list* je seznam souřadnic mračna bodů z laserového skenování a  $p$  je parametr, který udává výšku objektu (podle rozlišení mračna), ve které vedeme pomyslnou rovinu řezu, která nám oddělí budovu nebo stavby od ostatních objektů v terénu. Prakticky volíme  $p \geq 2$ .

Tuto filtraci nebylo bohužel možné provést v aplikaci GIS, protože v současné době prostředí GIS včetně skriptů v jazyku Python nepodporuje klauzuli pro agregaci „group by“. Pro tyto účely byl proto vytvořen speciální program v Delphi, který slouží k přípravě vstupních dat – viz obr. 7. Program má tyto funkce:

1. rozdělí vstupní mračno bodů do segmentů menšího objemu dat tak, aby tato data bylo možné zpracovávat na běžných technických a programových prostředcích (desktop computers),
2. v každém segmentu provede filtraci mračna bodů podle vztahu (14),
3. upraví výsledný datový soubor do vhodného formátu pro další zpracování v GIS (v případě QGIS je to formát CSV).

### 3.2.2 Experimentální výsledky

Metoda navržená v předchozí kapitole byla ověřena ve vybrané lokalitě v městě Brně. Na obr. 8a je znázorněno mračno bodů kolem budovy, na obr. 8b je surové mračno s vyznačením filtrace bodů. Body obarvené na obr. 8b červeně vyhovují agregační funkci v rovnici (14).

Výsledná detekce změn v katastrální mapě vzhledem k aktuálnímu stavu je uvedena na obr. 9. Jde o aplikaci vztahu (8) tj. skutečný stav v tomto případě vykazuje více objektů než kresba v mapě. Obr. 8 a obr. 9 obsahují pouze výřezy vstupních dat a slouží k demonstraci navržené metody.

V článku byla navržena metoda, která ve vstupním mračnu bodů identifikuje hrany půdorysů staveb pomocí agregační funkce. Po této filtraci se mračno porovná se stavem zakresleným v současné katastrální mapě v digitalizované podobě. Výsledkem je detekce změn zákresu objektů a lze ji využít pro aktualizaci změn terénních úprav a výskopisu v katastrálních a technických

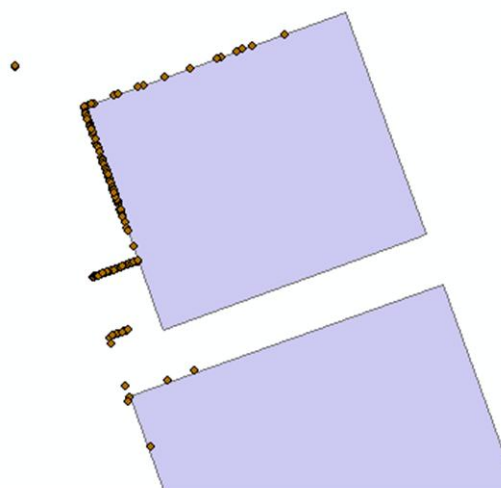
mapách v digitální formě. Proces probíhá v maximální možné míře automatizovaně ve 2 aplikacích:

1. ve speciálním programu v Borland Delphi a
2. v některém programu GIS (Geomedia, ArcGIS, QGIS).

Autor zvolil aplikaci QGIS, která je volně stažitelná z internetu. Navržená metoda je relativně rychlá, nenáročná na technické a programové vybavení a lze ji použít k detekci změn polohopisu v libovolných digitálních mapách ve spojení s laserovým skenováním. Výsledků lze využít pro aktualizaci příslušného digitálního mapového díla.

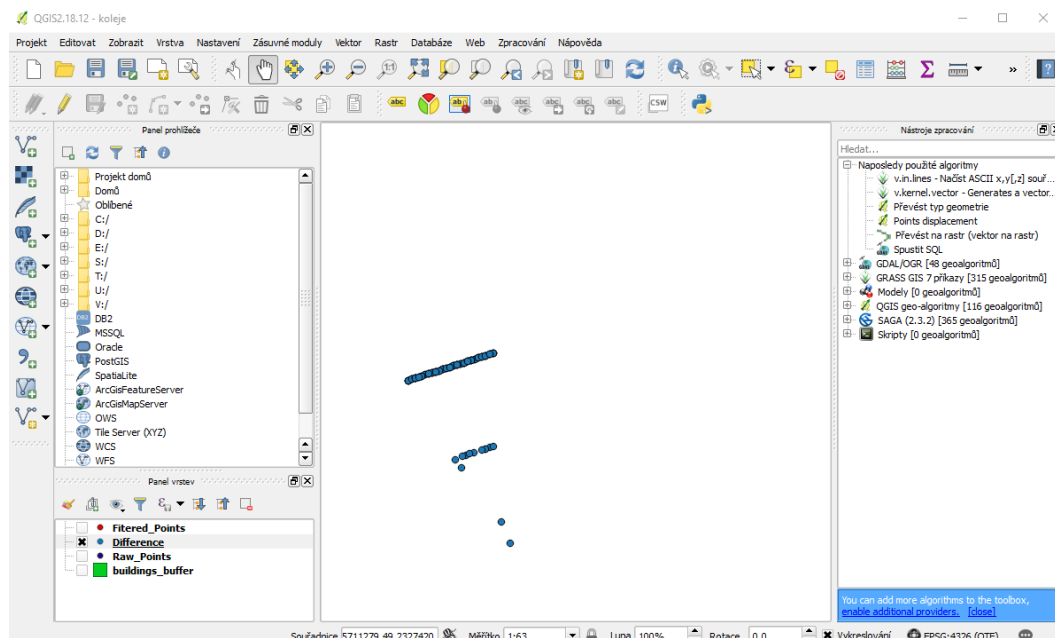


a) budova s mračnem bodů



b) mračno po filtraci s katastrální mapou

Obr. 8 - Příklad detekce změn v katastrální mapě



Obr. 9 - Výsledek (Difference) v mapovém okně QGIS (detail)

Projekt byl publikován ve článku [22].

### 3.3 HLEDÁNÍ OPTIMÁLNÍ CESTY PŘI MĚŘENÍ METODOU RTK

#### 3.3.1 Charakteristika problému

V tomto projektu nejde primárně o zpracování velkých dat, ale o řešení úloh s rozsáhlým stavovým prostorem. Stručně řečeno to znamená, že pro výsledné řešení je zapotřebí prohledávat rozsáhlou množinu možných kombinací vstupních dat. Je všeobecně známo, že pro tyto třídy úloh dosud nebyl objeven přijatelný algoritmus – viz např. [45]. Jde o tzv. NP-úplné problémy, které jsou charakterizovány tím, že s lineárním nárůstem vstupních dat roste exponenciálně doba řešení nebo nároky na paměť počítače. Doba řešení podle tzv. robustního algoritmu by trvala neúměrně dlouhou dobu i na speciálních superpočítačích.

Z tohoto hlediska je možné tyto problémy zařadit do skupiny úloh s velkým objemem dat. V územně vázaných systémech se tyto typy úloh vyskytují často. V tomto případě není možné použít metody rozdělení vstupních dat podle územního principu nebo využít filtraci dat s cílem redukovat jejich objem. Pro korektní řešení musíme pracovat se všemi vstupními daty. Příkladem tohoto problému je např. hledání minimální cesty mezi všemi body v dané lokalitě, na nichž se má měřit metodou GNSS – RTK.

#### 3.3.2 Formulace úlohy

##### *Zadání:*

Je dána množina územně rozmístěných objektů na nichž se má měřit. K dispozici jsou tyto atributy objektu:

ID, poloha, doba observace včetně tolerance, množina intervalů (datum, čas), v nichž je možné měřit. Poloha je dána vzájemnými vzdálenostmi mezi jednotlivými objekty v délkových i časových jednotkách. Do databáze se zadává vzdálenost ujetá vozidlem i vzdálenost, která se absolvuje pěšky. Na každém bodě je zadán i nutný časový odstup mezi bezprostředně následujícím měřením. Měření je možné opakovat, pro měření může být k dispozici i několik měřicích sad přístrojů.

##### *Úloha:*

Sestavte itinerář měření na všech objektech z hlediska:

1. předmětného (vlastní měření)
2. ekonomického (efektivita procesu)

ad 1) Sestavte optimální časový rozvrh měření na všech bodech tak, aby zpoždění vlivem přemísťování mezi body nebo čekáním na vhodný interval měření na daných bodech bylo minimální.

ad 2) Stanovte celkové náklady na měření tj. kalkulace cestovního vozidlem i pěšky, hodinovou sazbu měřičů, pronájem přístrojů. Z vypočtených nákladů na měření pak odvoďte efektivnost kampaně tj. optimální počet přístrojů, počet měřičů a vhodný čas pro měření.

#### 3.3.3 Teoretický model a metoda řešení

Nechť  $A$  je množina objektů. Každý objekt  $a \in A$  má tyto atributy:

- $a.p$  je poloha objektu v daném souřadnicovém systému,
- $a.t$  je časový interval, ve kterém může na daném objektu probíhat nějaká událost (i opakovaně),
- $a.s$  je časový odstup od poslední události. Během tohoto časového odstupu nemůže na daném objektu probíhat žádná událost.

Mějme množinu  $B$  časových intervalů, během nichž mohou na daných objektech  $a \in A$  probíhat události a množinu hodnot metrik  $D$ , jejíž prvky vyjadřují vzdálenost mezi jednotlivými objekty  $a \in A$  v délkových nebo časových jednotkách.

Nechť jsou dány funkce  $f: A \rightarrow C$ ,  $C \subseteq B$  a funkce  $g: D \rightarrow A \times A$ .

Funkce  $f$  přiřazuje objektům  $a \in A$  množinu vhodných intervalů  $a.t \in C$ , v během nichž mohou na objektech probíhat události, funkce  $g$  přiřazuje každé dvojici objektů  $a_i, a_j \in A$  metriku  $d_{ij} \in D$ .

Úkolem je optimalizovat účelovou funkci:

$$\sum_{i=1}^n \sum_{j=1}^{m_i} a_i.t_j + \sum_{i=1}^n \sum_{j=1}^{m_i-1} a_i.s_j + \sum_{i,j=1, i \neq j}^{n-1} d_{ij} = \min. \text{ pro } \forall a \in A. \quad (15)$$

kde  $a_i.t_j$  je doba trvání  $j$ . události na  $i$ . prvku ( $j$  vyjadřuje počet opakování události)  $d_{ij}$  je vzdálenost mezi objekty  $a_i$  a  $a_j$  v časových jednotkách,  $n$  je počet objektů,  $m_i$  je počet událostí, které nají proběhnout na objektu  $a_i$ .

### Datový model

Entity a atributy:

Na základě zadání byly identifikovány tyto entity (tučně) a jejich atributy (v závorce):

**Bod** (číslo bodu, Y, X, Z, čas\_blokování, poznámka)

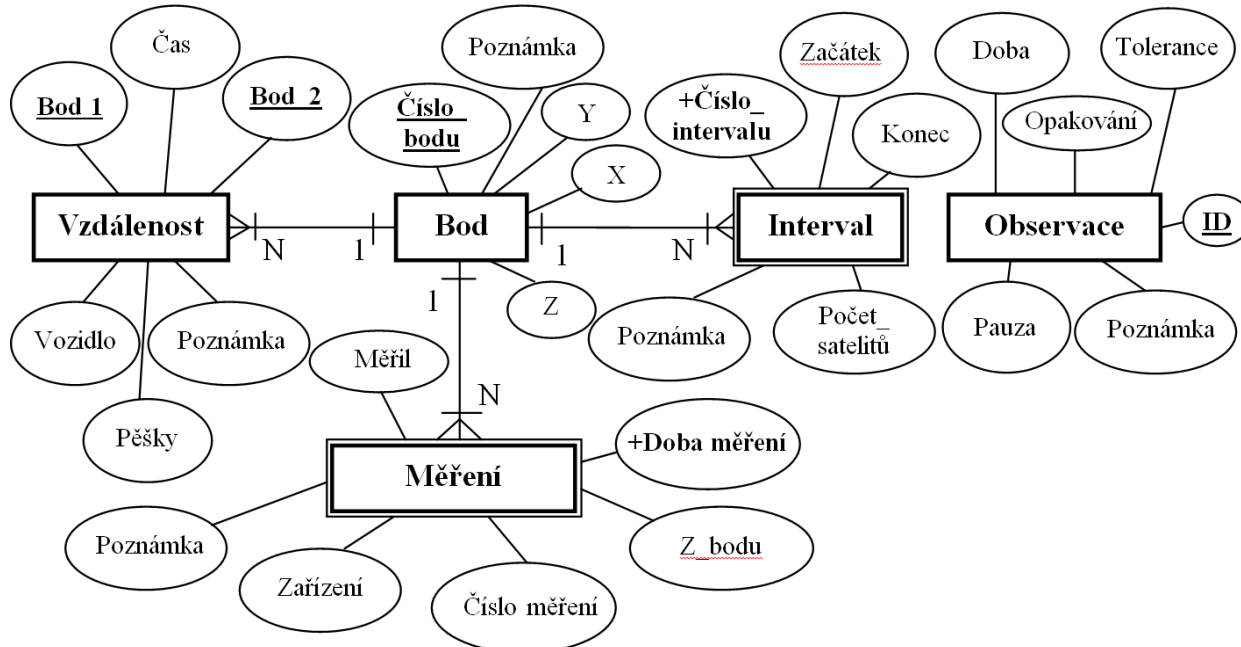
**Vzdálenost** (z bodu, do bodu, vzdálenost\_vozidlo, vzdálenost\_pěšky, čas, poznámka)

**Interval** (+číslo\_intervalu, začátek, konec, počet\_satelitů, poznámka)

**Měření** (+číslo\_měření, číslo\_předchozího\_bodu, čas\_měření, označení\_přístroje, jméno\_měřiče, poznámka)

**Observace** (ID, doba\_observace, odstup, tolerance, počet\_opakování, poznámka)

Z výše uvedeného prvotního soupisu byl vytvořen datový model, který je na obr. 10.

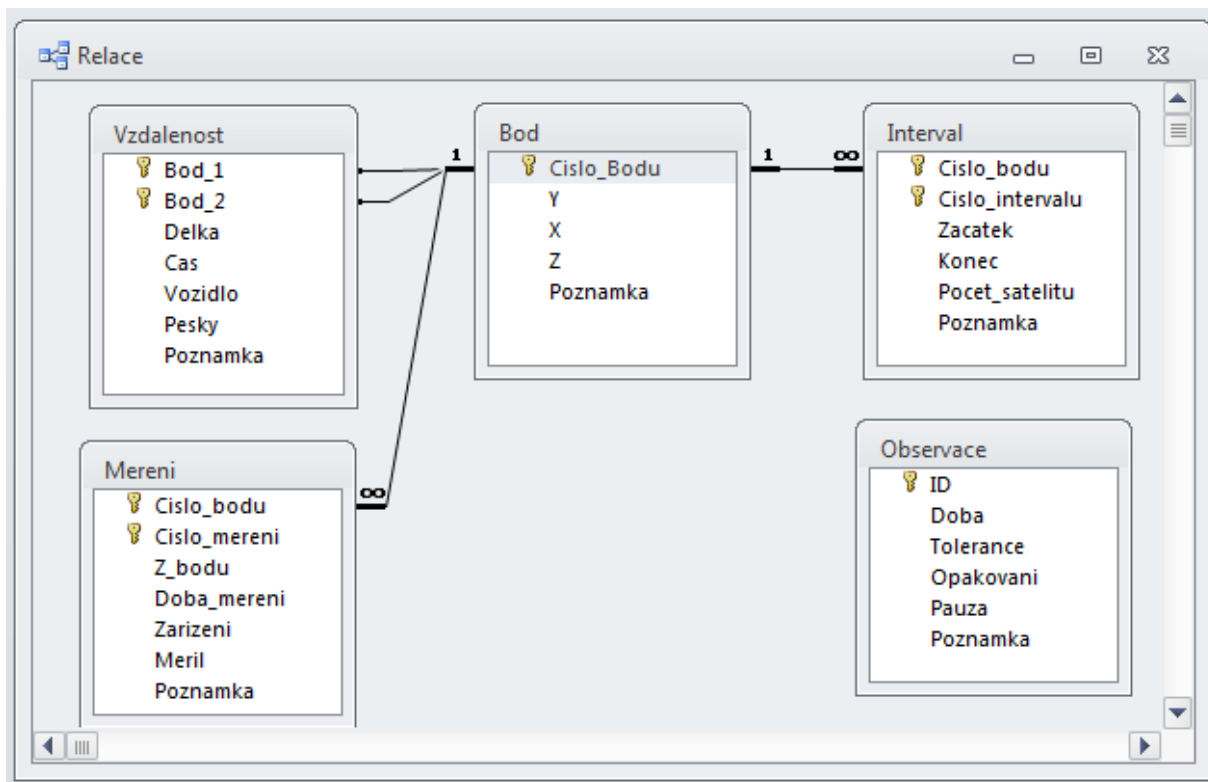


Obr. 10 Datový model projektu

### Transformace datového modelu do tabulek relačního databáze

Transformace z datového modelu do tabulek relačního modelu database proběhla podle standardních pravidel. Databáze byla realizována v aplikaci MS Access – viz obr. 11.





Obr. 11. Relační vazby v datovém modelu

### Algoritmus

1. Zadání výchozího bodu a systémového času.

Je-li zadán bod, začíná se od tohoto bodu, není-li zadán bod, začneme od libovolného bodu (implicitně prvního v pořadí).

Je-li zadán čas, nastaví se systémový čas od zadaného času. Není-li čas zadán, nastaví se jako systémový čas nejnižší čas uvedený v databázi (nejnižší začátek intervalů observací na jednotlivých bodech).

Zadá se počet měření  $N_m$  na jednotlivých bodech a počet aparatur  $N_a$ .

2. Testuje se, zda je úloha řešitelná (úplnost zadání).

- Na všech bodech musí být zadány intervaly vhodné pro měření.
- Zadaný čas musí nižší nebo roven začátku intervalů všech bodů.
- Všechny vzdálenosti mezi zadanými body musí být definovány v délkových i časových jednotkách.
- Musí být zadána doba měření, odstup od posledního měření a tolerance.

3. Vytvoří se dynamická struktura – kalendář událostí. Jde o seznam záznamů uspořádaných podle času. Struktura záznamu: Číslo bodu, systémový čas. Tato dvojice vyjadřuje, že na daném bodě se v daný čas může měřit po určitou dobu.

4. Vyprázdní se tabulka měření (odstraní se všechny záznamy o měření). Z této tabulky se pak sestaví itinerář měření na všech bodech. Tabulka měření se může vytvořit i dynamicky (dočasně).

5. Nastav výchozí bod =  $i$ . bod.

6. Obsluha události (s nejnižším systémovým časem) na  $i$ . bodě – tj. na  $i$ . bodě proběhlo měření. Zaznamenej do tabulky měření (číslo bodu, datum začátku měření, číslo předchozího bodu, označení přístroje a jméno měřiče), že na  $i$ . bodě proběhlo měření (událost). Událost vyřaď z kalendáře.

7. Aktualizuj systémový čas na hodnotu:  $t_{sys} = t_{sys} + t_m$

8. Aktualizuj čas blokování měření na  $i$ . bodě na  $t_{iblk} = t_{sys} + t_m + t_p + t_{tol}$

9. Plánování událostí:

Pro všechny  $j$ . následníky  $i$ . bodu vytvoř záznamy s údaji: číslo bodu následníka, čas události takto:

Je-li splněna podmínka:

$(t_{jbeg} \geq t_{sys}) \wedge (t_{jbeg} \geq t_{sys} + t_{ij}) \wedge (t_{jend} \leq t_{sys} + t_{ij} + t_m) \wedge (t_{sys} + t_{ij} \geq t_{jblok}) \wedge (t_{jblok} + t_m + t_{tol} \leq t_{jend}) \wedge (N_j < N_m)$  pak:

- naplánuj událost měření na  $j$ . bodě v čase  $t = t_{sys} + t_{ij}$
- událost zařaď do kalendáře podle času
- aktualizuj čas blokování měření na  $j$ . bodě  $t_{jblk} = t_{sys} + t_m + t_p + t_{tol}$

Vysvětlivky:

$t_{sys}$  je systémový čas,

$t_m$  je doba měření,

$t_{iblk}$  je čas kdy je možné na daném  $i$ . bodě opět měřit,

$t_p$  je čas odstupe od měření (pause),

$t_{tol}$  je čas tolerance,

$t_{jbeg}$  je čas začátku intervalu, kdy jsou vhodné podmínky pro měření na  $j$ . bodě,

$t_{jend}$  je čas konce intervalu, kdy jsou vhodné podmínky pro měření na  $j$ . bodě,

$t_{ij}$  je čas, který je zapotřebí pro přemístění z bodu  $i$  do bodu  $j$  (úměrný vzdálenosti mezi body),

$N_j$  je počet měření, která proběhla na  $j$ . bodě,

$N_m$  je počet opakování měření na bodech, které bylo zadáno.

10. Není-li splněna podmínka měření pro  $j$ . bod předchozím kroku, vyhledej pro  $j$ . bod v pořadí další takový interval observace, pro který je podmínka splněna a naplánuj událost.

11. Nachází-li v kalendáři událost pro daný bod ve vyšším systémovém čase než je právě plánovaná událost, událost s vyšším časem se v kalendáři zruší. V kalendáři může být pro každý bod jen jediná událost!

12. Cyklus pro daný počet aparatur  $N_a$ :

Opakuj obsluhu událostí v kalendáři:

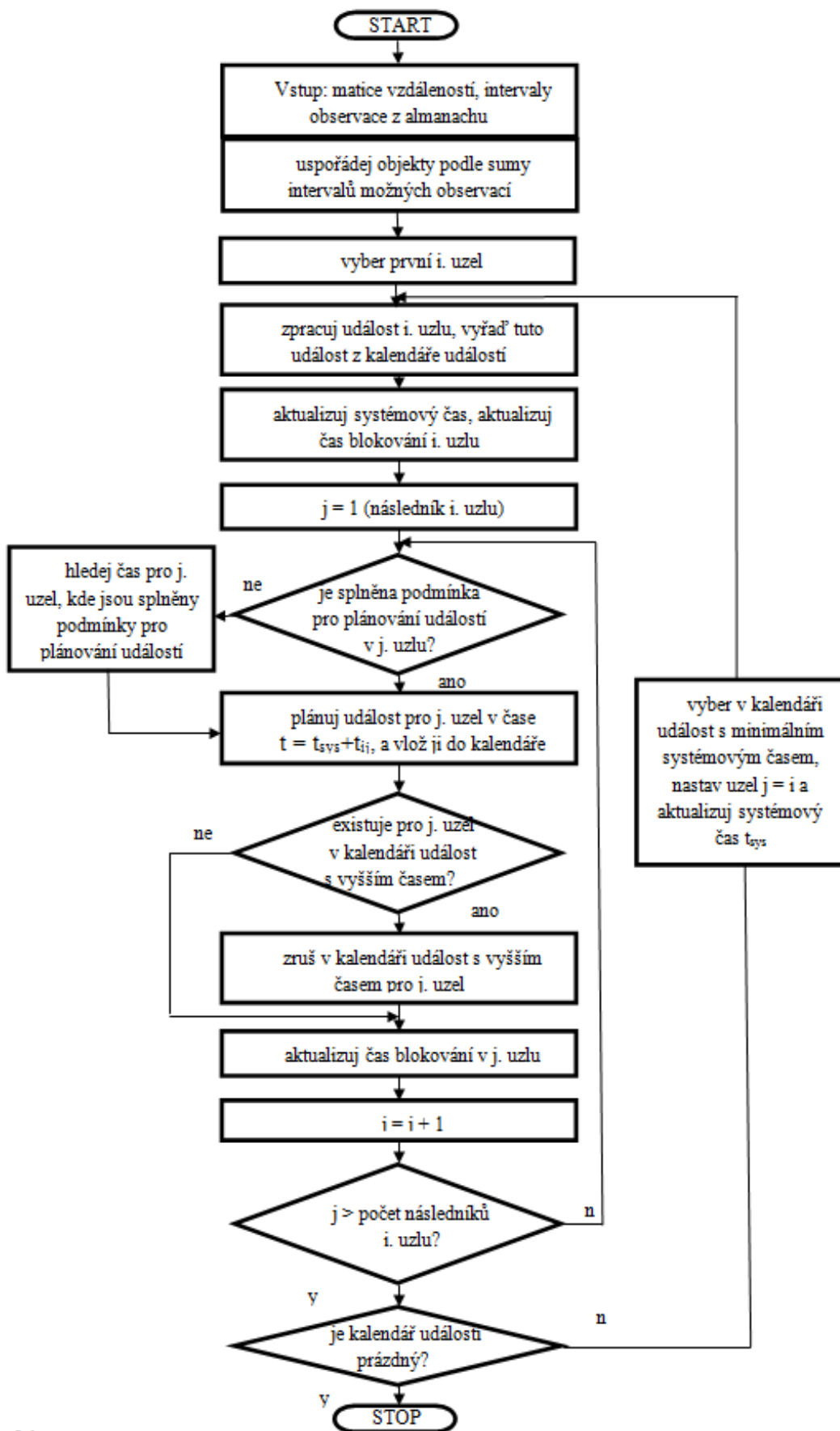
Je-li kalendář událostí prázdný, jdi na krok 13.

Není-li kalendář prázdný, vyber z něj událost s nejnižším časem. Nechť je to událost na bodě  $i$ . Aktualizuj systémový čas na čas této události. Jdi na krok 6.

13. Sestav itinerář plánu měření podle údajů v databázi (tabulky body, vzdálenosti a měření).

14. Konec algoritmu

Vývojový diagram algoritmu je na obr. 12.



Obr. 12. Algoritmus hledání optimální cesty při měření metodou RTK

### 3.3.4 Využití v praxi

Navržená metoda měření GNSS RTK byla aplikována na několika železničních tratích v ČR (Beroun – Rakovník, Opava – Krnov, hranice –Valašské Meziříčí). Ukázky v dalším textu jsou z úseku železniční tratě Hranice – Valašské Meziříčí. Na obr. 13 je výstup z aplikace, která byla vytvořena v Borland Delphi. Obsahuje varianty itineráře s pořadím bodů, na nichž se má měřit. Optimální varianta je na prvním řádku. Celková délka cesty v této variantě je 1563 m a celková doba měření činí 2 hod. 10 min. Všechny varianty jsou uvedeny v tabulce č. 1. Na obr. 14 je část on-line geo-web aplikace. Jde o výřez webového prohlížeče s mapou a značkami měřených bodů v dané lokalitě. Geo-Web aplikace se vyvolá z programu (viz obr. 13) a je vytvořena v Java Scriptu.

Variant	Duration/Dis	Time/Node	Time/Node	Time/Node	Time/Node	Time/Node	Time/Node	Time/Node	Time/Node	Time/Node	Time/Node	Time/Node
Optimal variant	Var. 1 - Time	2:10:00	1.5.2015 7:0	1.5.2015 7:1	1.5.2015 7:3	1.5.2015 7:4	1.5.2015 8:0	1.5.2015 8:1	1.5.2015 8:2	1.5.2015 8:4	1.5.2015 8:5	1.5.2015 9:1
Total distance = 1563 [m]	Var. 1 - Node	1563	0992	0991	0990	0960	0959	1111	1103	1102	1100	1099
Time duration = 2:10:00	Var. 2 Time	2:20:00	1.5.2015 7:0	1.5.2015 7:1	1.5.2015 7:4	1.5.2015 7:5	1.5.2015 8:1	1.5.2015 8:2	1.5.2015 8:3	1.5.2015 8:5	1.5.2015 9:0	1.5.2015 9:2
	Var. 2 Node	1789	0991	0992	0990	0960	0959	1111	1103	1102	1100	1099
	Var. 3 Time	2:30:00	1.5.2015 7:0	1.5.2015 7:1	1.5.2015 7:3	1.5.2015 8:0	1.5.2015 8:2	1.5.2015 8:3	1.5.2015 8:4	1.5.2015 9:0	1.5.2015 9:1	1.5.2015 9:3
	Var. 3 Node	2032	0990	0991	0992	0960	0959	1111	1103	1102	1100	1099
	Var. 4 Time	2:40:00	1.5.2015 7:0	1.5.2015 7:1	1.5.2015 7:3	1.5.2015 7:4	1.5.2015 8:3	1.5.2015 8:4	1.5.2015 8:5	1.5.2015 9:1	1.5.2015 9:2	1.5.2015 9:4
	Var. 4 Node	2159	0960	0990	0991	0992	0959	1111	1103	1102	1100	1099

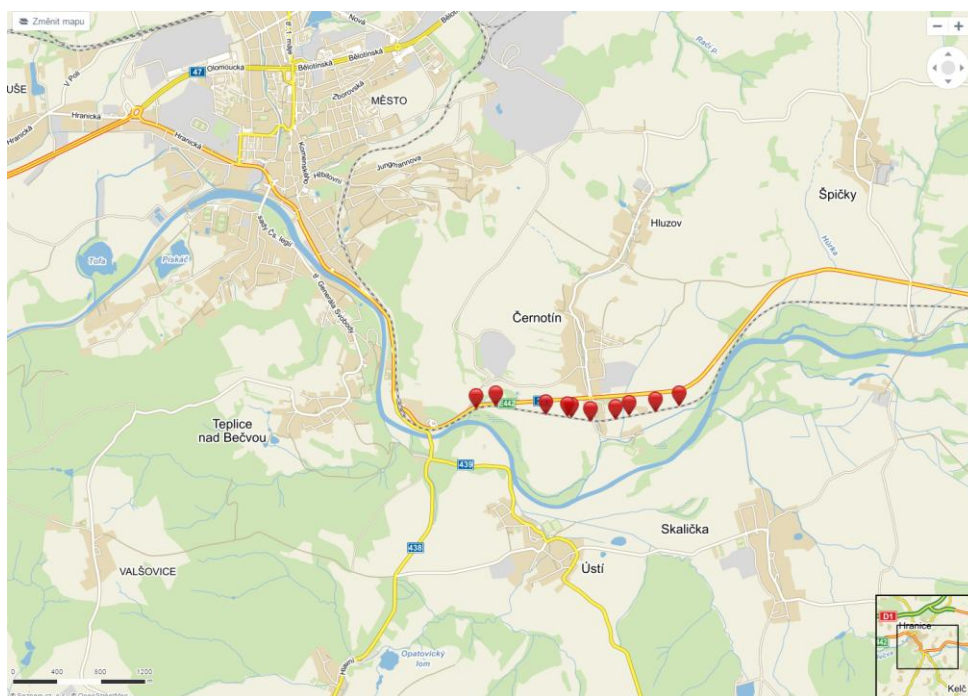
Obr. 13. Varianty itineráře v programu

Tabulka 1 obsahuje navržené varianty Hamiltonovské cesty na železniční trati Hranice na Moravě – Valašské Meziříčí. Varianty jsou seřazeny podle délky a času.

Tab. 1. Varianty Hamiltonovské cesty na železniční trati Hranice na Moravě–Valašské Meziříčí

Order	Variant of paths										Total distance [km]	Total time [hh:mm]
1	992	991	990	960	959	B	1103	1102	1100	1099	1,563	2:10
2	1099	1100	1102	1103	B	959	960	990	991	992	1,563	2:10
3	1100	1099	1102	1103	B	959	960	990	991	992	1,741	2:20
4	991	992	990	960	959	B	1103	1102	1100	1099	1,789	2:20
5	1102	1100	1099	1103	B	959	960	990	991	992	1,883	2:30
6	990	991	992	960	959	B	1103	1102	1100	1099	2,032	2:30
7	1103	1102	1100	1099	B	959	960	990	991	992	2,075	2:40
8	960	990	991	992	959	B	1103	1102	1100	1099	2,159	2:40
9	B	1103	1102	1100	1099	959	960	990	991	992	2,195	2:45
10	959	B	1103	1102	1100	1099	960	990	991	992	2,295	2:55

Metoda založená na kalednáři událostí prakticky představuje transformaci ze dvou domén (prostoru a času) do jedné – časové. Tento postup je vhodný pro přibližné lineární rozložení bodů, na kterých se má měřit. Pro obecné rozložení bodů jsou v aplikaci další 2 algoritmy pro hledání Hamiltonovské cesty v grafu. Jeden z nich je robustní [13], který předpokládá nízký stupeň propojení uzlů hranami, druhý je heuristický [17], který dává rychlé (ale jen přibližné) výsledky i při konfiguraci úplného grafu. Tuto problematiku publikoval autor v pracích [13] – [19].



Obr. 14. Geo-web aplikace (měřené body v terénu)

## 4 DISKUSE A ZÁVĚR

Ukázky řešených projektů v předchozí kapitole potvrdily, že za určitých podmínek lze řešit rozsáhlé úlohy nebo projekty v GIS i na běžných technických prostředcích a s podporou standardního programového vybavení. Prezentované projekty pokrývaly 3 základní typy úloh (viz kap. 2):

1. zpracování analýz na rozsáhlém území (např. celá ČR). V tomto případě se doporučuje rozdělit dané území na dílčí části a ty pak zpracovávat souběžně (paralelně). Přitom je nutné dodržet tyto podmínky:
  - výsledky z dílčích částí musí být na sobě navzájem relativně nezávislé,
  - výsledky analýz lze jednoduše sestavit do jednoho společného výsledku ve formě vhodné datové sady,
  - pro stanovení základních rozhodnutí je zapotřebí řešit zkušební (pilotní) projekt na stejném nebo podobném území jako se stejným zadáním a výsledky jako hlavní projekt,
  - počet dílčích částí (segmentů – rozklad podle množiny  $R$ ) se stanoví podle vztahů (3) a (4) tak, aby zpracování bylo efektivní (poměr doby vlastního zpracování a doby režie),
  - k odhadu časové náročnosti celkového projektu je možné použít vztah (7),
  - pro výběr vhodného technického a programového vybavení co do výpočetní výkonnosti lze použít koeficientu územní podrobnosti  $\delta$  – viz vztah (5),
  - ke zpřesnění výsledků prostorových analýz se doporučuje použít maximálně 1 referenční datovou sadu – viz vztah (6). Podle zkušeností autora využití dalších datových sad neúměrně zatěžuje prostředky, prodlužuje výpočet a na kvalitu výsledků nemá zásadní vliv,
  - po dokončení řešení v každém dílčím segmentu se doporučuje kontrola výsledků a v závěru pak sestavení všech výsledků do společného celku,
  - úspěšné řešení je podmíněno dobrou organizací práce všech členů řešitelského týmu.

2. úlohy s primárně značným objemem vstupních dat, ale s vysokou redundancí nebo s daty, která jsou na sobě relativně závislá. Tato část se týká zejména dat pořízených pozemním nebo leteckými laserovým skenováním. Vstupní data obsahují vysoký stupeň redundance, takže základní metodou, jak je zpracovat na běžných prostředcích je účelová filtrace. Vzhledem k omezení dat by filtrace měla být dostatečně rychlá a účinná. V projektu v kap. 3.2. bylo využito agregační funkce, čímž se objem dat zredukoval asi 100 x. Dále se předpokládá aplikace automatizované nebo poloautomatizované vektorizace. Tato problematika dosud nebyla zatím obecně uspokojivě vyřešena.
3. projekty s relativně malým objemem primárních dat, ale složitým algoritmem řešení, který je většinou založen na prohledávání rozsáhlého stavového prostoru, jenž je tvořen kombinací vstupních dat. Tato třída úloh se vzhledem ke své primární složitosti řeší optimalizačními nebo heuristickými metodami. V mnoha případech jde prakticky o transformaci z vyšší do nižší dimenze popř. o sloučení dimenzí. V demonstračním projektu popsáném v kap. 3.3. šlo o transformaci z prostorové a časové domény do časové domény – tj. sloučení domén. Postup je vhodný pro přibližné lineární uspořádání prvků (geografických objektů) v terénu. Pro obecnou konfiguraci prvků jsou připraveny další 2 algoritmy: jeden robustní dávající přesné výsledky, druhý heuristický, který poskytuje jen přibližné řešení, ale je velmi rychlý i pro složitou konfiguraci prvků s mnoha vzájemnými vazbami (např. pro úplný graf). V reálném světě se však velmi zřídka vyskytují úlohy se složitým stavovým prostorem, protože mnoho kombinací v dané úloze je buď relevantní nebo nenastanou vůbec. Proto je praktická složitost kombinatorických úloh vždy nižší než složitost teoretická a tedy každou kombinatorickou úlohu lze automatizovaně (na počítači) řešit aplikací vhodné metody.

## 5 KONCEPCE DALŠÍ ČINNOSTI

### 5.1 VĚDECKÁ ČINNOST

Jak již bylo uvedeno v úvodu, jsme v současné době svědky rychlého růstu technologické vyspělosti zařízení pro sběr dat v terénu. Paralelně s tímto jevem však metody, implementované v programovém vybavení v oblasti geografických informačních systémů, poněkud stagnují. Důsledkem je nárůst objemu prostorových dat, které jsou jen velmi málo využívány v praxi. Autor by se dále chtěl věnovat této problematice v dříve uvedených tématech. Jde především o metody, kterými by bylo možné efektivně zpracovávat velká data v rozsáhlých projektech GIS s podporou běžně dostupných technických a programových prostředků.

V případě, že lze modelované území rozdělit do menších územních celků (segmentů), pak bude zapotřebí navrhnout metody pro:

- rychlé rozhodnutí o tom, jakou velikost mají tyto segmenty mít,
- stanovení vhodného technického a programového vybavení pro paralelní zpracování,
- zpřesnění výsledků prostorových analýz,
- pokud možno co největší automatizovanou kontrolu dílčích výsledků,
- efektivní řízení celého projektu.

Pokud bude podstatou projektu velký objem vstupních dat s vysokou redundancí nebo s daty, která jsou na sobě relativně závislá, pak je zapotřebí navrhnout metody pro:

- rychlou účelovou filtraci vstupních dat,
- rychlou automatizovanou hranovou detekci ve 3D,
- automatizovanou vektorizaci z filtrovaných dat.

U projektů s relativně malým objemem primárních dat, ale složitým algoritmem řešení, který je většinou založen na prohledávání rozsáhlého stavového prostoru, jenž je tvořen kombinací vstupních dat se očekává:

- další vývoj nových rychlých optimalizačních metod,
- návrh nových rychlých heuristických metod,
- návrh nových metod vhodných pro transformaci problému do nižší domény s cílem zjednodušení problematiky a po vyřešení úlohy zpětnou transformací do vyšší domény získat požadovaný výsledek.

Výsledky vědecké a výzkumné činnosti budou prezentovány ve vědeckých a odborných časopisech, na tuzemských i mezinárodních konferencích a budou také využívány ve výuce.

## 5.2 PEDAGOGICKÁ ČINNOST

V pedagogické činnosti bude zapotřebí připravovat studenty na řešení výše uvedených problémů tj. zpracování velkých dat a souvisejících problémů. Jedním z nich je i tzv. zpětné (reverzní) inženýrství, které se v geodetické praxi projevuje při zaměření existujících staveb (např. zaměření skutečného provedení stavby). Stávající problém se projevuje tím, že zejména u zaměření historických staveb je velmi obtížné, importovat výsledky zeměměřické činnosti v terénu do vhodného software. Současně dostupné programy např. AUTOCAD nemají dostatečné funkce ve 3D pro interpretaci netypických geometrických tvarů (např. obecné plochy v prostoru, obtížně definovaná tělesa, aplikace textur a fotografií do volných ploch atd.). Autor vidí dočasné východisko z této situace využíváním Open-source programů, kde si chybějící funkce může uživatel naprogramovat sám. V této souvislosti je nutné vést studenty ke konstruktivnímu myšlení a algoritmizaci úloh, evokovat v nich nápady a netradiční řešení. Dalším aspektem pedagogické činnosti je propojení teorie s praxí. To znamená, že každý poznatek či nová myšlenka by měla být koncipována tak, aby pokud možno našla přímou využitelnost v praktických projektech a zakázkách souvisejících se zeměměřickou činností. Tento přístup je plně v souladu se státem podporovaným vzděláváním v oblasti prostorových dat (GeoInfoStrategie).

Takto připravení studenti budou schopni nejen vytvořit kvalitní závěrečné práce (bakalářské, diplomové nebo disertační), ale budou i dobře připraveni na úkoly, se kterými se setkají v praxi.

## POUŽITÁ LITERATURA

### Výběr z použité literatury včetně souvisejících prací autora

[1] ALLOMBERT, V.; MICHEA, D.; DUPROS, F.; et al.: An out-of-core GPU approach for accelerating geostatistical interpolation. Edited by: Abramson, D; Lees, M; Krzhizhanovskaya, V.; et al. INTERNATIONAL CONFERENCE ON COMPUTATIONAL SCIENCE, Book Series: Procedia Computer Science, Vol.: 29, pp.: 888-896, 2014.

[2] ARRA, T.: Knowledge, process and product information management to improve decision-making support at your pulp and paper mill. APPITA JOURNAL, Vol.: 56, Issue: 4, pp.: 259-261, 2003.

[3] BAJCSY, P.; VANDECREME, A.; AMELOT, J.; et al.: Terabyte-sized Image Computations on Hadoop Cluster Platforms. Edited by: Hu, X; Lin, TY; Raghavan, V; et al. IEEE International Conference on Big Data, pp. 125 – 131, 2013.

[4] BARTONĚK, D.; OPATŘILOVÁ, I.; BUREŠ, J.: Surface Classification Above Gas Pipeline Facilities. In: Lecture Notes in Geoinformation and Cartography (Editors: Kateřina Růžičková, Tomáš Inspektor). Springer ISBN 978-3-319-18406-7, 978-3-319-18407-4 (eBook), ISSN 1863-2246, 1863-2351, pp. 13-26, 2014.

- [5] BARTONĚK, D. BUREŠ, J., OPATŘILOVÁ, I.: Optimization of Pre-Processing of Extensive Projects in Geographic Information Systems. *Advanced Science Letters*, vol. 20, no. 10-11, pp. 2026-2029, 2014.
- [6] BARTONĚK, D.; BUREŠ, J.; OPATŘILOVÁ, I., Possibilities of Improvement of Image Classification via GIS Tools. 5th International Conference on Cartography and GIS, ISSN 1314-0604, Bulgarian Cartographic Association, Sofia, Bulgaria, 2014.
- [7] 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, In: *Thematic Cartography for the Society*, ISBN 978-3-319-08180-9, ISSN 1863-2246, Springer International Publishing, Switzerland, pp. 95 – 104, 2014.
- [8] BARTONĚK, D.; BUREŠ, J.; OPATŘILOVÁ, I.; VITULA, A., Method of error assessment in image classification. 14th SGEM GeoConference on Informatics, Geoinformatics and Remote Sensing, [www.sgem.org](http://www.sgem.org), SGEM2014 Conference Proceedings, ISBN 978-619-7105-12-4 / ISSN 1314-2704, Vol. 3, pp. 745-752, 2014.
- [9] BARTONĚK, D. BUREŠ, J., OPATŘILOVÁ, I.: Enhancement of Image Classification via GIS. *Advanced Science Letters* Vol. 21, No 12, ISSN 1936-6612, pp. 3737-3740, 2015.
- [10] BARTONĚK, D.; BUREŠ, J.; OPATŘILOVÁ, I.: Workflow for Analysis of Enormous Amounts of Geographical Data. *Advanced Science Letters*, Vol. 21, No. 12, ISSN 1936-6612, pp. 3680-3683, 2015.
- [11] BARTONĚK, D., BUREŠ, J.: Algorithmization and Optimization of Processing of Big Geographical Data. *Journal of Computational and Theoretical Nanoscience*, ISSN 1546-1955, EISSN 1546-1963, Vol. 13, No 12, pp. 9098–9104, 2016.
- [12] BARTONĚK, D. The possibilities of big data processing on the desktop computers. In: *The Rise of Big Spatial Data*, Editors: Igor Ivan, Alex Singleton, Jiří Horák, Tomáš Inspektor. Series Title: *Lecture Notes in Geoinformation and Cartography*. Springer International Publishing, eBook ISBN: 978-3-319-45123-7, Hardcover ISBN: 978-3-319-45122-0, Series ISSN: 1863-2246, vol: 1, pages: 408, pp 273 – 287, 2017.
- [13] BARTONĚK, D.: Algorithm for Travelling Salesman Problem. In: *The Role of Service in the Tourism & Hospitality Industry*. Edited by Ford Lumban Gaol and Fonny Hutagalung. CRC Press, Chap. 18, Pages 107–111, Print ISBN: 978-1-138-02736-7, eBook ISBN: 978-1-315-68852-7, 2015.
- [14] BARTONĚK, D.; BUREŠ, J.: Usage of heuristic algorithm for optimization of precise GNSS RTK measurement. *Proceedings of 15th International Multidisciplinary Scientific Geoconference SGEM, Informatics, Geoinformatics and Remote Sensing – Vol. I – Informatics, Geoinformatics, Photogrammetry and Remote Sensing*, pp. 899 – 906, 2015.
- [15] BARTONĚK, D.; BUREŠ, J., Optimization of Route in Field at Precise GNSS Positioning. 6th International Conference on Cartography and GIS, ISSN 1314-0604, pp 752 – 757, Bulgaria, 2016.
- [16] BARTONĚK, D.; BUREŠ, J.; ŠVÁBENSKÝ, O. Optimization of Process Field Measurement GNSS-RTK for Railway Infrastructure. *Solid State Phenomena, Materials Structure & Micromechanics of Fracture VIII*, Vol. 258, pp. 481-484, ISSN: 1662-9779, ISBN-13 978-3-03835-626-4, ISSN 1013-9826, 2017.
- [17] BARTONĚK, D.; BUREŠ, J.; PETRUCHA, J., Fast heuristic algorithm of searching Hamiltonian path in graph. 17th International Multidisciplinary Scientific GeoConference, SGEM Volume 17, Issue 21, Pages 895-902, 2017.
- [18] BARTONĚK, D.: Extended Algorithm for Travelling Salesman Problem with Conditions in Nodes. In: *Social Interactions and Networking in Cyber Society*. Editors: Gaol, Ford Lumban, Hutagalung, Fonny Dameaty (Eds.). Springer Nature Singapore Pte Ltd., eISBN: 978-981-10-4190-7, ISBN: 978-981-10-4189-1, Edition Number: 1, Part II, Pages: 199 – 207, VII, 2017.



[19] BARTONĚK, D.; BUREŠ, J.; ŠVÁBENSKÝ, O. Optimized GNSS RTK measurement planning for effective point occupation via heuristic analysis. *Engineering Computations*, Vol. 34, Issue: 1, pp. 90-104, 2017.

[20] BECKER, P.; PLESEA, L.; MAURER, T.: Cloud Optimized Image Format and Compression. Edited by: Schreier, G; Skrovseth, PE; Staudenrausch, H. 36th INTERNATIONAL SYMPOSIUM ON REMOTE SENSING OF ENVIRONMENT. Book Series: International Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences, Vol.: 47, Issue: W3, pp.: 613-615, 2015.

[21] BOEHM, J.; LIU, K.; ALIS, C.: SIDELOADING – Ingestion of large point clouds into the Aapache Spark Big Data Engine. Edited by: Halounova, L; Li, S; Safar, V; et al. XXIII ISPRS Congress, Commission II, Book Series: International Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences, Vol.: 41, Issue: B2, pp.: 343-348, 2016.

[22] BUREŠ, J.; BARTONĚK, D.; BÁRTA, L.; ŠVÁBENSKÝ, O.: Automated updating of selected technical and cadastral map components. *Proceedings, 7th International Conference on Cartography and GIS*, Sozopol, Bulgaria, ISSN: 1314-0604, Eds: Bandrova T., Konečný M., pp. 888 – 895, 2018.

[23] BYUN, H., R.; PARK, J., H.; JEONG, Y., - S.: Optional Frame Selection Algorithm for Adaptive Symmetric Service of Augmented Reality Big Data on Smart Devices. *SYMMETRY-BASEL*, Vol.: 8, Issue: 5, Article Number: 37, 2016.

[24] ČERBA, O.: Identické vazby propojených prostorových dat. Habilitační práce. Vysoké učení technické v Brně, Fakulta stavební, 211 str. + přílohy, 2017.

[25] DEHSANGI, M.; ASYABI, E.; SHARIFI, M.; et al.: cCluster: A Core Clustering Mechanism for Workload-Aware Virtual Machine Scheduling. Book Author(s): Awan, I; Younas, M; Mecella, M. 2015 3rd INTERNATIONAL CONFERENCE ON FUTURE INTERNET OF THINGS AND CLOUD (FICLOUD) AND INTERNATIONAL CONFERENCE ON OPEN AND BIG (OBD), pp.: 248-255, 2015.

[26] DENGEL, A.: The semantic desktop as a means for personal information management. Edited by: Fred, A. Kdir: *PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND INFORMATION RETRIEVAL*, pp.: 125-135, 2009.

[27] DIXON, M.: A Pattern Oriented Approach for Designing Scalable Analytics Applications. Edited by: Jann, J; Moreira, J; Kumar, M. 2nd Workshop on Parallel Programming for Analytics Applications (PPAA 2015), pp.: 4-8, 2015.

[28] ELLIOTT, J.; de SOUZA, R. S.; KRONE-MARTINS, A.; et al.: The overlooked potential of Generalized Linear Models in astronomy-II: Gamma regression and photometric redshifts. Author(s): COIN Collaboration. *ASTRONOMY AND COMPUTING*, Vol.: 10, pp.: 61-72, 2015.

[29] ELSTER, A., C.: High-performance computing: Past, present, and future. Edited by: Fagerholm, J; Haataja, J; Jarvinen, J; et al. *APPLIED PARALLEL COMPUTING: ADVANCED SCIENTIFIC COMPUTING*. Book Series: LECTURE NOTES IN COMPUTER SCIENCE, Vol.: 2367, pp.: 433-444, 2002.

[30] GAO, S.; LI, L.; Li, W.; et al.: Constructing gazetteers from volunteered Big Geo-Data based on Hadoop. *COMPUTERS ENVIRONMENT AND URBAN SYSTEMS*, Vol.: 61, Special Issue, pp: 172-186, 2017.

[31] GODZARIDIS, E.; BOISVERT, S.; XIA, F.; et al.: Human analyst at superhuman scales: What Has Friendly Software To Do? *Big Data*, Vol.: 1, Issue: 4, pp.: 227-236, 2013.

[32] GOPU, A.; HAYASHI, S.; YONUG, M., D.; et al.: Trident: Scalable Compute Archives - Workflows, Visualization, and Analysis. Edited by: Chiozzi, G; Guzman, JC. *SOFTWARE AND CYBERINFRASTRUCTURE FOR ASTRONOMY IV*, Book Series: Proceedings of SPIE, Vol.: 9913, 2016.

[33] GUNAWAN; F. X.; SETIAWAN, E., I.: GeoJSON Web Service Based Road Assets Management System for Surabaya City Using Mobile GPS. Book Group Author(s): IEEE, 20th

INTERNATIONAL COMPUTER SCIENCE AND ENGINEERING CONFERENCE (ICSEC), Book Series: International Computer Science and Engineering Conference, 2016.

[34] HELBIG, C.; BAUER, H.-S.; RINK, K.; et al.: Concept and workflow for 3D visualization of atmospheric data in a virtual reality environment for analytical approaches. ENVIRONMENTAL EARTH SCIENCES, Vol.: 72, Issue: 10, pp.: 3767-3780, 2014.

[35] HERMAN, N., S.: An introduction to spatial and mobile object detection, spaial analysis and visualization using geographical information system and remote sensing. Author(s): IEEE, PROCEEDINGS OF 11th INTERNATIONAL CONFERENCE ON INFORMATION & COMMUNICATION TECHNOLOGY AND SYSTEMS, Book Series: International Conference on Information & Communication Technology and Systems, pp.: 1-2, 2017.

[36] HWIRIM, B.; HYUK, P., J.; YOUNG-SIK, J.: Optional Frame Selection Algorithm for Adaptive Symmetric Service of Augmented Reality. Big Data & Smart Devices, IEEE, Vol.: 8, Issue: 5, Article Number: 37, 2016.

[37] HYUN-WOO, K.; HYUK, P., J.; YOUNG-SIK, J.: Human-intelligence workflow management for the big data of augmented reality on cloud infrastructure. NEUROCOMPUTING, Vol.: 279, Special Issue, pp.: 19-26, 2018.

[38] KHOMOTCHOUK, B., B.; HENNESSY, J., R.; WAHLESTEDT, C.: shinyheatmap: Ultra fast low memory heatmap web interface for big data genomics. PLOS ONE, Vol.: 12, Issue: 5, Article Number: e0176334, Published: 2017.

[39] KIM, H.-W.; PARK, J. H.; MAJIGSUREN, D.; et al. Efficient Sustainable Operation Mechanism of Distributed Desktop Integration Storage Based on Virtualization with Ubiquitous Computing. SUSTAINABILITY, Vol.: 7, Issue: 6, pp.: 7568-7580, 2015.

[40] KIM, H.-W.; PARK, J., H.; JEONG, Y.-S.: Human-centric storage resource mechanism for big data on cloud service architecture. JOURNAL OF SUPERCOMPUTING, Vol.: 72, Issue: 7, pp.: 2437-2452, 2016.

[41] LI, J.; ZHANG, T.; LIU, Q.; et al.: Predicting the visualization intensity for interactive spatio-temporal visual analytics: a data-driven view-dependent approach. INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE, Vol.: 31, Issue: 1, pp.: 168-189, 2017.

[42] LLORACH, G.; AGENJO, J.; EVANS, A.; et al.: Web-based data visualization of an MMO virtual regatta using a virtual globe. Edited by: Spencer, SN. 20th International ACM Symposium on 3D Web Technologies (Web3D), pp.: 97-100, 2015.

[43] LONGLEY, P., A.; GOODCHILD, M.; MAGUIRE, D., J.; RHIND, D., W.: Geographic Information Systems and Science.. John Wiley & Sons, Aug. 9, 2010 - Technology & Engineering, 560 pages, ISBN 978-0-470-72144-5.

[44] LOPEZ, V., M.; SANZ CONTRERAZ, M. C.; VAZQUEZ, V., A.; et al.: Drawing by computer BIM, architects and architecture firms tend to rely on 3D printing. Edited by: Chova, LG; Martinez, AL; Torres, IC. 11th INTERNATIONAL TECHNOLOGY, EDUCATION AND DEVELOPMENT CONFERENCE, Book Series: INTED Proceedings, pp: 3582-3590, 2017.

[45] MANNA, Z.: Matematická teorie programů. SNTL Praha, 468 str., DT: 519.681(075.8), 1981.

[46] MURACEVIC, D.: Development of Web GIS application using open geospatial standards. Book Group Author(s): ASME. 4th INTERNATIONAL CONFERENCE ON ADVANCED COMPUTER THEORY AND ENGINEERING ( ICACTE 2011), pp.: 299-303, 2011.

[47] NAIK, N.; JENKINS, P.; SAVAGE, N.; et al.: Big Data Security Analysis Approach Using Computational Intelligence Techniques in R for Desktop Users. PROCEEDINGS OF 2016 IEEE SYMPOSIUM SERIES ON COMPUTATIONAL INTELLIGENCE (SSCI), 2016.

[48] NESHATPOUR, K.; MALIK, M.; HOMAYOUN, H.: Accelerating Machine-Learning Kernels in Hadoop Using FPGAs. Book Group Author(s): IEEE. 15th IEEE/ACM INTERNATIONAL SYMPOSIUM ON CLUSTER, CLOUD AND GRID COMPUTING, Book

Series: IEEE-ACM International Symposium on Cluster Cloud and Grid Computing, pp.: 1151-1154, 2015.

[49] NIGGEMANN, F.; APPEL, F.; BACH, H.; et al.: Heterogeneous access and processing of geo-data on a cloud based infrastructure delivering operational products. Edited by: Schreier, G; Skrovseth, PE; Staudenrausch, H. 36th INTERNATIONAL SYMPOSIUM ON REMOTE SENSING OF ENVIRONMENT. Book Series: International Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences, Vol.: 47, Issue: W3, pp.: 663-667, 2015.

[50] PERDIKARIS, P.; VENTURI, D.; KARNIADAKIS, G., E.: Multifidelity information fusion algorithms for high-dimensional systems and massive data sets. SIAM JOURNAL ON SCIENTIFIC COMPUTING, Vol.: 38, Issue: 4, pp.: 521-538, 2016.

[51] RAMOS-POLLAN, R.; GONZALEZ, F., A.; CAICEDO, J., C.; et al.: BIGS: A Framework for Large-Scale Image Processing and Analysis Over Distributed and Heterogeneous Computing Resources. Book Group Author(s): IEEE 8th INTERNATIONAL CONFERENCE ON E-SCIENCE (E-SCIENCE), Book Series: Proceeding IEEE International Conference on e-Science (e-Science), 2012.

[52] SHANG, Y.; SHANG, L.; GAO, C.; et al.: Using the High-Level Based Program Interface to Facilitate the Large Scale Scientific Computing. SCIENTIFIC WORLD JOURNAL, Article Number: 914514, 2014.

[53] SEDLMAYR, M.; WUERFL, T.; MAIER, Ch.; et al.: Optimizing R with SparkR on a commodity cluster for biomedical research. COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE, Vol.: 137, pp.: 321-328, 2016.

[54] SHE, B.; BOULANGER, P.; NOGA, M.: Real-Time Rendering of Temporal Volumetric Data on a GPU. Edited by: Banissi, E; Bertsch, S; Burkhard, R; et al. 15th INTERNATIONAL CONFERENCE ON INFORMATION VISUALISATION (IV 2011), Book Series: IEEE International Conference on Information Visualization, pp.: 622-631, 2011.

[55] SUN, S.; MA, D.; SHEN, H.: A Web-based medical image collaborative processing system using mobile devices. Book Group Author(s): ASME. 4th INTERNATIONAL CONFERENCE ON SOFTWARE TECHNOLOGY AND ENGINEERING (ICSTE 2012), pp.: 329-335, 2012.

[56] VASILE, M.-A.; POP, F.; TUTUEANU, R.-I.; et al. Resource-aware hybrid scheduling algorithm in heterogeneous distributed computing. FUTURE GENERATION COMPUTER SYSTEMS-THE INTERNATIONAL JOURNAL OF E-SCIENCE, Vol.: 51, pp.: 61-71, 2015.

[57] VIDLER, J.; RAYSON, P.; ANTHONY, L.; et al.: Dealing With Big Data Outside Of The Cloud: GPU Accelerated Sort. Edited by: Calzolari, N; Choukri, K; Declerck, T; et al. 9th International Conference on Language Resources and Evaluation (LREC), pp. 26-31, 2014.

[58] WU, C.; FU, J.; ZHANG, Z.; et al.: Efficient On/Off-line Query Pre-Processing for Telecom Social Streaming Data. Edited by: Wang, KIK; Jin, Q; Zhang, Q; et al. IEEE 14th INTL CONF. ON DEPENDABLE, AUTONOMIC AND SECURE COMPUTING, 14th INTL CONF ON PERVASIVE INTELLIGENCE AND COMPUTING, 2nd INTL CONF ON BIG DATA INTELLIGENCE AND COMPUTING AND CYBER SCIENCE AND TECHNOLOGY CONGRESS (DASC/PICOM/DATACOM/CYBERSC), pp.: 827-834, 2016.

[59] ZHANG, J.; YOU, S.; GRUENWALD, L.: Tiny GPU Cluster for Big Spatial Data: A Preliminary Performance Evaluation. Book Group Author(s): IEEE. IEEE 35th International Conference on Distributed Computing Systems Workshops (ICDCSW). Book Series: IEEE International Conference on Distributed Computing Systems Workshops, pp.: 142-147, 2015.

[60] ZHANG, Y.; JI, Z.: Membrane Models in Big Data Process on GPU-accelerating Systems. Book Group Author(s): PROCEEDINGS OF 2016 SIXTH INTERNATIONAL CONFERENCE ON INSTRUMENTATION & MEASUREMENT, COMPUTER, COMMUNICATION AND CONTROL (IMCCC 2016), pp.: 422-425, 2016.

[61] <http://www.vugtk.cz/slovník>

## ABSTRACT

At the present time we witness a great development of digital information technologies. This trend can be seen on all the levels of our social life; individual, institutional and the level of state administration etc. The situation encroached on the sphere of the so called spatial data which are the data containing, apart from the attributive part, also localization part which unequally determines the position within an obligatory coordinate system.

These changes resulted in rapid growth of digital data, significantly supported by technical advance of the devices producing these data. At the same time with the growth of advanced technology for making spatial data, methods as well as software for big data processing rather fall behind. Paradoxically out of the total volume approximately about 2% of the data is being made use of. The main reason for this state is the fact that big data processing require in many cases hardware and software of high computational performance. Unfortunately, such an information infrastructure is owned by very limited number of the users.

A proportion between amount of processed data and its total volume would mark an improvement in favour of the used data if big data could be processed even by common users of information technologies. In geographical information systems (GIS) we meet with above mentioned problems either when solving projects related to the extensive territory or considerable secondary complexity which eventually lead to big data processing.

Theoretical part deals with some proposals of methods for processing of extensive projects in GIS on available desktop computers. In order to solve extensive projects in GIS the author has proposed these methods:

1) in case of a large territory it is recommendable to split input data into segments relatively independent on one another and for further processing to make use of parallelism (this method has been made use of in the project „Classification of the surface over gas-pipeline within the Czech Republic“.

If this division is out of question it is necessary to work with the whole set of input data (e.g. in case of network analysis) we can use:

2) purpose – built reduction or optimization of input data by

a) data filtration (this method was used in case of laser scanning)

b) decrease of dimension of problem ie. solution of the task in lower dimension and then the transformation back to a higher dimension. This method has been used when searching an optimal route in field at RTK method of measuring.

In order to get a qualified estimation of relevant parameters (time solution, capacity of sources and the structure of research team), both of these methods require the testing task (pilot project).

The author would like to focus on the creation and verification of methods by which it would be possible to process effectively extensive projects in GIS supported by desktop hardware and software. It is a project of new quick methods for functional reduction of the data volume, optimization of processing, edge detection in 3D and automated vectorization.

As far as a pedagogical activity is concerned it will be necessary to prepare students for the capability to solve the above described problems ie. big data processing and other related tasks. One of them is the so called „reverse engineering“ which, in practice, shows at surveying of existing constructions. In many cases current software is unable to realize more complicated shapes. The author can find the way out of this situation by using open-source software where missing functions can be created by the user himself.

In this context it is important to make the students think constructively, support their own ideas and non-traditional solutions. Another aspect of pedagogical activity is interconnection of theory with practice ie. the usage of pieces of knowledge in practical projects and work to order related to geodesy. This attitude is fully in accordance with the state supported education in the field of spatial information (GeoInfoStrategy project).