

1. Úvod

Kriminalita a trestná činnost se jako socioekonomická aktivita stala nedílnou součástí života kolem nás a naší společnosti, jejímž nerovnoměrným sociálním rozvrstvením je přirozeným následkem. Na rozdíl od jiných ostatních socioekonomických projevů je však chápána jako ryze negativní činnost, omezující a zastašující jev. Zřejmým přáním každého člověka, nebo alespoň většiny, je tedy v rámci naší bezpečnosti její vymýcení. Eliminovat kriminalitu je však samozřejmě nemožné a tak jedním z mála účinných nástrojů zůstává prevence a predikce.

Tam, kde se dříve vyšetřovatelé pátrající po zločincích museli spolehnout na svůj instinkt, se dnes, po několik desítek let trvajícím vývoji geoinformačních technologií používají moderní prostorové analýzy kriminalistických dat, využívající matematické algoritmy a modely k potvrzení či vyvrácení různých hypotéz a určení inkriminovaných míst, na která je potřeba zaměřit preventivní pozornost. S pomocí nástrojů geografických informačních systémů pak tyto analýzy získávají náležitou vizuální podobu a tvoří tak společně účinnou zbraň v boji s kriminalitou.

Hlavním cílem mé práce je jednu z takovýchto analytických pomůcek představit, seznámit s možnostmi jejího použití a na reálných datech spolu s výstupy v programu GIS (geografických informačních systémů) prezentovat. Jedná se o prostorově statistický program **CrimeStat** vytvořený Nedem Levinem a jeho kolegy v národním právním institutu ve Washingtonu. Grafické výstupy analýz budou pak graficky zobrazovány pomocí softwaru **ArcView GIS 3.2a** firmy ESRI (Environmental Systems Research Institute).

2. Literární rešerže

2.1 Geografické informační systémy – GIS

Zkratka GIS je odvozena z anglického výrazu Geographical Information Systems, který se v české literatuře a prostředí vyskytuje pod názvem geografické informační systémy. Vzhledem k tomu, že jde o poměrně mladou a obsáhlou vědní disciplínu, zahrnující široké spektrum činností, neexistuje v současné době jednotná definice tohoto oboru. Z více než desítky, s nimiž jsem se za dobu studia této problematiky setkal, vybírám a za obecně použitelnou považuji tuto:

„GIS je organizovaný soubor počítačového hardware, software, geografických údajů (naplněné báze dat) a lidských zdrojů navržený na efektivní získávání, ukládání, upravování, obhospodařování, analyzování a zobrazování všech forem geografických informací“ (internetová stránka firmy ESRI).

Po uvedení této definice jistě vyvstane na mysli otázka, co je to geografická informace. Ještě před vymezením tohoto pojmu je potřeba uvést pojem úzce související a to **geoobjekt – prostorový objekt**. Geoobjekt je reálný, nebo imaginární objekt, který se vztahuje (popisuje) k části prostoru na povrchu Země. Je ho možné odlišit od jiných geoobjektů pomocí jeho prostorové polohy (geometrie), polohových vztahů k jiným objektům (topologie), relevantních charakteristik (atributů) a temporálních změn (dynamiky).

Geografická informace – je geometrický, topologický, atributový a dynamický popis geoobjektu s ohledem na potřeby subjektivně určené aplikace. Informace přitom (podle téhož autora) může být uložena a zpracována v různých formách a to **v numerické formě** pro statistickou analýzu, **v textových dokumentech**, **v obrazových informacích** či **v multimediální formě**. V konečném pojetí je však možné všechny uvedené formy zredukovat, za využití výpočetní techniky, do **digitální formy** (Streit, 1997).

Geografické informační systémy lze rovněž chápat ve třech odlišných rovinách. Jako **technologie**, jako **aplikační nástroj** a jako **vědecký obor**. Pod GIS jako **technologii** rozumíme prostředky nevyhnutelné pro realizaci a provoz aplikace: hardwarové a softwarové vybavení. Pod GIS jako **aplikací** si představujeme informační systém „geografického typu“, který je součástí řízení jisté organizační jednotky (např. podniku, městského úřadu). Složitý je pohled na GIS jako **vědecký obor** (Tuček, 1998).

V souvislosti s **vědecko-technickým** chápáním GIS je třeba zdůraznit klíčovou úlohu GIS v geoinformatice, environmentálních vědách a prakticky ve všech oblastech monitorování, plánování a řízení životního a přírodního prostředí (Streit, 1997).

GIS umožňuje pořizovat, využívat a aktualizovat rozsáhlé databáze s tématicky různorodými prostorovými daty. Dovoluje analyzovat a kombinovat údaje obsažené v databázi, a tím určovat polohu, kvalitu, stav, vývoj či možné důsledky nejrůznějších modelových situací. Tyto rozsáhlé dovednosti jsou povětšinou vyjadřovány čtyřmi základními otázkami:

Poloha: co se nachází na ...?

Dotaz se snaží zjistit, co se nachází na konkrétním místě. Polohu lze popsat mnoha různými způsoby, např. místopisným názvem, zeměpisnou šířkou a délkou, souřadnicemi X,Y,Z.

Podmínka: kde je to?

Zodpovězení tohoto dotazu již vyžaduje prostorovou analýzu. Na základě zadání podmínky týkající se kvality či stavu je vyhledáno místo, které tuto podmínku splňuje.

Trendy: co se změnilo od ...?

Dotaz v sobě zahrnuje oba předchozí dotazy a snaží se zjistit změny v dané oblasti v průběhu času.

Modelování: co když ...?

Tento dotaz se pokládá při potřebě předpovědi důsledků (Tollingerová, 1996).

2.2 Mezníky vývoje GIS

Vzhledem k výše naznačeným faktům je patrné, že jsou GIS, jaké je známe dnes, velmi úzce spojeny s výpočetní technikou a proto patří mezi velmi mladé vědní disciplíny, jejíž historie se vyvíjela ruku v ruce s rozmachem počítačů. S tímto rychlým rozvojem vědy a techniky vzrůstá množství informací i požadavky na jejich zpracování.

Druhá polovina 20. století je často charakterizována jako období informační exploze spojené s pronikáním počítačů do každodenního života. Stávající databanky jsou převáděny do digitální podoby, která umožňuje rychlejší a pohodlnější manipulaci s velkým objemem informací. Vyvíjejí se nové technologie pro zpracování dat. Mezi ně patří i geografické informační systémy. V současnosti zájem o GIS rychle roste a jejich aplikace pronikají do

nejrůznějších sfér lidské činnosti (Tollingerová, 1996). Počítačové GIS se začaly používat v 60. letech, přičemž jejich manuální vzory se objevily o 100 let dříve. Je doloženo, že většina, ale to neznamená že všechna, významná řešení v oblasti GIS se uskutečnila v severní Americe. Mimořádnou úlohu sehrály organizace jako US Bureau of the Census, US Geological Survey, Harvard Laboratory for Computer Graphics nebo Environmental Cartography Unit. Příkladem působení komerčního sektoru může být firma Environmental System Research Institute (Tuček, 1998).

Historii GIS je možné rozdělit na 4 období. **První pionýrské** začíná počátkem 60. let a trvá přibližně do roku 1975. V tomto období je mimořádně důležitý vliv jednotlivých průkopnických osobností a institucí, zvláště univerzit, na dosažené výsledky. V **druhé fázi**, přibližně od roku 1973 do začátku 80. let, dochází k ujednocení pokusů a činností agenturami a institucemi na lokální úrovni. Ve **třetí fázi**, přibližně od roku 1982 do konce 80. let, dominuje komercializace problematiky. Ve **čtvrté, současné** fázi dominuje přístup uživatelský, soutěž mezi prodejci (producenty), začínají pokusy o standardizaci. Zvyšuje se důraz na uživatelské chápání GIS a možnosti jeho použití (Maguire, Goodchild, Rhind, 1991).

2.3 GIS a jiné počítačové systémy

Díky širokému rozsahu použití geografických informačních systémů je třeba uvést na tomto místě několik příbuzných a podobných systémů, které se v praxi hojně využívají, přičemž je velmi těžké stanovit, a autoři různých prací se v názoru na tuto problematiku velmi liší, zda jsou GIS nadmnožinou těchto systémů či jejich podmnožinou. Mezi tyto příbuzné systémy patří **CAD** (computer aided designing) a **CAM** (computer aided mapping), dále **DBMS** (systémy řízení bází dat), systémy pro **DPZ** (dálkový průzkum země) a **statistické programy**.

CAD (computer aided designing)

Systémy pro počítačem podporované navrhování a konstruování byly vytvořeny pro návrh a kreslení dvoj a trojrozměrných objektů. Prostřednictvím interaktivní grafiky uživatel buduje a zobrazuje objekty. Tyto systémy se vyznačují tím, že mají obvykle jen symbolické propojení s databází (obsahující např. seznamy materiálů a cen) avšak díky kvalitnímu grafickému editoru a příjemnému uživatelskému prostředí jsou považovány za atraktivní a dokonale

nástroj pro tvorbu a používání grafických údajů. Zde je nutno uvést, že je velký rozdíl mezi grafickými a geografickými údaji, se kterými pracují GIS. V ČR dochází k častému splývání právě těchto dvou pojmů. Důležitým rysem těchto systémů je rovněž fakt, že neumějí pracovat s topologií dat. Nemohou tedy provádět prostorové analýzy.

CAM (computer aided mapping)

Počítačová kartografie se soustřeďuje na sběr, klasifikaci a zobrazování kartografických a geografických údajů. Pozornost je více soustředěna na sběr údajů a přesné geodetické výpočty, než na prostorové analýzy a modelování. CAM systémy mají k dispozici mnoho výkonných technických a programových prostředků na tvorbu kvalitních mapových výstupů.

DBMS (systémy řízení bází dat)

Tyto systémy patří v současnosti ke standardnímu programovému vybavení. Byly navrženy a optimalizovány na vstup, ukládání a výběr neprostorových a negrafických údajů. Možnosti zobrazování a zpracování geografických údajů jsou limitované. Ve větší míře používají datovou strukturu.

Systémy pro DPZ (dálkový průzkum země) a statistické programy

Systémy pro **DPZ** se používají k uchování, zpracování a interpretaci digitálních leteckých a kosmických snímků. Vyznačují se přepracovanými metodami matematického (zvláště statistického) zpracování a klasifikace snímků. Úlohou **statistických programů** je zpracování souborů číselných údajů. Mají mnohé možnosti zpracování, postrádají však prostorové funkce. Často se používají jako výkonný doplněk databázového systému (Tuček, 1998).

2.4 Využití geografických informačních systémů

Jak již bylo zmíněno, nalézají v dnešní době geografické informační systémy širokou škálu uplatnění v mnoha oborech lidského konání. Patří sem zejména tyto:

- 1) Archeologie – systémy se používají k mapování a archivaci archeologických objevů a pro umístění existujících nalezišť do mapy.

- 2) Armáda – využití v oblasti tvorby map terénu, sledování polohy strategických objektů v terénu a při tvorbě simulátorů.
- 3) Doprava – použití při navrhování, stavbě a rozvoji dopravních komunikací. Dále při plánování a údržbě dopravní infrastruktury, plánování přepravy nebezpečných nákladů a navigace vozidel pomocí systémů umístěných přímo ve vozidlech, tzv. AVL (Automatic Vehicle Location).
- 4) Katastr nemovitostí – převod katastrálních map a ostatních informací do digitální podoby a s tím spojené jejich využívání při návrhu pozemkových úprav a přípravě stavebních projektů.
- 5) Lesní hospodářství – zde systémy umožňují existenci lesnických map v digitální podobě a jejich propojení s údaji lesního hospodářského plánu.
- 6) Městské informační systémy – jsou stěžejním prvkem při sledování hustoty dopravy, míry zaměstnanosti, evidence obyvatelstva městských částí, struktury všech spojů a zásobování, ekologických aspektů a napomáhají tak rozhodování v jejich plánování a řízení.
- 7) Obchod a finance – využití k mapování nákupního chování spotřebitelů a spádových oblastí obchodu, čímž napomáhá k optimalizaci rozmístění a velikosti prodejen. Umožňuje též prezentovat výsledky marketingového výzkumu a marketingových studií či preference spotřebitelů. Velký význam má i v pojišťovnictví při evidenci oblastí se zvýšeným pojišťovacím rizikem.
- 8) Policie – policie využívá GIS jak při mapování dopravních problémů a průjezdnosti jednotlivých komunikací, tak i v oblasti kriminalistiky k monitoringu oblastí dle úrovně jejich kriminality a faktorů s kriminalitou spojených.
- 9) Státní správa – použití pro optimální řízení jednotlivých státních resortů, organizaci požární a záchranné služby a policie. Sleduje vývoj jednotlivých regionů a následky příslušné regionální politiky.
- 10) Životní prostředí – využívá se zejména pro vyhledávání zdrojů znečištění, monitorování stavu koncentrace škodlivých látek v ovzduší, sledování a modelování dalších přírodních jevů jako např. postup povodňové vlny, přívalové vlny tsunami, eroze, odlesňování atd. (Tollingerová, 1996).

2.5 Činnosti v rámci projektu GIS

Názory na rozdělení činností v rámci projektu GIS a tím i používaných nástrojů systému na jejich vykonání se liší od aplikace k aplikaci podle potřeb uživatele. Rozdíly mezi nimi však nejsou velké. Kompromisem se pak jeví být toto členění projektu GIS:

1. Stanovení cílů projektu
2. Budování databáze, včetně podrobnějšího členění na:
 - navržení struktury databáze
 - naplnění databáze – vstup údajů
 - kontrola údajů a odstraňování chyb
3. Restrukturalizace nebo manipulace s údaji
4. Vykonání analýz a syntéz:
 - analýzy a syntézy geografické, modelů terénu, statistické, analýzy obrazů
5. Prezentace výstupů

2.5.1 Stanovení cílů projektu

Geografické informační systémy se používají buď v tzv. databázových nebo analytických aplikacích. Podle toho, do které kategorie je možné příslušný projekt zařadit, mohou existovat určité odlišnosti v přístupu k řešení uvedených kroků. V **databázových aplikacích** nám jde hlavně o vytvoření databáze, využitelné širším okruhem uživatelů, databáze průběžně aktualizovatelné, snadno prohledávatelné. Důraz tedy bude na jejím správném navržení a tvorbě univerzálně využitelných výstupů.

V **analytických aplikacích** jsou motivem právě specifické analýzy bez potřeby vazeb na širší okolí, potřeby jiných uživatelů. Uvedená skutečnost do určité míry ovlivňuje i výběr vhodného systému. Obvykle je třeba při stanovování cíle odpovědět na následující otázky:

- Jaký problém budeme řešit?
- Jak se řeší klasicky, bez použití GIS?
- Z jakých zdrojů a za kolik můžeme získat vstupní údaje pro řešení?
- Jaké budou výsledné produkty řešení navrhovaným postupem – zprávy, mapy, tabulkové přehledy?
- Jak často se tyto postupy použijí a s jakým časovým odstupem?
- Budou zpracovávány údaje potřebné i pro jiné účely?
- Komu budou určeny výstupy – specialistům, laické veřejnosti? (Tuček, 1998)

2.5.2 Budování databáze

Jde o kritickou etapu, která vyžaduje velké časové i finanční investice. Vyžaduje zvážení dostupných zdrojů údajů, použitých zařízení a pracovních postupů, především však správné **navržení struktury a parametrů databáze**. Vzhledem k možným negativním dopadům nesprávných rozhodnutí pro používání databáze je třeba, aby cílem správného návrhu bylo:

- zabezpečit splnění cílů aplikace GIS a podpořit její funkčnost
- databáze by měla obsáhnout všechny potřebné údaje, s minimem těch nadbytečných
- struktura by měla umožnit užívání dat různým uživatelům
- aplikace pro vstup a údržbu dat by měly být odděleny od ostatních

Naplnění databáze je časově nejzdlouhavějším, nejkomplicovanějším i finančně nejnáročnějším krokem v rámci projektu GIS. Odhaduje se, že z celkových nákladů na GIS připadá obvykle až 80% na data. Pro vstup lze obecně použít **různé zdroje údajů**. V úvahu přicházejí zvláště mapy různých druhů v analogové i digitální podobě, náčrty v daném souřadnicovém systému, fotogrammetrické podklady, údaje z DPZ v analogové i digitální podobě, údaje z geodetických měření, statistické údaje, digitalizované údaje z jiných systémů, výkresy z CAD systémů atd. Použitelné jsou všechny, je třeba jen zvolit vhodný pracovní postup a použít potřebná technická zařízení. Obecně lze vstup údajů do GIS rozdělit na vstup údajů z **primárních zdrojů** a na vstup údajů ze **sekundárních zdrojů**.

1. Vstup údajů z primárních zdrojů

Primární zdroje jsou zdroje údajů, získané přímým měřením a zjišťováním na geografických objektech. Pokud jde o geometrickou část popisu, jedná se hlavně o výsledky **geodetického měření** různého druhu. Za další, již klasický zdroj údajů pro mapování, se považuje **fotogrammetrie**, jejíž možnosti rozšířilo začátkem 70.let zavedení technologií **DPZ** a otevřelo tak možnost získat primární údaje o rozsáhlých územích.

a) geodetická měření – geodetická měření i v současnosti zůstávají nejpřesnějšími a nejdokonalejšími zdroji geometrické části prostorových údajů. Tak, jak je možné je využít v klasických postupech mapování, jsou vhodné i pro budování digitálních reprezentací reálných objektů. Výsledky měření mohou existovat ve

více formách, zejména však jako **terénní zápisník údajů, elektronický zápisník a GPS (global positioning system)**. Jde o satelitní technologii, ve které se poloha přijímače na libovolném místě na Zemi vypočítává na základě jeho relativní polohy vůči více satelitům.

b) fotogrammetrické metody – jedná se o metody, při nichž se data o rozměru, tvaru a poloze objektů v prostoru nezískávají přímým měřením v terénu, nýbrž vyhodnocením měřičských snímků. Při zpracování a vyhodnocování snímků, představujících **centrální projekci**, je třeba tyto snímky numericky nebo analogově transformovat na tzv. **ortogonální projekci**. To se provádí klasicky graficko-opticko-mechanickými vyhodnocovacími metodami a zařízeními, početními metodami a v poslední době i počítačovými metodami. Základní úlohou fotogrammetrie je tedy převod měřičského snímku do ortogonálního průmětu (mapa, plán).

c) metody DPZ (dálkový průzkum Země) – dálkovým průzkumem se všeobecně míní každé získávání informací o objektu zkoumání z dálky – bez přímého kontaktu s ním. Podle definice přijaté na 16. kongresu ISPRS (International Society for Photogrammetry and Remote Sensing) v roce 1988 je „dálkový průzkum Země umění, věda a technologie na získávání spolehlivých informací o fyzikálních objektech a jejich okolí pomocí záznamu, měření a interpretace snímků a digitálních záznamů, které se získávají pomocí nekontaktních snímacích systémů“.

Přenos informací o zkoumaných objektech zemského povrchu nebo atmosféry k snímacím sensorům, jenž jsou uloženy převážně na létajících tělesech – satelitech a letadlech – vykonává elektromagnetické záření.

2. Vstup údajů ze sekundárních zdrojů

Sekundární zdroje jsou údaje obsažené zejména v kartografických podkladech nejrůznějšího druhu. Označení „sekundární“ znamená, že byly primárně získány měřením, či zjišťováním v procesu mapování – tvorby mapy. Ve finální kartografické podobě může být obsah mapy ovlivňován účelem mapy, použitým postupem mapování, generalizací, měřítkem atd. Bez ohledu na výhody, které mají prvotní údaje, jsou nejčastějším zdrojem informací pro GIS právě již existující mapy. Techniky získávání údajů z kartografických zdrojů ovlivňuje několik faktů. Mapy jsou přesné, vyhotovené v měřítku a jsou snímány zařízeními schopnými rozlišit i velmi jemné detaily. Dále mapy obsahují mnoho symbolů,

jsou mnohobarevné, zobrazení je děleno do mapových listů a podle druhu a účelu mají různou kvalitu. S ohledem na historický vývoj, tradice, velikost území apod. mohou v různých zemích sehrávat uvedené zdroje různé úlohy. V zemích s velkými a relativně řídko osídlenými územími (Kanada, Rusko, USA) mají mimořádnou důležitost údaje z DPZ – jsou aktuální a mohou i pro obrovské území pocházet z jednoho časového horizontu. Avšak v zemích s bohatými tradicemi detailního zjišťování mnoha dat (sem patří i ČR) je vhodné mapová díla převést do digitální polohy.

2.5.3 Restrukturalizace a manipulace s údaji

Postupy restrukturalizace nebo manipulace s údaji se uplatňují na všech součástech geografické informace (popisu geoobjektů). Může se jednat o prostorovou informaci (geometrii), popisné informace (atributy) i prostorové vztahy mezi objekty nebo stavebními prvky reprezentací objektů.

Mezi hlavní úpravy údajů v databázi patří: **změna topologických vztahů, dělení na listy, ztotožnění okrajů, změna velikosti buňky rastru**. V této souvislosti se je také vhodné zmínit o **generalizaci**, tedy zevšeobecnění geografických informací, které se používá při jejich zpracování ze čtyř hlavních důvodů:

- a) Ekonomické požadavky – proces sběru dat v projektu GIS je velice nákladný, proto se z ekonomických důvodů často přikračuje k vytváření databází z již existujících map nebo na základě statistických zjišťování.
- b) Požadavky redukce objemu dat – při vstupu, manipulaci, analýze a syntéze i vytváření výstupů dochází k chybám. Zdroje chyb jsou lidské, instrumentální, stejně jako metodologické a procesní. Generalizace v tomto případě tedy slouží k eliminaci chyb.
- c) Víceúčelovost požadavků pro údaje – oficiální mapovací organizace musí poskytovat informace pro mnohé uživatele, takže je třeba získané informace filtrovat a modelovat podle účelu použití a významnosti pro dané měřítko, aby se dodržovaly ekonomické principy jednoho získávání a mnohoúčelového použití.
- d) Požadavky zobrazování a vnímání dat – jde o potřebnou redukci nebo kompresi dat, protože objem získaných dat je podstatně větší, než je možno přímo vnímat. Takže je nutné použít generalizaci pro výběr, zjednodušení a zvýraznění informace pro účely jejího vnímání a pochopení.

2.5.4 Vykonání analýz a syntéz

Možnost pracovat s nahromaděnými informacemi, provádět na nich analýzu a syntetizovat údaje nové je srdcem celého GIS. Vzhledem k faktu, že je tato kapitola velice obsáhlá a je taktéž hlavní náplní mé práce, rozhodl jsem se tuto problematiku, zejména pak prostorovou analýzu, podrobně rozebrat v další kapitole a zde pouze úzce zmiňuji nejčastější a nejpoužívanější způsob analýzy a to **dotazy do databáze**. Dotazováním vybíráme z databáze údaje, které odpovídají specifickému kritériu nebo podmínce. Dotazovací operace má obecně tři hlavní komponenty:

- specifikaci údajů, kterých se týká
- formulaci podmínek, kterým musí údaje vyhovovat
- instrukci, co se má na vybraných údajích vykonat

Nejzákladnějšími nástroji, které GIS poskytují jsou tedy ty, které umožňují dotazy do databáze, prohledávají ji. Ptáme se zjednodušeně buď na polohu tzv. **prostorový dotaz na databázi** nebo na vlastnosti tzv. **atributový dotaz na databázi**. Ve většině případů bývají dotazy zkombinovány do soustavy podmínek.

2.5.5 Presentace výstupů

Poslední fází budování GIS je prezentace dat. Prezentace se děje buď pouhým výběrem požadovaných dat, nebo demonstrováním výsledků analýz. Data lze prezentovat v digitální formě na obrazovce počítače nebo ve formě grafické. Ke zpřístupnění digitálních dat uživatelům slouží speciální prohlížeče, jednoduše ovladatelné programové produkty, pomocí nichž se data z počítače efektně zobrazí na obrazovce počítače. Grafické výstupy pak představují především mapy, plány, výkresy, atlasy, tabulky, statistické přehledy, databázové reporty, realizované na plotru a tiskárně.

Nejčastěji vytvářeným typem výstupů v GIS jsou mapy, zejména **tématické mapy**, jež jsou zaměřeny na prostorové změny jednotlivého fenoménu nebo jednotlivého vztahu fenoménů. V rámci tématických map můžeme vyčlenit:

- a) choropletové (kartografické) mapy – typicky se využívají k vyjádření hodnot proměnných, které se vyskytují v rámci dopředu definovaných plošných jednotek např. počet dopravních nehod v jednotlivých regionech nějakého území.

b) proximální (izogradační mapy) – soustřeďují se na polohu a hodnotu mapovaného atributu na relativně homogenních plochách. Na rozdíl od předešlých se hranice identifikují na základě změn v datových hodnotách.

c) vrstevnicové nebo izoliniové mapy – používají se k reprezentaci kvantity (kvantitativního atributu) pomocí čar, spojujících body se stejnou hodnotou a zvýraznění gradientů mezi hodnotami. Nejčastěji se aplikují na znázornění povrchů a tlakových a teplotních poměrů v meteorologii.

Paletu vytvářených výstupů rozšiřují ještě specifické druhy map jako např. symbolické mapy, kartodiagramy či liniové (čárové) mapy (Tuček, 1998).

2.6. Prostorová analýza dat

2.6.1 Definice prostorových analýz

Většina informací, se kterými se setkáváme a které využíváme, má prostorový charakter. Jistým způsobem jsou vázány k určitému místu, jež reprezentují. Toto místo je třeba chápat v širším slova smyslu - může to být bod, sada bodů, linie, sada (kolekce, svazek) linií, areál. V této souvislosti pak hovoříme o geoinformacích. Formalizací informace získáváme data, obdobně formalizací geoinformace získáváme geodata (resp. prostorová data).

Prostorové analýzy představují kolekci technik, které vznikly v různých oborech a jejichž cílem byla analýza dat s důrazem na jejich prostorové vztahy. Významné postavení mezi těmito obory zaujímá statistika, ale řada postupů byla odvozena v geografii, geostatistice, ekonometrii, epidemiologii, v územním plánování a urbanizmu. Tyto postupy jsou používány v ještě širší škále aplikací včetně např. zdravotnictví a kriminalistiky. Prostorové analýzy můžeme definovat následovně:

Prostorové analýzy jsou souborem technik pro analýzu a modelování lokalizovaných objektů, kde výsledky analýz závisí na prostorovém uspořádání těchto objektů a jejich vlastností. Objektem pro tento účel rozumíme geografické objekty a jiné objekty s prostorovou lokalizací (např. hvězdy nebo útvary v obraze), at' již fyzické nebo abstraktní povahy, velmi často i události a jevy.

Vymezení pojmu prostorové analýzy nebylo dříve tak univerzálně chápáno a často se vztahovalo jen k určité oblasti aplikací či použitých postupů. Jako příklad můžeme uvést dvě starší definice:

- Unwin (1981): „Prostorové analýzy se zabývají uspořádáním prostorových dat na mapách (tedy bodů, linií, ploch, povrchů)“.
- Johnston, Gregory, Smith (1994): „Prostorové analýzy jsou kvantitativní (hlavně statistické) procedury a techniky aplikované v lokalizačních úlohách“.

Někteří autoři chápou termín „prostorové analýzy“ jako synonymum pojmu kvantitativní geografie, část z nich pak tento pojem uplatňuje jen pro tu část prostorových analýz, která využívá stochastické povahy jevů. Bez ohledu na konkrétní vyjádření je zjevné, že prostorové analýzy představují sadu analytických metod, vyžadujících přístup k atributům studovaných objektů i k informaci o jejich lokalizaci (Klufová, 2000).

2.6.2 Cíle prostorových analýz

Cíle prostorových analýz se opět značně liší podle oblasti aplikací a je obtížné nalézt univerzální rozdělení. Většina autorů se pouze omezuje na výčet cílů, které reprezentují oblast zájmu pro danou aplikaci. Při obecnějším vymezení cílů prostorových analýz můžeme rozlišit následující cíle:

1. Popis objektů resp. událostí ve sledovaném prostoru (včetně popisu uspořádání - tj. textury). Zahrnuje odvození statistických charakteristik pozorovaného uspořádání geoprvků (bodů, linií nebo areálů) a jejich srovnání; dále testování, zda je pozorovaná distribuce významně odlišná od určité hypotetické textury (což je významné pro následující interpretaci procesů); zkoušení prostorových vztahů a vazeb mezi entitami, ale i běžný popis vývoje pole např. výpočet hodnoty v neznámých místech (interpolace).

Zajímá nás, proč jsou určité fenomény více seskupeny v některých místech, zda to není jen vliv náhody, jak lze porovnat texturu v různých oblastech, jak lze takový rozdíl kvantifikovat, zda dochází ke změnám v čase. Někteří autoři kritizují tento cíl, protože většina analýz končí u takového popisu a už se nezabývá vysvětlením procesů, které vedly k pozorovanému uspořádání. Navíc málokdy v přírodě odpovídá vzorek teoretickému modelu. Zde však nečekáme, že situaci bude přesně vystihovat teoretická distribuce, popis nám ale slouží k nalezení klíčových faktorů, které vedou ke vzniku určitého uspořádání.

2. Výběr určitého místa na základě splnění jisté sady podmínek (či obecněji podle jistého rozhodovacího schématu) nebo zkoumání míry splnění daných podmínek v určitém místě nebo území.

3. Interpretace procesů, které vedly k pozorovanému stavu uspořádání objektů či událostí ve sledovaném prostoru (systematický průzkum), např. interpretace vzniku pozorovaného uspořádání bodů, vysvětlení vývoje území v čase (jak střední hodnoty tak variability).

4. Optimalizace uspořádání objektů / jevů ve sledovaném prostoru např. lokalizační a alokační úlohy a volba způsobu distribuce toků.

5. Zlepšení schopností předpovídat a kontrolovat objekty či události ve sledovaném prostoru (využití prediktivních modelů).

6. Redukce původního množství dat do menší, úspornější a přehlednější sady dat. Provádíme např. generalizaci původních dat pro lepší popis sledovaného jevu nebo jen za účelem snadnější manipulace.

Uvedený přehled cílů prostorových analýz jistě není a ani nemůže být úplný, protože s rozvojem geoinformačních technologií se nacházejí nové formy uplatnění prostorových analýz a tím i nové cíle jejich použití (Klufová, 2000).

2.6.3 Statistické prostorové analýzy

Statistické prostorové analýzy zahrnují metody založené na stochastické (náhodné) povaze uspořádání a vztahů. Nabízejí poměrně široké spektrum aplikací. Někdy jsou přímo označovány jako prostorová statistika. Podle počtu současně zkoumaných charakteristik můžeme používané statistické metody rozdělit na monovariační (jednorozměrné) a multivariační (analýzy vícerozměrných objektů/událostí). Používání alternativních názvů, uvedených v závorce, není příliš vhodné vzhledem k jejich snadné záměně za počet dimenzí u geometrické reprezentace objektů.

Monovariační statistické analýzy pracují současně pouze s jednou charakteristikou objektu, zatímco multivariační statistiky využívají více charakteristik současně. U multivariačních technik lze ještě specifikovat, zda studují více charakteristik současně jen vizuálně (a pak analýza a interpretace probíhá vizuálně a mentálně), nebo zda jde skutečně o multivariační metody.

Dělení podle povahy statistických technik využívá analogie z tradičního rozdělení statických technik na popisné (typicky výpočet statistických charak-

teristik) a indukční (na základě studia výběru usuzujeme na vlastnosti celku, provádíme testování hypotéz). Podle tohoto principu se dělí techniky statistické prostorové analýzy na popisné a inferenční.

popisné (centrografické techniky) – jedná se především o kvantitativní měření, charakteristiky polohy a charakteristiky rozptýlení.

inferenční (analýzy textury) – ty určují, zda distribuce je nebo není náhodná, popisují vztahy mezi 2 a více veličinami.

Pokud hovoříme o distribuci geoprvků, myslíme texturu (vzor), kterou vytváří geoprvky svým rozmístěním ve sledované části prostoru. Z hlediska textury geoprvků se rozlišují tři základní typy distribuce:

- shluková (clustered), event. skupinová
- pravidelná (regular)
- náhodná (random)

Podrobnější přehled základních technik pro provádění statistických prostorových analýz zahrnuje:

1. Jednoduché deskriptivní analýzy, transformace dat a sumarizace
2. Metody nejbližších vzdáleností (nearest neighbor) a K-funkce
3. Kvadrantové, jádrové a Bayesovské vyhlazovací metody
4. Prostorová autokorelace a kovariační struktury
5. Geostatistické a prostorové ekonometrické modelování
6. Prostorové generalizované lineární modelování
7. Multivariační techniky
8. Prostorové interakční modely

1. Jednoduché deskriptivní analýzy, transformace dat a sumarizace

Tyto operace často nejsou samostatně uvedeny, avšak tvoří základní nástroje pro řadu dalších technik. Zahrnujeme do nich jednoduché grafické a numerické metody sumarizace dat a manipulace s daty (histogramy, vyrovnání histogramu jádrovým odhadem, rankilové grafy, rozptylogramy, projekce multivariačních dat do 2D zobrazení, výpočet základních statistických ukazatelů, zjišťování korelace, transformace dat). Popis těchto technik lze najít v základních statistických učebnicích nebo publikacích popisujících průzkumovou analýzu dat. Zařazujeme zde i základní popisnou statistiku pro

prostorové objekty a jevy (např. určení středu pro shluk bodů, tvorba elipsy disperze).

2. Metody nejbližších vzdáleností (nearest neighbor) a K- funkce

Metody nejbližších vzdáleností a K-funkce jsou určeny pro posouzení umístění událostí či objektů (především bodová reprezentace) a určení typu pozorované textury (náhodná, nenáhodná). Metody nejbližších vzdáleností jsou založeny na grafickém srovnání pozorované distribuční funkce vzdáleností mezi objekty (nebo vzdáleností mezi náhodně umístěným bodem a pozorovaným objektem) s jinými pozorovanými distribučními funkcemi nebo s očekávanou (teoretickou) distribuční funkcí získanou z modelu vytvořeného z náhodných dat.

Funkce $K(h)$ je definována jako očekávaný počet dalších objektů do vzdálenosti h od určitého objektu. Provede se grafické nebo statistické srovnání naměřené K funkce s K funkcí odvozenou z teoretických modelů a posuzuje se typ pozorovaného vzorku.

3. Kvadrantové, jádrové a Bayesovské vyhlazovací metody

Kvadrantové, jádrové a Bayesovské vyhlazovací metody jsou založeny na neparametrických technikách a slouží k transformaci dat z diskrétní reprezentace do kontinuální, tedy k výpočtu hustoty událostí, k vyhlazení textury apod.

V prostorovém kontextu mohou být použity např. jako průzkumné metody pro identifikaci odlišných míst (tzv. hot spots, tedy více variabilních nebo naopak více homogenních míst), pro identifikaci vhodných modelů, pro analýzu shody modelů s naměřenými daty. Kvadrantové metody představují nejjednodušší způsob transformace dat z diskrétní reprezentace do kontinuální. V případě bodů se počítají výskyty bodů v jednotlivých buňkách překrývající mřížky.

Pokročilejší metody vycházejí z myšlenky jádrového (kernelového) vyhlazování. Vyhlazená hodnota v daném bodě je vypočtena jako vážený průměr z hodnot v okolních bodech, kde váhy jsou odvozeny z distribuce pravděpodobnosti se středem v příslušném bodě.

Jádrový odhad hustoty pracuje s lokalizačními daty a pak vyjadřuje prostorově vyhlazený odhad lokální intenzity výskytu objektů/událostí. Tuto

lokální vyhlazenou intenzitu je možné chápat i jako povrch rizika výskytu těchto objektů/událostí. Druhou možností je aplikace na atributová data a výpočet vyhlazeného odhadu sledovaných hodnot.

4. Prostorová autokorelace a kovariační struktury

Při geostatistické analýze se považuje rozložení hodnot modelované veličiny (např. obsah olova) za tzv. regionalizovanou proměnnou, která je vyjadřována jako funkce souřadnic X, Y, Z . V každém bodě jistým způsobem vymezeného prostoru nabývá určité hodnoty. Je evidentní, že v případě přírodních objektů je hodnota v daném místě výsledkem řady procesů, z nichž některé mají výrazně náhodný charakter. Výsledkem je značná prostorová variabilita hodnot a jejich nespojitost.

Geostatistická analýza se snaží popsat chování takovéto regionalizované proměnné. Jejím základním nástrojem jsou strukturální funkce. Prostorové kovariační struktury se zjišťují v atributových datech a popisují závislost mezi rozptylem (resp. korelací) hodnot a vzdáleností měření. Uvádějí, zda a jak souvisí umístěním blízké hodnoty jedna s druhou.

5. Geostatistické a prostorové ekonometrické modelování

Geostatistické modelování je založeno především na provádění lokálních odhadů s využitím výsledků strukturální analýzy (aplikace interpolačních procedur). Představují prostorové rozšíření standardních lineárních regresních modelů. Parametry jsou odhadovány pomocí funkce maximální věrohodnosti nebo zobecněnou metodou nejmenších čtverců.

6. Prostorové generalizované lineární modelování

Prostorové generalizované lineární modelování představuje zobecnění prostorových regresních modelů na případy, kdy modelovaná atributová data přísluší k výčtové doméně. Vycházejí z prostorového zobecnění myšlenek log-lineárního modelování kontingenčních tabulek a modelování poissonových nebo binomických proměnných.

7. Multivariační techniky

Většina multivariačních technik není speciálně orientována na prostorově závislá data, ale i tak mohou být velmi užitečná jako nástroj pro redukci dat a pro identifikaci významné kombinace proměnných. Metody se využívají např. v klasifikačních postupech při zpracování dat dálkového průzkumu Země.

8. Prostorové interakční modely

Prostorové interakční modely jsou založeny na modelování pozorovaného toku ze sady zdrojů do soustavy cílů. Pro soustavu zdrojů se definují požadavky, pro sadu cílů se popisuje atraktivnost. Měřítkem vzdálenosti může být i čas nebo náklady. Modely jsou běžně odvozovány ze zobecněných gravitačních modelů, zaměřených např. na minimalizaci procestovaných podmínek nebo na optimalizačních problémech (Klufová, 2000).

3. Metodika

3.1 Program CrimeStat

CrimeStat je prostorově statistický program umožňující analýzu lokalit trestných činů, navržený na konci 80.let Nedem Levinem a jeho kolegy ve státě Texas s podporou národního právního institutu ve Washingtonu. Jeho účelem je poskytnout paletu nástrojů pro prostorovou analýzu zločinů a ostatních místně určených incidentů. Jedná se o soubor nástrojů, navržený k poskytování statistických přehledů a modelů kriminalistických dat. Soubor vybavuje analytiku a výzkumníky z oblasti kriminality širokou škálou prostorově statistických metod, zastoupených ve spektru od jednoduchých, až po špičkové nejmodernější metody. Důvod je prostý a to ten, že uživatelé se ve svých potřebách a požadavcích liší. Program by tedy měl být přínosem pro rozličné organizace.

CrimeStat je samostatný, pod Windows běžící program, který lze snadno propojit s většinou nástrojů GIS. Ačkoliv je navržen zejména pro zpracování obrovského množství dat uchovávaného policejními odděleními, lze jej rovněž aplikovat na rozličné analýzy dalších polohově určených dat, jako jsou lokality zatčení, dopravní nehody či místa zásahu záchranné služby a hasičů.

Většina GIS programů, jako MapInfo, ArcView, ArcGIS, ARC/INFO, disponují velmi sofistikovanými databázovými operacemi, které občas dokážou při kriminální analýze poskytnout velkou službu. Nastávají však i situace, kdy je nezbytný kvantitativnější přístup. Pokud bude chtít například analytik prošetřit pouliční krádeže za určité časové období, nutně bude potřebovat nástroje k zobrazení jejich posunu. Při přidělování policejních hlídek do určité oblasti bude zase potřeba zjistit minimální dojezdovou vzdálenost, aby byla doba přesunu v případě zásahu co nejkratší.

Ač tedy některé z těchto a jim podobných analýz mohou být provedeny v některém GIS, kvantifikace v programu CrimeStat umožňuje přesnější identifikaci a schopnost porovnat různé typy incidentů, což činí tento statistický program důležitým.

3.2 Získání potřebných dat a jejich úprava

Naplnění databáze potřebným množstvím dat je jednou z nejdůležitějších činností každého projektu GIS. Bez dostatečného množství a struktury dat na vstupu nelze pomýšlet na kvalitní a relevantní údaje na výstupu. V případě mé

vlastní práce, tedy analýzy kriminálních incidentů, se jevila být podmínka kvalitních vstupů ještě klíčovější, neboť informace tohoto typu nejsou vůbec snadno dostupné. Při hledání zdroje potřebných informací jsem se nejdříve obrátil na český statistický úřad, ale povaha dat, uchovávaných touto institucí, neodpovídala potřebám mé práce. Statistické přehledy o počtu a druhu zločinů jsou jistě zajímavým počinem, avšak pro vlastní analýzu nedostačují. Pro analýzu bylo potřeba získat data s místním určením incidentů, tedy s určením polohy každého z nich v jistém souřadnicovém systému a navíc v digitalizované podobě.

S těmito neskromnými požadavky jsem se obrátil na tiskové oddělení českobudějovické policie a obeznámil tamější pracovníky se záměry mé práce. Jelikož policie v Českých Budějovicích pravidelně spolupracuje se studenty informatiky, nebylo poskytnutí malého souboru místně určených dat problémem.

Data však musela být případ po případu vyhledána ve spisech a spolu s místem incidentu vypsána, neboť hotové, místně určené databáze případů v digitální podobě, jak je znají kriminalističtí analytici stojící za zrodem programu CrimeStat, se u naší policie zatím nevytvářejí. Po vzájemné dohodě jsem tedy ve výsledku nakonec obdržel soubor dat zahrnující lokality 48 dopravních nehod z období mezi 22.4.-27.4. 2005 (včetně údajů o druhu zranění a přítomnosti alkoholu), 96 dopravních nehod z období mezi 18.8.-31.8 2005 a konečně 9 loupeží z období květen - červen 2005 a 11 loupeží z období říjen - listopad 2005 včetně informací o napáchaných škodách. Všechny incidenty se přihodily na území Českých Budějovic.

Získaná data bylo pro analýzu potřeba převést do digitální podoby, což jsem provedl přímo v programu ArcView, kde jsem editací digitální mapy Českých Budějovic přidával nová zobrazení v podobě bodů, představujících místa jednotlivých incidentů. Takto vytvořené body přijmuly souřadnicový systém mapy a získaly své geografické souřadnice X a Y, čímž byl soubor dat digitálně lokalizován a připraven k vložení do programu CrimeStat.

Vzhledem k faktu, že množství a struktura dat, která se mi podařilo pro analýzu získat, nebyla zejména kvůli nedokonalému systému sběru a uchování dat v České Republice dostatečná, rozhodl jsem se v některých případech pro větší názornost a pestrost použít k práci rovněž data 1063 pouličních krádeží za rok 1995 města Baltimore v USA, která byla součástí manuálu k programu.

3.3 Vkládání dat do programu CrimeStat

Vzhledem k tomu, že je program cílen pro širokou skupinu uživatelů, operující s daty v mnoha různých institucích, nevyžaduje ovládání programu nadstandardní znalost problematiky, a po krátkém seznámení se s ním v něm může vykonávat základní analytické operace téměř každý, což bezesporu představuje velkou výhodu.

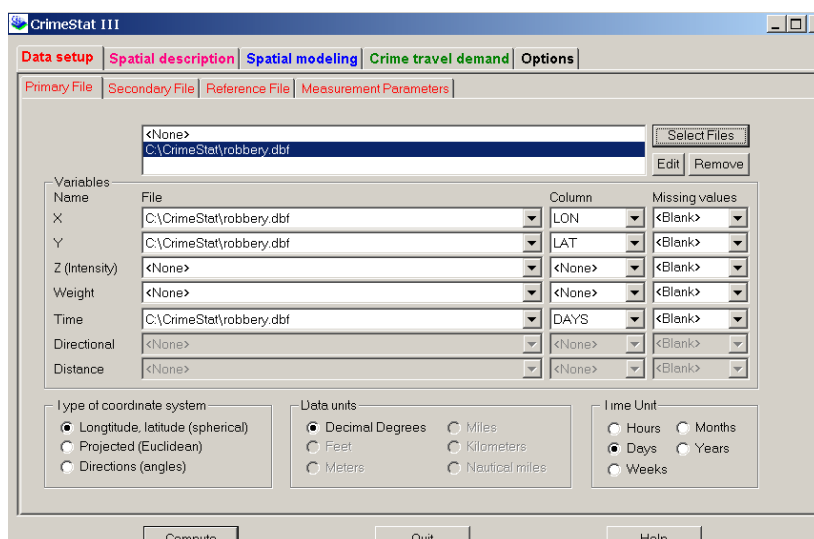
Vložení dat v požadovaném formátu, správné nastavení jejich parametrů a rovněž pak nastavení parametrů okolního prostředí, v němž bude analýza probíhat, je však klíčové a nezbytné k dosažení správných dat na výstupu. Chyby v takovémto nastavení program samozřejmě nerozpozná a na výstupu pak poskytne informace, které nemají s realitou nic společného. Proto je třeba věnovat vkládání dat maximální pozornost.

V programu CrimeStat provádí uživatel vkládání a nastavení dat kliknutím na položku *Data setup*, odhalující čtyři hlavní složky pro vkládání a upřesnění údajů k analýze. Nazývají se **primární složka**, **sekundární složka**, **referenční složka** a **parametry měření**.

Primární složka:

Tato složka je nutně požadována, neboť poskytuje souřadnice jednotlivých incidentů. Data jsou do ní vkládána v několika formátech – ASCII, databázovém formátu 'dbf', ArcView 'shp', MapInfo 'dat' a jiné, přičemž je důležité, aby měla data jako součást své struktury X a Y souřadnice. Uživatel kliknutím zvolí, jaký typ souboru bude vkládat a pak jej vyhledá a otevře, přičemž se mu název souboru promítne do celkem pěti řádků, patřícím různým proměnným. Ty jsou názorně zobrazeny na obrázku 1.

Obr. 1: Okno nastavení primární složky dat



První dva řádky patří souřadnicím **X** a **Y** sledované proměnné, tedy místům, kde k incidentům došlo. Tyto dvě hodnoty jsou bezpodmínečně nutné a pro jejich konečné nastavení je třeba vybrat v položce sloupec (column) odpovídající pole, ke kterému v databázi přísluší. Takže pokud se např. hodnoty souřadnice **X** nacházejí v databázi ve sloupci atribut s názvem longitude (zeměpisná délka), přiřadí uživatel v položce sloupec (column) k proměnné **X** název sloupce longitude, který ji reprezentuje.

Následuje proměnná **intenzita**. Intenzita je jednoduše hodnota, kterou je možno přidělit spolu s hodnotami **X** a **Y** ke každému incidentu a představuje tak další proměnnou. Pokud je tedy např. místem s **X** a **Y** souřadnicemi policejní stanice, intenzita jako další proměnná pak může vyjadřovat počet volání na tuto stanici za jeden měsíc. Já jsem proměnnou intenzity použil na případech loupeží, kde představovala výši škody v korunách českých.

Další proměnnou je **váha (weight)**. K vážení v tomto smyslu dochází, pokud se má dostat rozličným incidentům rozdílného statistického zacházení. Záchraná služba se např. může rozhodnout, z důvodu rozdílné hustoty osídlení a vzdáleností, přidělit oblastem v centru města váhu 1, okrajovým částem města váhu 2 a vesnicím přiléhajícím k městu váhu 3. Během mnoha operací v programu CrimeStat pak budou tyto hodnoty zahrnuty do kalkulace s tím, že incidenty, které se staly v oblasti s váhou 3 budou považovány za důležitější, než lokality a váhou 2. Váhy tak v tomto případě budou rozlišovat a kompenzovat rozdíly dojezdových vzdáleností k incidentům.

Kromě oblastního rozlišení je možné váhy použít i jiným způsobem, záleží pouze na potřebách uživatele. V mé práci jsem vážil dopravní nehody v rozmezí hodnot 1-3 z hlediska počtu a typu zúčastněných vozidel v kombinaci s vážností zranění a přítomností alkoholu (i když tyto případy byly vzácné). Takže např. zatímco nehodě osobního a nákladního automobilu či kolizi tří vozů či jakékoliv nehodě se zraněním nebo přítomností alkoholu. byla pro přirozenou závažnost přidělena váha 3, srážce osobního automobilu s pevnou překázkou bez vzniku zranění byla přidělena váha 1.

Poslední proměnnou je **čas**. Této hodnoty je využíváno zejména při složitých modelovacích metodách, sledujících chování pachatelů opakovaných trestných činů. Požadavky na podrobnost dat jsou v tomto případě enormní, proto jsem tuto proměnnou nikde do svých kalkulací nezahrnul.

Velkou pozornost je také třeba přikládat **souřadnicovému systému**, jež je součástí primární složky, a jeho jednotkám. CrimaStat uznává tři základní souřadnicové systémy:

- sférický (zeměpisná šířka a délka), který měří lokality pomocí úhlů z referenčních míst na Zemi.
- projektovaný (euklidovský), jehož souřadnice jsou libovolné souřadnice založené na specifické projekci Země na plochou rovinu. Souřadnice mají libovolný počátek(místo, kde $X=0$ a $Y=0$) a jsou téměř vždy definovány ve stopách či metrech.
- směrový (polární), kde jsou místa identifikována pomocí úhlů od libovolné referenční linie, obvykle na severu a hodnoty se pohybují mezi 0° - 360° .

Souřadnicový systém mojich dat je projektovaný (S-STJK), neboť v tomto systému jsem obdržel i digitální mapu Českých Budějovic, do které jsem incidenty zanášel. Pro správnost a soulad všech velikostních jednotek jsem nejdříve v ArcView nastavil jednotky mapy a vzdáleností na metry a v programu CrimeStat zatrhnul nejdříve projektovaný systém a poté měrné jednotky rovněž na metry.

Sekundární složka:

CrimeStat rovněž umožňuje vložení dalšího souboru dat a to právě do sekundární složky. V primární složce tak mohou být například místa odcizení motorových vozidel, zatímco v sekundární mohou být místa, kde byla vozidla nalezena. Podobně může primární složka obsahovat soubor dat s místy loupeží a sekundární místa policejních stanic.

CrimeStat dokáže se sekundární složkou zkonstruovat dva druhy studií. V první řadě dokáže spočítat vzdálenost od každého bodu primární složky ke každému bodu sekundární složky. Za druhé, CrimeStat umí využít jak primární, tak sekundární složku k odhadu trojrozměrné hustoty povrchu. Pokud tedy například obsahuje primární složka místa rezidenčních loupeží a sekundární složka populaci dle jednotlivých sčítacích bloků, nastavenou jako proměnnou intenzity, pak CrimeStat spočítá hustotu loupeží s ohledem na hustotu populace (riziko vloupání).

Jelikož je k použití sekundární složky zapotřebí velmi specifických dat, z nichž některá jsou součástí pouze amerického systému (sčítací bloky pro

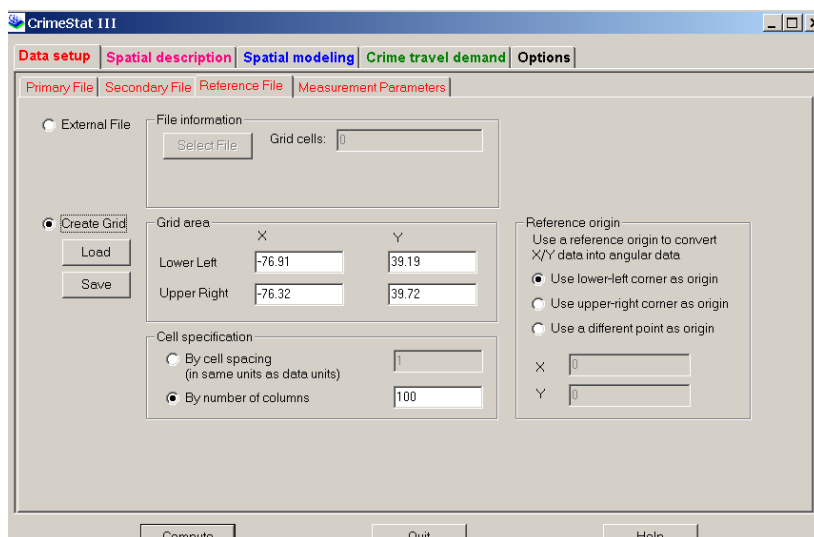
evidenci obyvatel), prezentoval jsem metody vyžadující sekundární složku na datech policie v Baltimore, jejichž součástí byla i databáze počtu obyvatel v jednotlivých okrscích města. V případě analýzy dopravních nehod a loupeží na území Českých Budějovic nebyla sekundární složka použita. Nastavení sekundární složky se provádí naprosto stejným způsobem jako u složky primární s tím rozdílem, že uživatel nezadá souřadnicový systém. Data sekundární složky musí být již ve stejném systému, jako data složky primární.

Referenční složka:

Některé metody v programu CrimeStat generalizují bodová data na všechna místa analyzované oblasti, k čemuž používají referenční soubor umístěný přes zkoumanou oblast. Jde o obdélníkovou mřížku s buňkami, definovanými pomocí sloupců a řádků. CrimeStat nabízí dvě základní možnosti jak si mřížku opatřit. Buď se mřížka načte do programu z externího souboru, anebo se přímo v programu vytvoří. Pro svá data jsem zvolil možnost druhou, tedy vytvoření mřížky. Jde o velmi jednoduchou proceduru, kdy se mřížka vytvoří zadáním souřadnic jejího spodního levého a horního pravého rohu. Tyto údaje jsem zjistil přímo v ArcView, kde jsem tahem myši ohraničil kolem katastru Českých Budějovic obdélník budoucí mřížky a poté vyčetl potřebné souřadnice, které jsem zanesl do referenčního souboru programu CrimeStat.

Konečná úprava mřížky se provede nastavením jejich buněk na požadovanou hustotu a to buď rozměrem buňky či počtem sloupců mřížky. Rozhodl jsem se pro nastavení počtu sloupců a upravil hodnotu z výchozích 100 sloupců na 250, aby byl rastr mřížky dostatečně jemný a poskytoval vizuálně kvalitnější výstupy. Takto vytvořenou mřížku si bylo možno přímo v programu následně uložit.

Obr. 2: Okno nastavení referenční složky dat



Parametry měření:

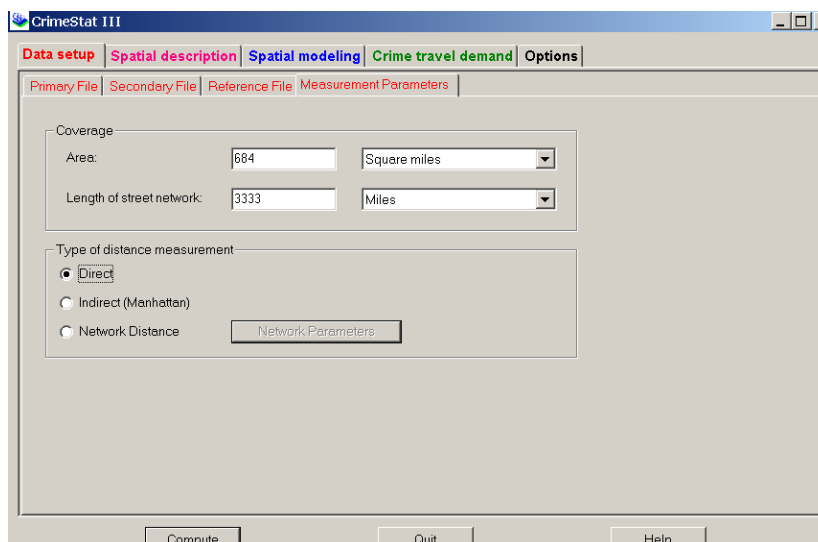
Posledními vlastnostmi završujícími definici dat jsou parametry měření. V této složce se definuje geografická oblast (rozloha), délka silniční sítě studované oblasti a určuje se, má-li být při měření použito vzdálenosti přímé nebo nepřímé. Pokud jde o první dvě zmíněné charakteristiky, je třeba dodat, že jejich určení je požadováno pouze u dvou metod, se kterými jsem nepracoval, tudíž jsem na jejich zjištění nelpěl. Co se týká používaných vzdáleností, měl jsem jako uživatel na výběr dvě.

Přímé vzdálenosti, představující nejkratší vzdálenost mezi dvěma body. Na klasické rovině, tedy v případě projektovaného souřadnicového systému, je nejkratší vzdáleností mezi dvěma body přímka. Avšak ve sférickém souřadnicovém systému je nejkratší vzdálenost oblouk. CrimeStat se proto na základě souřadnicového systému rozhodne, kterou z těchto dvou přímých vzdáleností bude počítat.

Nepřímá vzdálenost představuje přibližný odhad trasy na pravidelné síti ulic. V USA se pro ni vžil název Manhattanovská (bloková) metrika, podle pravidelné sítě ulic ostrova a i mnohá americká či jiná nepříliš stará města mají svým uspořádáním a strukturou ulic k tomuto vzoru blíže a užití tohoto typu vzdálenosti potom mnohem více odráží realitu. Vzdálenost je v tomto případě měřena jako suma horizontálních a vertikálních úseků při cestě mezi dvěma body. Pro své analýzy jsem však přeci jenom ponechal vzdálenost přímou, jež lépe reflektuje uspořádání Českých Budějovic a českých měst obecně.

Systémová (síťová) vzdálenost představuje cestovní vzdálenost v rámci konkrétní reálné silniční sítě vložené uživatelem. Měření je tak, na rozdíl od dvou předchozích případů, prováděno po reálně dostupných trasách, což přináší nespočet výhod a upřesňuje měření.

Obr. 3: Okno nastavení parametrů měření



4. Propojení GIS a programu CrimeStat

Tento oddíl mé práce bude věnován postupnému nastínění problematiky většiny statistických metod fungujících pod programem CrimeStat, demonstraci jejich použití na získaných datech a interpretaci výsledků provedených analýz. Analýzy jsou v programu rozčleněny do třech základních celků. Jde o skupiny deskriptivních metod, shlukových metod a prostorového modelování. Každý z celků bude náležitě v samostatných kapitolách představen. Grafické výstupy pro danou skupinu měření budou zařazeny vždy na závěr každé kapitoly.

4.1 Prostorová distribuce (rozmístění)

V této kapitole se budu zabývat prostorovým rozdělením a statistickými metodami, jež jsou využívány při popisu prostorových rozmístění případů. Ty budou nejprve náležitě vysvětleny a posléze ilustrovány na příkladech v programu CrimeStat. Nejzákladnějším typem ukazatelů pro prostorové rozdělení zločinů jsou centrografické statistiky. Patří sem: **průměr, střed minimální vzdálenosti, standardní odchylka vzdálenosti, elipsa odchylek a odchylka souřadnic X a Y.**

Průměr (mean center)

Je jedním z nejjednodušších ukazatelů, jedná se pouze o průměr souřadnic X a Y. V případě jedné proměnné je to bod, ve kterém se součet všech vzdáleností mezi ním a všemi ostatními body rovná nule. V případě dvou proměnných, jako je prostorové rozdělení incidentů, je bohužel situace o něco složitější. I tak může být průměr chápán jako bod, kde se suma všech vzdáleností mezi průměrem souřadnic X k ostatním hodnotám X rovná nule a současně se rovná nule suma všech vzdáleností mezi průměrem souřadnic Y a ostatními hodnotami Y.

Vzorec vypadá následovně:

$$\bar{X} = \sum_{i=1}^N \frac{X_i}{N} \quad \bar{Y} = \sum_{i=1}^N \frac{Y_i}{N}$$

Kde X_i a Y_i jsou souřadnice umístění jednotlivých incidentů a N je celkový počet incidentů. Grafickým výstupem průměru v Arcview je bod se svými souřadnicemi.

Vážený průměr (weighted mean center)

Váženého průměru je dosaženo vážením každé ze souřadnic X a Y další proměnou. Jeho výhodou je, že bod asociovaný v určitém prostředí může být ovlivněn vlastnostmi tohoto prostředí. Jeho užitečnost tedy spočívá v tom, že popisuje prostorové odlišnosti zkoumaného prostředí a s ním i faktory, které mohou mít vliv na rozmístění kriminality jako je např. hustota obyvatel různých městských částí, zda se jedná o rize rezidenční či komerční čtvrť nebo jaké sociální a etnické skupiny obyvatelstva v dané čtvrti žijí. V programu CrimeStat se vážený průměr počítá z atributů vah primární složky dat.

Vzorec je o váhu rozšířený průměr:

$$\bar{X} = \sum_{i=1}^N \frac{W_i X_i}{N} \quad \bar{Y} = \sum_{i=1}^N \frac{W_i Y_i}{N}$$

Střed minimální vzdálenosti (centre of minimum distance)

Další centrografickou statistikou je centrum minimální vzdálenosti. Co činí tuto metodu jedinečnou je fakt, že definuje bod, ve kterém je suma vzdáleností k ostatním bodům nejmenší. Stanovuje tedy místo, odkud je vzdálenost k ostatním definujícím místům nejkratší.

Ačkoliv program CrimeStat počítá pouze s přímou a nepřímou vzdáleností a čas dosažení bodů tedy nemůže být kalkulován, jeví se býti tato statistika velice užitečnou zejména v rozhodování umístění policejních hlídek. Postavíme-li např. policejní vůz do bodu centra minimální vzdálenosti, zajistíme tak nejkratší možnou dojezdovou dobu ke všem zásahům. V případě aplikace na dopravní nehody se tato statistika týká a může ji využít zejména záchranná služba a hasiči.

Vzorec vypadá následovně:

$$C = \sum_{i=1}^N d_{ie} \dots \min$$

Kde d_{ie} je vzdálenost k jednomu bodu i a C je centrum minimální vzdálenosti s X a Y souřadnicemi. Bohužel však neexistuje vzorec, který by tuto vzdálenost spočítal. Grafickým výstupem centra minimální vzdálenosti v Arcview je opět bod se svými X a Y souřadnicemi.

Standardní odchylka souřadnic X a Y (standard deviation of X and Y)

Kromě průměru a centra minimální vzdálenosti vykonává CrimeStat rozličná měření prostorového rozdělení prvků, jež popisují rozptýlení, orientaci a tvar rozdělení proměnné. Nejjednodušší z nich je právě standardní odchylka X, Y. Použité vzorce jsou standardní vzorce z většiny statistických učebnic:

$$S_x = SQRT \left[\sum_{i=1}^N \frac{(X_i - \bar{X})^2}{N-1} \right] \quad S_y = SQRT \left[\sum_{i=1}^N \frac{(Y_i - \bar{Y})^2}{N-1} \right]$$

Kde X_i a Y_i jsou souřadnice X a Y jednotlivých bodů, \bar{X} a \bar{Y} průměry souřadnic X a Y a N je celkový počet bodů. Standardní odchylka souřadnic X a Y naznačuje stupeň rozptýlení a jejím výstupem je obdélníkový polygon.

Standardní odchylka vzdálenosti (standard distance deviation)

V souvislosti se standardní odchylkou souřadnic X a Y, jež poskytuje informace o rozptylu případů, se vyskytují dva problémy. V první řadě neposkytuje souhrnnou statistiku rozptylu v místech jednotlivých incidentů a je vlastně dvěma oddělenými statistikami (rozptyl v X směru a rozptyl v Y směru). Za druhé provádí měření pouze v jednotkách souřadnicového systému. Prostředek, který toto překonává je standardní odchylka vzdálenosti. Ta je odchylkou vzdálenosti každého bodu od průměru vyjádřená ve velikostních jednotkách (stopách, metrech, kilometrech).
její vzorec je následující:

$$S_{XY} = SQRT \left[\sum_{i=1}^N \left\{ \frac{(d_{iMC})^2}{N-2} \right\} \right]$$

Kde d_{iMC} je vzdálenost mezi každým bodem i a průměrem a N je celkový počet bodů. Výstupem této veličiny v Arcview je kruhový polygon.

Standardní elipsa odchylek (standard deviational ellipse)

Standardní odchylka vzdálenosti je velmi dobrým příkladem rozptýlení případů okolo průměru (mean center). Avšak ve dvourozměrném prostoru jsou rozdělení často zkreslená, směřující v jednom či druhém směru a nevystihující realitu, což se jeví jako problém. Proto je zde další statistika zajišťující rozptyl ve

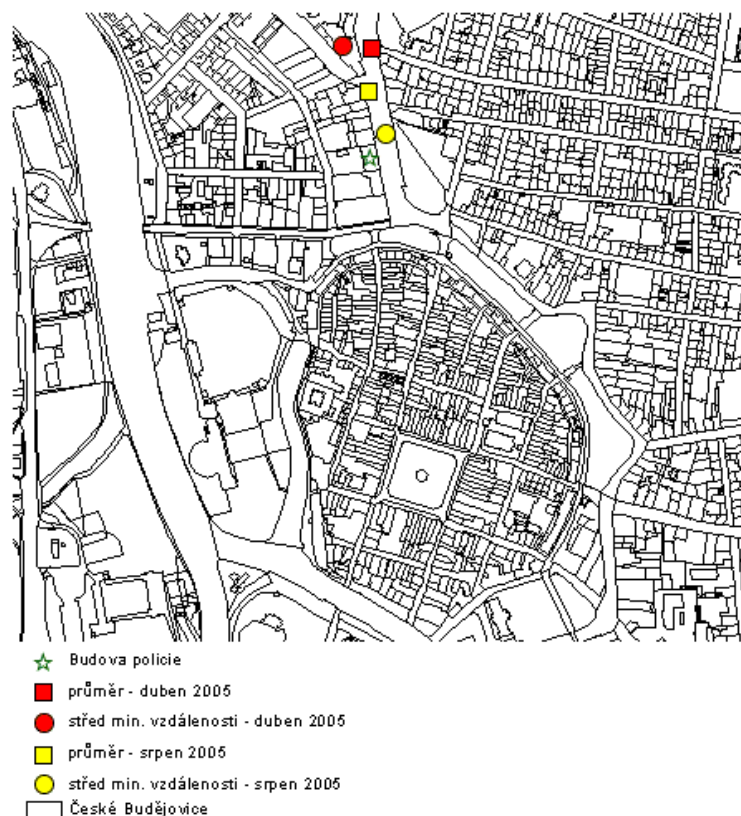
dvou rozměrech. A tou je právě elipsa odchylek. Její algoritmus je odvozen z bivariační funkce a to následujícím způsobem. Výstupem je polygon elipsy.

$$\text{Bivariační distribuce} = \text{SQRT} \frac{[\sigma_x^2 + \sigma_y^2]}{2}$$

4.1.1 Grafické výstupy prostorově popisných analýz

Na obrázku 4 je možné vidět prostorový popis dopravních nehod z dubna a srpna nejzákladnějšími charakteristikami, zmíněnými v úvodu této kapitoly, tj. průměrem a centrem minimální vzdálenosti. Z polohy obou párů bodů na mapě je patrná výhodná pozice policejní stanice v Českých Budějovicích, neboť jak průměry, tak centra minimálních vzdáleností nehod se nacházejí v naprosté blízkosti policie na Pražské třídě. Zejména centrum minimální vzdálenosti srpnových nehod je se sídlem policie téměř totožné. Toto vše tak dokazuje maximálně strategicky výhodnou polohu policie ve městě, umožňující co nejkratší možnou cestovní vzdálenost k dopravním nehodám.

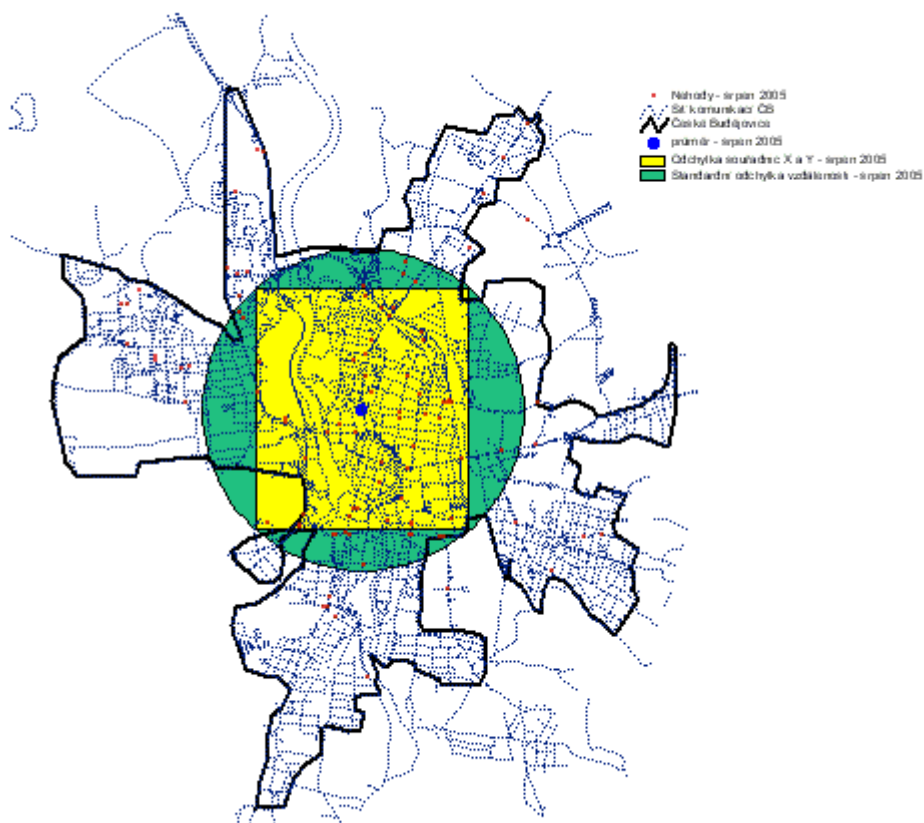
Obr. 4: Průměr a centrum minimální vzdálenosti dopravních nehod



Budeme-li uvažovat podobným způsobem v případě záchranné služby a nemocnice či umístění základny jednotek požární ochrany, dojdeme nutně k závěru, že by jejich pozice v blízkosti souřadnic těchto dvou charakteristik byla rovněž výhodná. Případy, kdy je dojezdová doba záchranné služby rozhodující,

však při nehodách ve městě nejsou z důvodu malé závažnosti zraněních tak časté, jako v případě nehod mimo město. Následující obrázky znázorňují, jak jsou v praxi analytiky využívány ostatní zmíněné centrografické charakteristiky. Jejich základním znakem je, že popisují rozdělení proměnné v dvojrozměrném prostoru, identifikují její rozptyl a orientaci.

Obr. 5: Standardní odchylna vzdálenosti a odchylna X a Y srpnových nehod



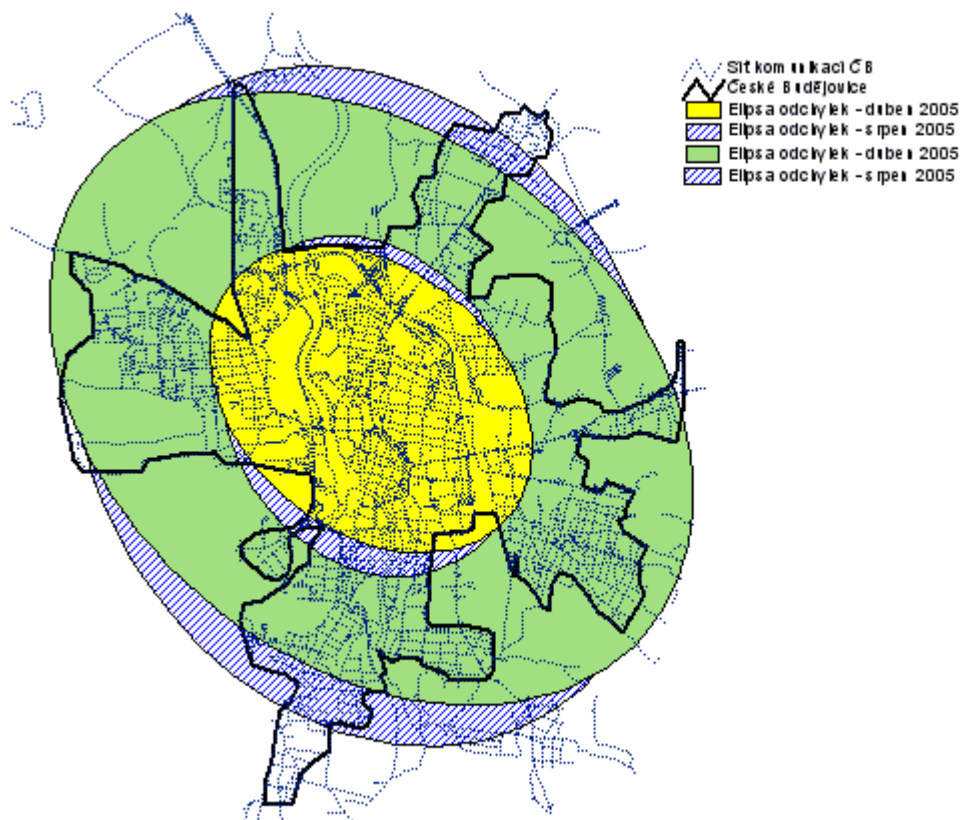
Obrázek 5 zobrazuje dvě nejjednodušší z metod popisu prostorového rozdělení incidentů a to standardní odchylnu vzdáleností a odchylnu souřadnic X a Y, znázorňující rozptyl okolo průměru. Z obrázku je patrné, že odchylna souřadnic, znázorněná obdélníkem, modeluje rozptyl nehod dosti nevěrohodně pouze ve dvou směrech (X a Y), díky čemuž se stává omezenou a nevystihuje důkladně realitu. Dle jejího rozdělení bychom se mohli domnívat, že se většina dopravních nehod stala v celkem nevelké oblasti čítající centrum města a jeho přidruženou severní část. Druhá metoda, standardní odchylna vzdáleností, modelující rozptyl incidentů pomocí kruhu, je již názornější a ve svém zobrazení

naznačuje i existenci nehod, ke kterým došlo v jižnější části města kolem Mánesovy ulice a dále ve Čtyřech Dvorech. Z jejího zobrazení lze též vyčíst, že jsou nehody orientovány obecně v severozápadní části města.

Na obrázku 6 je znázorněno prostorové rozdělení dubnových a srpnových nehod pomocí elips odchylek. Jejich nespornou výhodou je vystižení skutečné orientace souborů incidentů. Z obrázku je zřejmé, že rozdělení incidentů je v obou měsících téměř shodné, vykazující pouze v měsíci srpnu mírný odklon severovýchodním směrem, pravděpodobně kvůli většímu počtu nehod v severní či jižní části města. K žádné zásadní změně orientace lokalit nehod ale nedošlo a obě rozdělení sledují totožný trend rozptýlení incidentů na ploše města severozápadním směrem.

Tato skutečnost má své důvody v umístění nejdůležitějších městských komunikací a poloze největších křižovatek ve městě. Hlavní dopravní tepny bývají, co se hustoty dopravy týče, nejfrekventovanější a tato vlastnost se bez rozdílu vztahuje i na jejich křižovatky, kde přirozeně dochází v důsledku zvýšeného počtu vozidel k vyšší nehodovosti. A významné křižovatky se v Českých Budějovicích nacházejí právě na severozápadě, jihovýchodě a severu města. Jedná se o křižovatky ulic Mánesova – Lidická, Husova – Na Dlouhé Louce, Husova – Strakonická, Pražská – Strakonická, Pražská – Karoliny Světlé, Nádražní – Pekárenská, Nádražní – Rudolfovská atd.

Obr. 6: Standardní elipsy odchylek dopravních nehod (duben a srpen)



Na dvou ze zmíněných křižovatek (konkrétně Husova – Na Dlouhé louce a Pražská – K. Světlé) byl v měsíci září na zakázku magistrátu města zahájen testovací provoz velice zajímavého a výkonného měřicího zařízení schopného zaznamenávat rychlost projíždějících vozidel, jejich přestupky a vyfotografovat poznávací značky vozidel zákon porušujících řidičů. Ti, jak prokázaly tyto radary, předpisy příliš nedodržují. V jihočeské metropoli zaznamenali za necelý měsíc (od poloviny září 2005 do poloviny října 2005) 97 000 přestupků. Dopustil se ho každý desátý řidič. Reakce představitelů města České Budějovice a společnosti Czech Radar byly následující.

Juraj Thoma (ODS), náměstek primátora Českých Budějovic: Toto číslo je opravdu alarmující. V Českých Budějovicích policisté od začátku roku 2005 zaznamenali 2 500 dopravních nehod. Zahynulo při nich 5 osob a dalších 52 bylo těžce zraněno.

Milan Bednář, provozující společnost Czech Radar: Je to asi tím, že to je skutečně tranzitní město. Ta doprava je tam poměrně hustá. Ze všech zjištěných přestupků tvoří zhruba 20% jízda na červenou.

Juraj Thoma (ODS), náměstek primátora Českých Budějovic: Jsou zaznamenány i řidiči linkových autobusů, kteří projíždějí křižovatkami na červenou. Téměř 80 000 řidičů nedodrželo při průjezdu křižovatkou předepsanou rychlost 50 km/h. Nejrychlejší řidič tam jel asi 122 km/h.

Alena Toningerová, mluvčí okresního ředitelství Policie v Českých Budějovicích: Nejčastějšími příčinami je nedodržení bezpečné vzdálenosti, nedání přednosti v jízdě a samozřejmě i nepřiměřená rychlost.

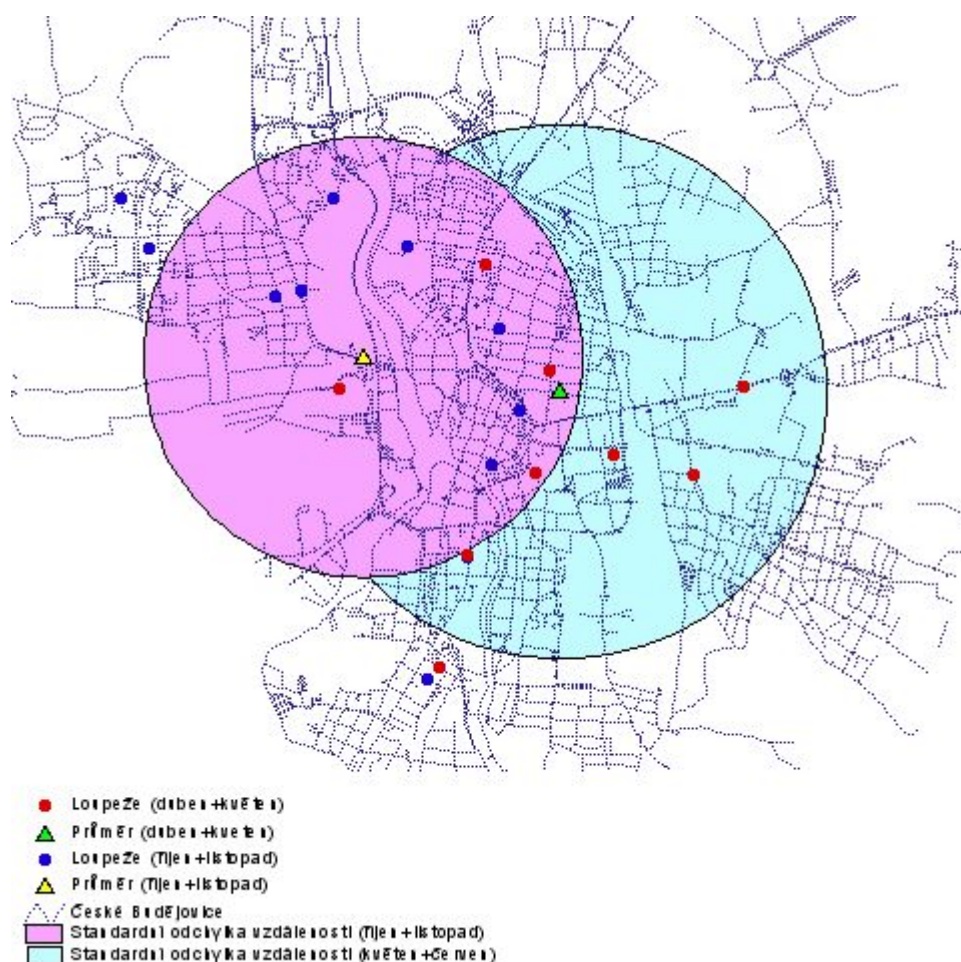
Juraj Thoma (ODS), náměstek primátora Českých Budějovic: Teď jde jenom o to, aby se ve zkušebním provozu osvědčilo to, zda bude příslušný přestupkový odbor schopen všechna ta čísla a ty údaje zpracovat.

Radary v tříměsíční zkušební době zatím na své náklady provozuje soukromá společnost. Pokud se zařízení osvědčí, město podle náměstka Tomy zřejmě prodlouží s firmou smlouvu. Radnice rovněž uvažuje o tom, že začne neukázněné řidiče pokutovat. Veškeré výdaje na provoz zařízení by pak mohly být pokryt vybrané pokuty (internetové stránky firmy Czech Radar, 17.10.2005).

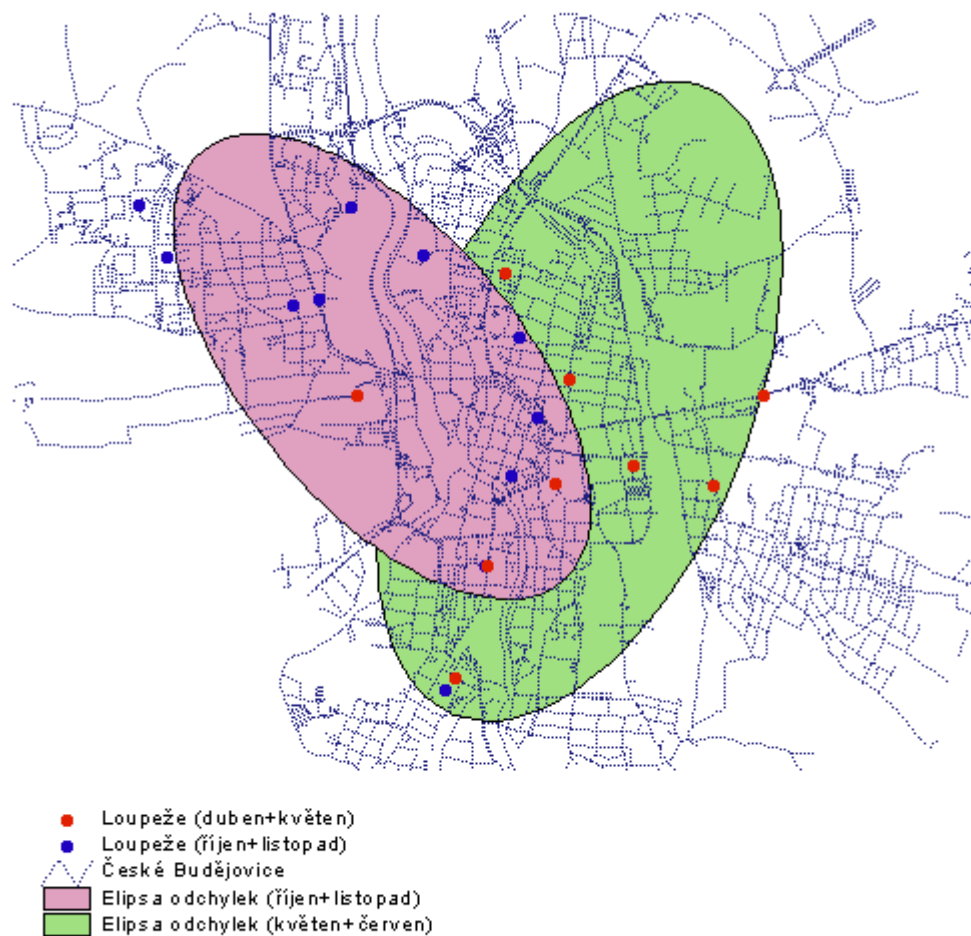
V době mé poslední návštěvy přestupkového odboru magistrátu města České Budějovice v lednu 2006 ještě město žádá data od firmy Czech Radar neobdrželo. V té samé době se rovněž mělo na zasedání vedení města rozhodnout o případném prodloužení smlouvy na provoz radarů.

Obrázek 7 zobrazuje prostorové rozdělení loupeží, ke kterým došlo na území Českých Budějovic ve dvou časových obdobích roku 2005. Popis rozptylu incidentů je v tomto případě proveden pomocí průměrů a standardních odchylek vzdáleností. Již z těchto dvou prostých charakteristik je zřejmý významný prostorový posun lokalit loupeží z východní části města více na západ, k čemuž jistě přispěl fakt, že se téměř polovina loupeží v období října a listopadu stala ve Čtyřech Dvorech, které byly v květnu a červnu kriminalitou tohoto typu téměř nedotčeny. Významnost posunu podtrhuje i vzdálenost mezi průměry za obě časová období. Pokud budeme chtít popsat tuto změnu ještě podrobněji, použijeme na výstupu k popisu souboru incidentů elipsy odchylek. Výsledek je zobrazen na obrázku 8.

Obr.7: Lokality, průměry a standardní odchylky vzdálenosti loupeží



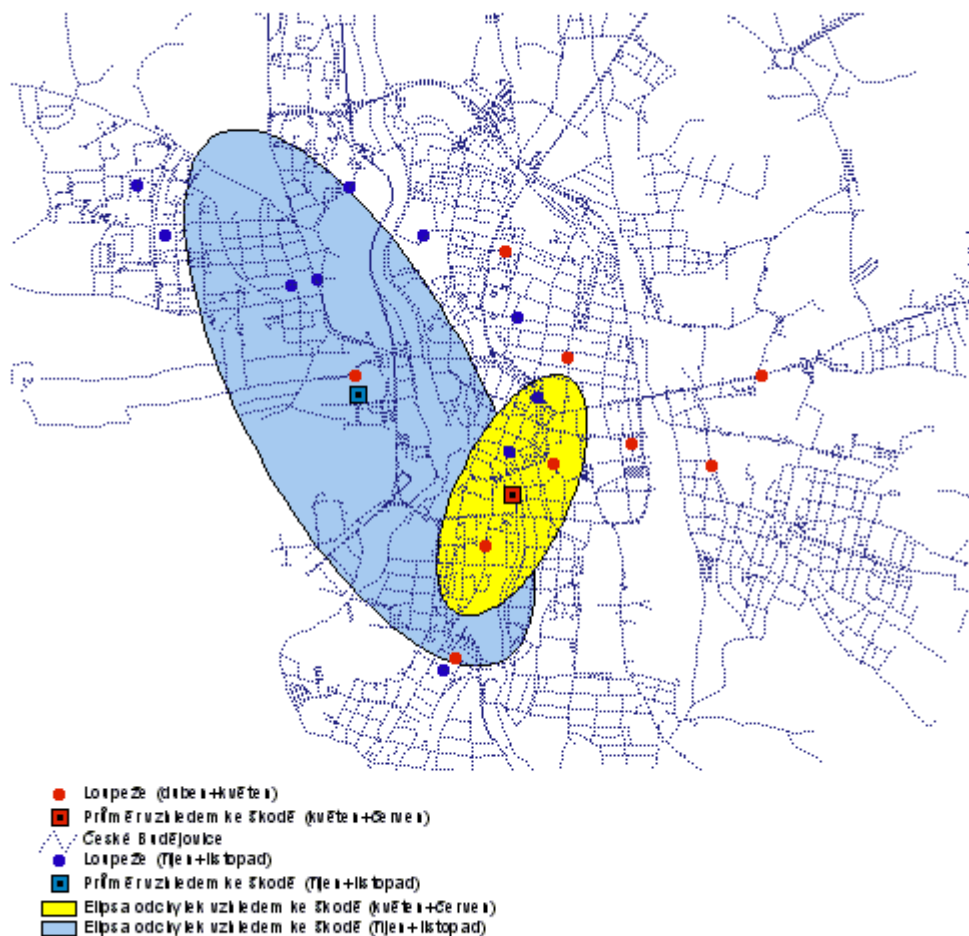
Obr. 8 : Standardní elipsy odchylek lokalit loupeží



Z tohoto obrázku je již jasný nejenom prostorový posun, který jsme mohli identifikovat pomocí předchozích charakteristik, ale hlavně naprosto rozdílná orientace souborů incidentů. Květnové a červnové loupeže se odehráli v relativní blízkosti centra s orientací na severovýchod, zejména kvůli loupežím v blízkosti Rudolfovské třídy a loupeži v obci Borek, která se do zobrazení nevešla.

V říjnu a listopadu byl naproti tomu zaznamenán enormní nárůst loupeží ve Čtyřech Dvorech a okolí, což přispělo k významné změně orientace rozmístění incidentů. Pro policii představuje takováto vyzorovaná změna jasnou indicii, kam zacílit své hlídkové vozy a kde zintenzivnit preventivní opatření. Obrázek 9 ukazuje tutéž situaci s tím rozdílem, že v tomto případě byla na soubor loupeží aplikována proměnná intenzity, představující výši vzniklé škody v korunách.

Obr. 9 : Elipsy odchylek lokalit loupeží s ohledem na vzniklou škodu v KČ



Již na první pohled je patrné k jak velkým změnám došlo a že se nejpodstatněji dotkly incidentů z května a června, jejichž rozdělení je reprezentováno elipsou minimální rozlohy, situované jižně od centra města. Orientace zůstala zachována, průměr se však výrazně posunul. Co tuto změnu vysvětluje, je fakt, že se mezi loupežemi z tohoto období nacházejí takové, které vykazují nulovou škodu (nic nebylo odzveno), takové, které vykazují škodu poměrně nízkou (řádově do tří tisíc korun) a dále ty, které způsobily škodu velmi vysokou. Tyto nejvýznamnější loupeže se staly pouze dvě a byly tudíž z hlediska intenzity škody přirozeně upřednostněny. Malá elipsa je tedy zobrazením lokalit loupeží vztažených na způsobenou škodu. V tomto případě je výsledkem loupeží v Multicasinu na Lidické třídě (škoda 56 000 Kč) a v baru Cotton v Otakarově ulici (škoda 15 300 Kč). U druhého souboru incidentů k tak zásadním změnám

nedošlo, neboť nejvážnějších loupeží bylo v tomto případě více a hlavně byly rovnoměrně lokalizovány od jižní části města (kancelář Tipsport v Šumavské ulici, škoda 32 300 Kč) až po severozápadní část (restaurace Verneovka, N.Frýda, škoda 17 300 Kč; prodejna PC Husova, škoda 37 102 Kč). Zajímavostí je, že Multicasino v Lidické ulici bylo v listopadu vyloupeno podruhé a to bez jednoho dne přesně půl roku po první loupeži, při které bylo odcizeno 56 000 Kč.

V listopadu se však pachatelům již tak velkou částku odcizit nepodařilo a způsobili škodu pouze za 3 200 Kč. Lze se jen domnívat, zdali to bylo způsobeno kvalitnějším zabezpečením podniku, neskušeností pachatele či efektivním zásahem policie. Podnik se tímto jistě zařadí mezi místa, na která bude policie v pozdních nočních hodinách zaměřovat větší pozornost.

Z výše provedených analýz je zřejmé, že pokud disponujeme a popisnou statistikou zpracujeme data rozličných kriminálních případů, můžeme na základě těchto popisných analýz usuzovat, jak se prostorově mění, sledovat změny orientace jejich rozptylu a z těchto analýz si dotvořit informace o trendech v pohybu pachatelů či výběru oblíbených lokalit trestné činnosti nebo o tom, jsou-li série loupeží spíše dílem jednoho nebo dvou různých pachatelů. Dle vysledovaných změn orientace činů za dvě odlišná období může rovněž policie zefektivnit směr cesty svých hlídkových vozů a zaměřit je na správné oblasti.

4.2 Analýza frekventovaných míst (Hot Spots)

Hot spots, neboli frekventovaná, „horká místa“, představují místa limitovaného geografického určení, kde se kriminalita či jakýkoli jev po čase neustále znovu vyskytuje. Policie již ze zkušeností ví, že existují místa, která k páchání trestné činnosti přitahují více jako např. chudší odlehlé čtvrtě, sídliště, bary, diskotéky apod. Ať jsou důvody pro koncentraci zločinů v těchto místech jakékoliv, jsou skutečné a jsou v povědomí většiny policejních oddělení. Koncept je velice užitečný. Policisté hlídkující v určitém obvodu tak mohou zaměřit svou pozornost na specifická místa, u kterých vědí, že se zde trestná činnost postupně znovu objeví a jednotky prevence kriminality mohou zacílit své úsilí s vědomím, že dosáhnou kladného efektu s použitím minimálních finančních zdrojů.

Jak již bylo zmíněno, představují tyto techniky velmi účinný nástroj v prevenci kriminality, avšak jejich měření je také bohužel velkým problémem. Existují doslova tucty rozličných statistických technik navržených k identifikaci těchto míst. Všechny techniky jsou závislé na optimalizaci statistických kritérií, ale liší se navzájem ve své metodologii a stejně tak v kritériích použitých k identifikaci frekventovaných míst. Rozdělení metod proto provádím do těchto pěti základních skupin:

1. Metody dle polohy bodu – CrimeStat disponuje dvěma a to *Mode* a *Fuzzy Mode*. Obě metody budou postupně probrány a demonstrovány.
2. Hierarchické techniky – jedná se o metody hierarchicky shlukující jednotlivá místa výskytu zločinů, od primárních shluků, obsahujících pouze několik případů, až po shluky vyšších řádů. V programu CrimeStat bude demonstrována metoda *nearest neighbor hierarchical clustering*.
3. Dělicí techniky – tyto techniky dělí incidenty do specifického počtu seskupení, obvykle definovaných uživatelem. Každý bod je tedy přiřazen k jedné jediné skupině.
4. Techniky hustoty – tyto techniky identifikují shluky vyhledáním míst s hustou koncentrací incidentů. CrimeStat má dva typy jádrových metod k hledání hustoty. Demonstrovány budou *single kernel density* a *dual kernel density*.
5. Techniky uplatňující riziko – tyto techniky identifikují shluky ve vztahu s další proměnou jako např. populace, zaměstnanost atd. Demonstrována bude metoda *risk-adjusted nearest neighbor hierarchical clustering*.

V této kapitole popíši a na příkladu dat názorně předvedu několik nástrojů k identifikaci shluků zločinných incidentů. Dvě metody k hledání jádrové hustoty budou popsány a demonstrovány v další kapitole.

4.2.1 Mode a Fuzzy Mode

Mode

Je jednou z nejprostších analýz typu „hot spots“ a představuje prostorovou analogii klasického statistického modu. Pracuje tak, že zaznamenává místa s největším počtem kriminálních případů, tedy jejich frekvence (četnosti) v každém jedinečném místě (bodě s X a Y souřadnicemi), vytvoří seznam a vyhodnotí výsledky v tabulce od nejfrekventovanějších míst po ta nejméně frekventovaná.

Výstup je v databázovém formátu „dbf“ a obsahuje čtyři údaje.

1. celkovou pozici
2. frekvenci incidentů
3. X souřadnici místa
4. Y souřadnici místa

Fuzzy mode

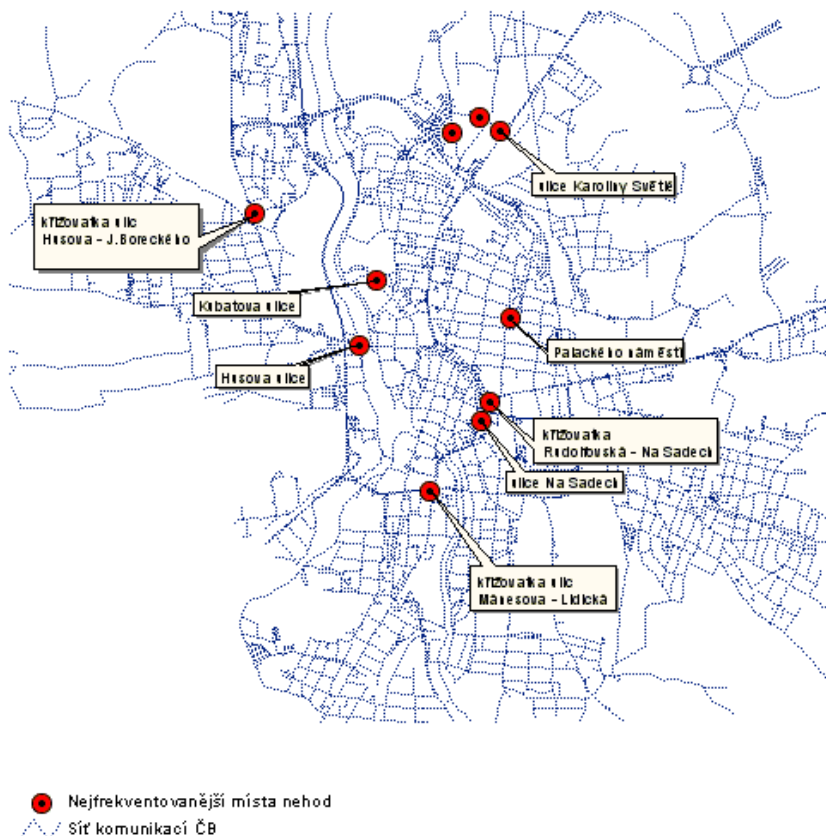
Užitečnost předchozí techniky je však závislá na stupni rozlišení geografického původu místa incidentu. U mnohých případů se stává, že jim je přiřazeno pouze jedno místo, jediný geografický statut, který čin identifikuje, který však úplně přesně neodráží realitu. Dojdeli například ke krádeži vozu na parkovišti, je velmi často tato krádež přiřazena k adrese domu majitele ukradeného vozu, jenž se může nacházet i o několik bloků dál. Z této informace by pak policie mohla nabýt dojmu, že se v dané oblasti vyskytuje mnoho frekventovaných míst, i když ve skutečnosti se jedná jen o jedno a tím je společné parkoviště, sdílené mnoha lidmi s různými adresami.

Proto CrimeStat přichází s další metodou z řady bodových lokací nazvanou **Fuzzy mode**. Ta umožňuje uživateli nastavit malý hledací radius okolo každého místa tak, že proces zahrne v potaz všechny události, které se staly jak v místě určení, tak i v blízkém okolí vymezeném kružnicí daného poloměru. Nastaví-li tedy uživatel hledací radius 100m, bude mu ve výstupu zobrazen počet činů v požadovaném místě a okruhu 100m kolem tohoto místa. Cílem metody je

tedy dovolit měření v místech s možným vysokým výskytem trestných činů a vysokou nepřesností při měření. Také je si třeba ale povšimnout, že nastavením hledacího radiusu jsou některé případy započítávány několikrát, pokaždé pro ten radius, pod který spadají. Ty se samozřejmě překrývají. Proto je třeba být při nastavování opatrný a volit spíše menší hodnoty. Jedině tak může tato metoda poskytnout lepší výsledky v identifikování nejčtetnějších míst, než předchozí metoda. Pokud jde o výstupy, poskytuje tato metoda rovněž databázový formát a stejný druh výstupů jako zmíněný modus.

Pomocí této metody s hledacím radiusem 200m byla provedena analýza 48 lokalit dubnových nehod, která je znázorněna na obrázku 10. Z celkového počtu 45 míst sestavených programem jsem vybral a dle souřadnic do mapy zanesl prvních deset, z čehož první tři lokality čítaly tři nehody a dalších osm nehody dvě. Mezi první tři nejčtetnější místa dubnových nehod patří ulice Karoliny Světlé, zejména pak její křížení s Pražskou třídou, dále ulice Na Sadech v úseku od Rudolfové ulice k Žižkově ulici a křižovatka ulic Mánesova – Lidická.

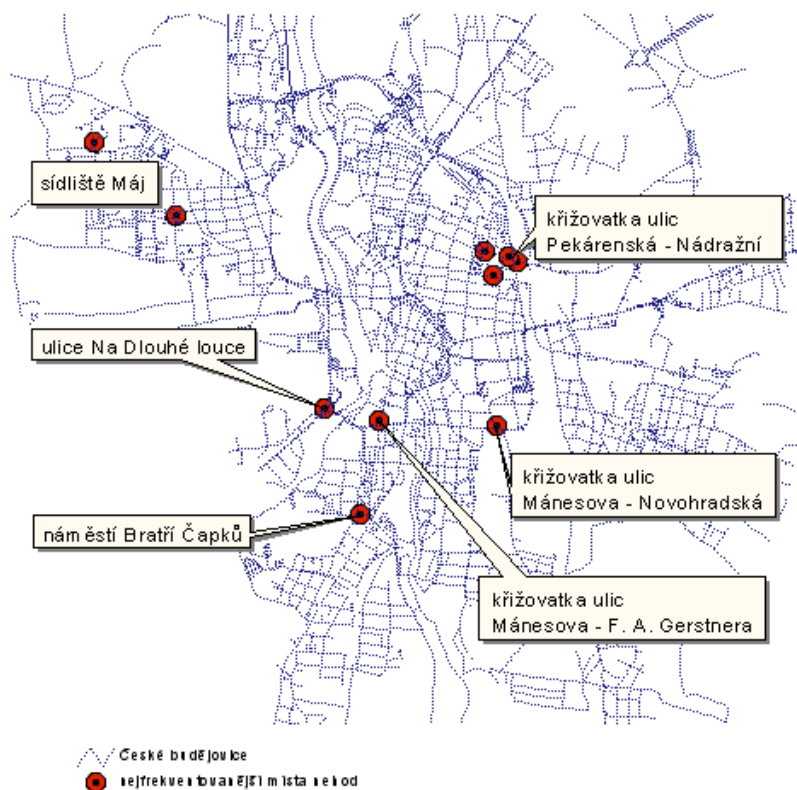
Obr. 10: deset nejfrekventovanějších míst dubnových dopravních nehod



Na obrázku 11 je vidět výsledek aplikace metody fuzzy mode na 96 srpnových dopravních nehod při nastavení hledacího radiusu 100m (z důvodu většího množství případů byl zmenšen). Program po kalkulaci opět vyhodnotil 45 nejfrekventovanějších lokalit, z nichž jsem vybral a dle souřadnic do mapy zanesl prvních deset. První čtyři nejčetnější místa, vykazující po čtyřech incidentech, se nacházejí v okolí křižovatky ulic Pekárenská a Nádražní, kde byla koncentrace nehod opravdu vysoká a s největší jistotou jsou zde některé případy zdvojeny kvůli překrývání hledacích radiusů. Následují místa s dvěma incidenty, jako náměstí Bratří Čapků, ulice Na Dlouhé louce, dvě lokality na Mánesově ulici a sídlišti Máj.

Metoda modus nenašla v žádném ze dvou souborů dopravních nehod jediné stejné místo, kde by došlo alespoň ke dvěma či více incidentům, a tudíž ve výsledku nabídla pouze seznamy 45 míst s frekvencí po jednom incidentu.

Obr. 11: deset nejfrekventovanějších míst srpnových dopravních nehod



4.2.2 Hierarchické shlukování metodou nejbližšího souseda (nearest neighbor hierarchical clustering)

Tato metoda identifikuje v programu CrimeStat skupiny incidentů, jež jsou si prostorově blízké, na základě kritérií je shlukuje dohromady a pokračuje tvořením skupin z již vytvořených shluků. Shlukujícím kritériem je v tomto případě metoda nejbližšího souseda. Ta definuje tzv. prahovou vzdálenost, kterou porovnává se všemi vzdálenostmi všech párů bodů. Do shluků jsou pak vybrány pouze ty body, jejichž vzdálenost k jednomu či více bodům je kratší než daná prahová vzdálenost. Uživatel může navíc specifikovat minimální počet bodů, jež mají být zahrnuty v každém shluku.

Pak pouze body, které splňují obě kritéria – jsou blíže, než prahová vzdálenost a patří do skupiny s minimálně tolika body, kolik bylo nastaveno uživatelem, jsou zařazeny do shluku prvního řádu (first-order cluster). Rutinně je takto posléze dopočítáno další shlukování produkující hierarchii. Body z prvního řádu jsou shlukovány do druhých řádů, ty do shluků třetího řádu atd. až do doby, kdy již další shlukování není možné. Uživatel rozhoduje před užitím metody o následujících parametrech, které pak mají vliv na vzhled výstupu.

Pravděpodobnost (nebo p-hodnota) náhodného výběru bodu do shluku (ona prahová vzdálenost). P-hodnota se nastavuje v programu posunutím měřítka po stupnici v rozmezí hodnot od 0.00001 ($p=0.001\%$) do 0.999 ($p=99.9\%$). Čím menší pravděpodobnost zvolíme, tím menší bude prahová vzdálenost. Menší prahové vzdálenosti produkují následně více shluků malých velikostí a naopak. Využití tohoto nastavení spočívá v možnosti uzpůsobení výstupu velikosti zkoumané oblasti.

Pro zkoumání konkrétních lokalit, jež nepředstavují velkou geografickou oblast, jako např. ulice či obchodní centrum, se jeví odpovídající zvolit malou prahovou vzdálenost, jež vyprodukuje přímo mikroregiony kriminality. Na druhou stranu např. pro hlídkující policejní vůz je taková oblast příliš malá a je lepší se přiklonit k větším vzdálenostem, aby výsledný shluk incidentů zasahoval alespoň přes několik bloků či dokonce čtvrtí.

Minimální počet incidentů zahrnutých do shluku. Tímto kritériem nastaví uživatel minimální počet bodů, který musí shluk obsahovat a zabrání tím tomu, aby program generoval shluky o dvou či třech incidentech, které by měli nulovou vypovídací hodnotu. Je na uvážení uživatele a hlavně na velikosti souboru dat, kterým disponuje, jakou hodnotu si zvolí. Programem automaticky

nastaveno je minimum deset bodů ve shluku. Ještě je třeba dodat, že ze dvou výše zmíněných kritérií má na generované množství shluků a jejich velikost větší vliv právě specifikace minimálního počtu bodů.

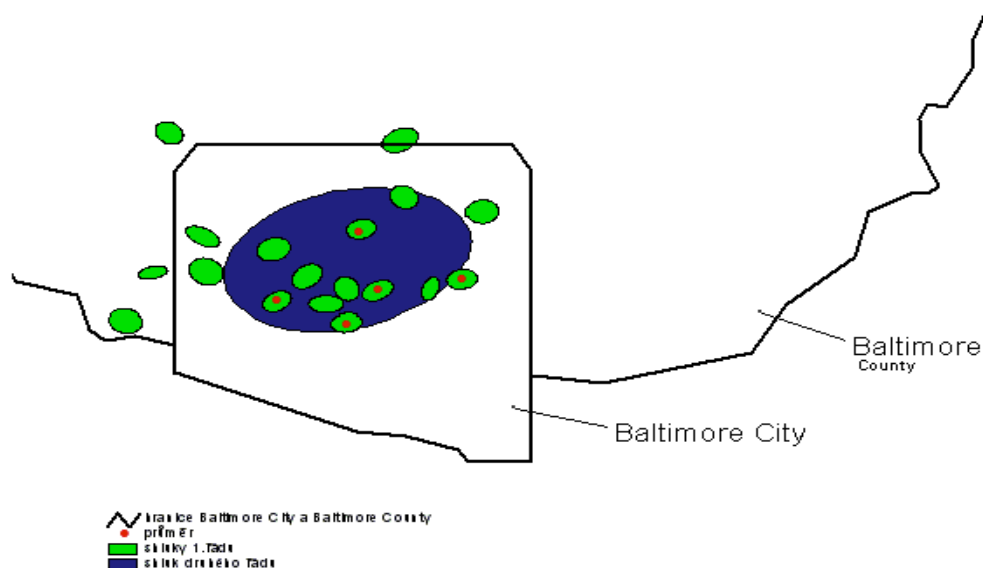
Metoda poskytuje celkem šest druhů výstupů:

1. Pro každý určený shluk jeho hierarchické pořadí a číslo shluku
2. Pro každý shluk spočítá program centografický průměr X a Y souřadnic
3. Počet incidentů v každém shluku
4. Hustota shluku (počet incidentů dělený plochou)
5. Elipsy jednotlivých shluků a jejich velikost
6. Plochy jednotlivých elips

Nejdříve si metodu předvedeme na příkladu 1061 krádeží ve městě Baltimore. Parametry, které jsem zadal, jsou následující. Pravděpodobnost výběru jsem zvolil 75%, abych zabránil vytvoření zbytečně vysokého počtu shluků malé velikosti a spíše získal větší regiony s více incidenty. Minimální počet incidentů ve shluku jsem ponechal 10. Program následně vygeneroval 19 nejčetnějších shluků incidentů, z toho 18 z nich prvního řádu a 1 shluk druhého řádu. Jejich poloha, tvar a umístění jsou reprezentovány elipsami na obrázku 12.

Počet krádeží v každé frekventované oblasti, znázorněné elipsou, leží v rozpětí 10-22 případů, rozloha jejich území pak mezi 390 až 870 čtverečními metry. Shluk druhého řádu, poskytující obecnější informace a využívající se pro rozlehlá území, má rozlohu téměř 40 čtverečních kilometrů. Prvních pět lokalit prvního řádu s největším počtem činů, tj. 19-22, jsem pomocí souřadnic průměru jejich elips označil. I když se shluky prvního řádu používají hlavně k identifikaci menších oblastí, tak i v tomto případě poskytují dobrý přehled o koncentraci krádeží v Baltimore, které jsou soustředěny ve středu a severní části města. V komerčním centru na jihu města (downtown) se vyšší koncentrace nevyskytují.

Obr. 12: Použití metody hierarchických shluků k určení frekventovaných míst krádeží ve městě Baltimore



Jistě to souvisí s faktem, že se tato centra amerických měst po skončení pracovní doby téměř vylidňují. Z hlediska praktického uplatnění potvrzují tyto domněnky do jisté míry procesy, kterým byla vystavena západní města v daném období. Jednalo se o restrukturalizaci měst v souvislosti s měnícími se výrobními procesy a z ní vyplývajících dopadů na využití jednotlivých zón města a jeho obyvatele, což bylo podrobena četným výzkumům a existuje celá řada teorií, použitých k vysvětlení daných procesů. Právě během 80. let minulého století se v USA zvýšil dramaticky počet bezdomovců a tím také i kriminalita. Toto zvýšení souviselo se znovuoživováním a gentrifikací* amerických měst, kdy byly v centrech měst bourány domy obyvatel s nízkými příjmy, aby tak udělaly místo pro nové mrakodrapy (Sýkora, 1993).

S otázkou populace souvisí i nadstavba metody hierarchického shlukování s uplatněním tzv. faktoru rizika. Jde o to, že některé shluky, generované

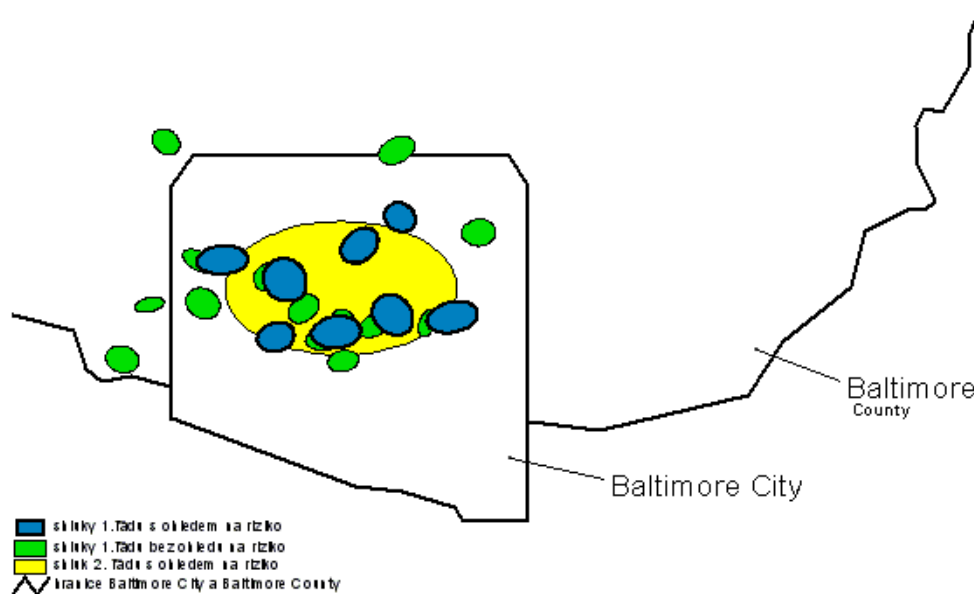
* gentrifikace je proces, při němž dochází k rehabilitaci obytného prostředí některých čtvrtí v centrálních částech velkých měst a k postupnému vytlačování a nahrazování původního obyvatelstva příjmově silnějšími vrstvami nově příchozích „gentrifierů“ (Sýkora, 1993)

demonstrovanou metodou, jsou někdy více či méně výsledkem vysoké koncentrace osob v blízkém okolí. Populace není totiž nikdy rozmístěna rovnoměrně, je spíše orientována do populačních center, kde je jednoduše vždy větší pravděpodobnost páchaní trestné činnosti. Proto je koncentrace kriminality ve vztahu k hustotě obyvatel zajímavým tématem a pro mnoho účelů nabývá prvořadé důležitosti především riziko zločinu, spíše než jeho množství. Pokud je tedy cílem odhadnout, kde se nacházejí nejrizikovější shluky, není metoda hierarchických shluků dle nejbližšího souseda odpovídajícím nástrojem.

Princip nadstavby metody pracující s rizikem je jednoduchý. Místo, aby program počítal s nastavenou konstantní prahovou vzdáleností, neustále ji mění a přizpůsobuje s ohledem na počet obyvatel jednotlivých sčítacích bloků obsažený a nastavený jako proměnná intenzity v sekundární složce dat. To vše se děje pomocí algoritmu interpolace (ten je uveden v další kapitole).

Obrázek 13 tedy jasně demonstruje rozdíl mezi frekventovanými a rizikovými místy. Frekventovaná místa představují pouze objem již uskutečněného zločinu, riziková místa pak potenciální riziko zločinu v lokalitě s ohledem na množství obyvatel jednotlivých městských částí. Rizikových oblastí je mnohem méně, jsou však o něco rozsáhlejší a koncentrují velké množství případů (23-38). Čtyři oblasti z obou analýz se překrývají, tzn. že ve čtyřech částech města odpovídá množství krádeží rizikovosti těchto míst.

Obr. 13: Použití metody hierarchických shluků k určení rizikových lokalit loupeží s ohledem na hustotu populace města Baltimore



Analýzu tohoto typu by bylo jistě zajímavé provést v námi známém prostředí Českých Budějovic a docílili bychom bezpochyby zajímavých zjištění. K provedení je však zapotřebí soubor dat sekundární proměnné, ke které bychom mohli riziko vztáhnout. V našem systému sběru a uchování dat, jenž zatím za vyspělými zeměmi výrazně pokulhává, však něco takového zatím není možné. Na obrázku 14 je vyobrazena aplikace metody hierarchických shluků na srpnové dopravní nehody v Českých Budějovicích. Jediné nastavení parametrů, které generuje shluky je maximální pravděpodobnost výběru a naprosto minimální počet bodů ve shluku (nastavil jsem pouze 2), čímž ztrácí analýza vypovídací hodnotu. Při jakémkoliv střídmejším nastavení mi program oznámil, že nenašel žádné shluky.

Tento fakt jen podtrhuje mou předchozí domněnku o nedostatečném souboru dat (v analýze pro Baltimore jsem pracoval s 1061 případy). Přestože je víceméně tato metoda na data dopravních nehod neaplikovatelná, i tak je však možno z obrázku odhalit frekventovaná místa, jako jsou křižovatky ulic Pekárenská – Nádražní či Mánesova – Čechova , čítající tři nehody a další dvě místa čítající nehody dvě. To je však bohužel vše, k čemu může momentálně sloužit.

Obr.14: Nejčastější místa dopravních nehod použitím metody hierarchických shluků dle nejbližšího souseda



4.2.3 Prostorová časová analýza kriminality (Spatial and Temporal Analysis of Crime - STAC)

Jako poslední metodu ze skupiny „hot spots“ jsem si vybral metodu prostorové časové analýzy kriminality, jež byla vyvinuta jako jedna z prvních metod k identifikaci frekventovaných míst v roce 1989 kriminalisty ve státě Illinois, USA. Ti posléze souhlasili s integrací STAC do programu CrimeStat. Metoda STAC v programu CrimeStat vyhledává a určuje nejhustší shluky incidentů na základě reálného rozložení bodů na mapě, tvoří z bodových dat prostorové prvky a identifikuje tak největší koncentraci bodů pro dané rozdělení. Každá oblast s určitou hustotou je následně reprezentována vhodnou elipsou, jež jsou snadno zobrazitelné v některém software GIS.

STAC není omezen žádnými vytyčenými hranicemi jako jsou policejní obvody nebo územím sčítání lidu, což je velice důležité, vezmeme-li v úvahu, že místa kriminálních případů (místa obchodu s drogami, násilím vyhlášené hospody, teritoria gangů nebo graffiti) nikdy nutně nekončí na hranici policejního obvodu, ale spíše se prolínají.

Rovněž se při prohlížení mapy některé oblasti může zdát, že v celé čtvrti je vysoká (nebo nízká) kriminalita, ačkoliv možná může být veškerá zločinnost pouze dílem jednoho či dvou míst s velmi častým výskytem trestné činnosti. Proto je STAC založen na aktuálních shlučích případů nebo místech na mapě. Při práci s touto metodou nastavuje uživatel opět několik parametrů, které ovlivňují v konečné fázi vzhled výstupu. Jedná se o následující parametry:

Poloměr hledání – poloměr hledání je klíčovým nastavením celé metody a všeobecně platí, že čím větší radius je volen, tím více incidentů je zahrnuto do shluku a velikost elipsy je rovněž větší. S menším poloměrem vzniká naopak více malých elips s menším počtem incidentů. Při rozhodování je třeba zvážit, pro jak velkou geografickou oblast je měření prováděno. Tvůrci programu doporučují zvolit pro městské okrsky radius 200m a pro města celá 750m. Vezmeme-li však v úvahu, že celým městem autoři miní např. Chicago, je jasné že pro potřeby českého uživatele, zkoumajícího stotisícové krajské město, se budou parametry poněkud lišit.

Minimální počet bodů ve shluku – toto nastavení je shodné jako u předchozí metody. Uživatel určuje minimální počet incidentů, jež mají být zahrnuty do shluku. Minimum jsou dva, program doporučuje standardně deset.

Hranice měření – uživatel musí pomocí mřížky vymežit oblast, ve které metoda proběhne. Buď zvolí mřížku vytvořenou samotnými daty (maxima a minima souřadnic X a Y) nebo mřížku vytvořenou v programu CrimeStat. Ta je tvořena uživatelem v referenčním souboru při vkládání dat do programu. Doporučuje se právě tato.

Typ prohledávání – tímto způsobem se volí metoda zkoumání mřížky. V nabídce jsou dvě možnosti. Obdélníkový vzor, jestliže má oblast zkoumání spíše pravidelný charakter (pravidelnou síť ulic) nebo vzor trojúhelníkový pro oblasti analýzy s převážně nepravidelnou strukturou.

Metoda generuje těchto sedm výstupů:

1. Shluk – identifikační číslo každé elipsy
2. Souřadnice X a Y centrografického průměru elipsy
3. Rotace – stupně, o kolik je elipsa pootočená (0 je horizontálně)
4. Délka x a y os elipsy
5. Plocha elipsy ve čtverečních jednotkách. Elipsy jsou seřazeny dle plochy
6. Počet incidentů ve shluku, elipse
7. Hustota shluku – počet incidentů na jednotku plochy

Výhody metody STAC: STAC má jako shlukový algoritmus mnoho výhod

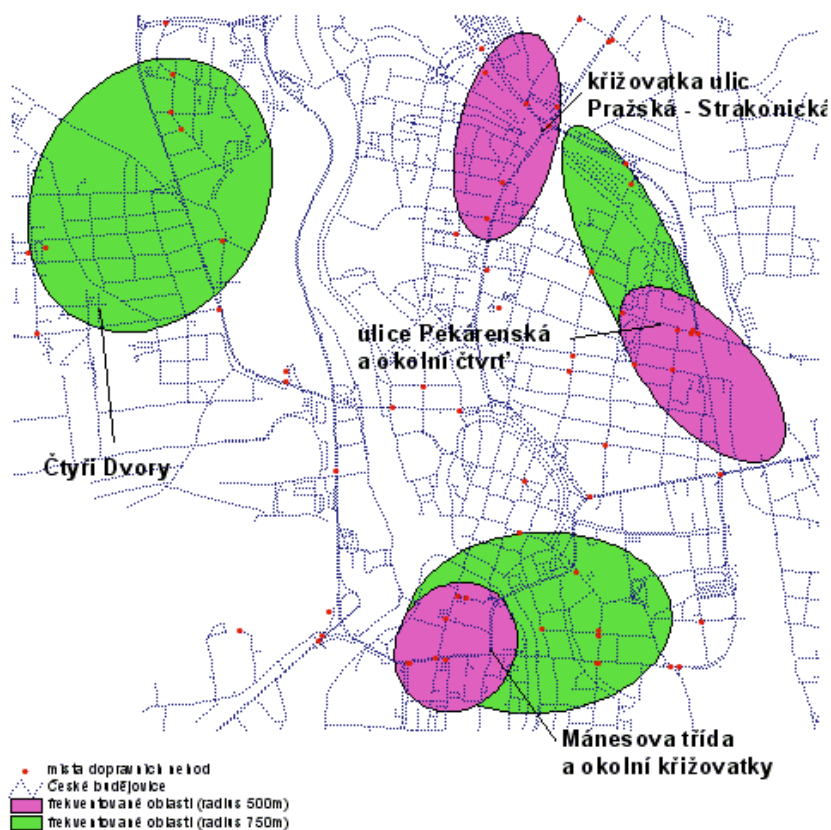
- STAC dokáže rychle analyzovat ohromné množství případů. Pro databáze čítající až několik tisíc případů je tedy výborným analytickým nástrojem. Manuál programu pracoval například s 13000 loupeží, což je enormní počet.
- uživatel metody má pod kontrolou přibližnou velikost elips, minimum bodů v elipse a studovanou oblast. Tyto rysy dovolují rozsáhlé pátrání po Hot Spots v celém městě či velkém územním celku.
- STAC a hierarchické shlukování se doplňují. Hierarchické shlukování nejprve vyvodí malé elipsy a ty pak zahrne do větších. Doporučený postup ve STAC je vyvodit elipsy velkých měřítek a tyto takticky analyzovat.
- Frekventovaná místa nemusí být nutně limitována určením pro jeden druh zločinu, místa atd. Elipsy drogových zločinů mohou být například proloženy elipsami vloupání a kriminalisté tak mohou vypátrat jistou příčinnost mezi činy tam, kde se elipsy překrývají.

Nevýhody metody STAC:

- STAC je založen na distribuci bodových dat a proto není počítáno jak s faktory rizika, tak dalšími jinými faktory
- malé změny v mřížce ohraničující studovanou oblast mohou vést k dosti rozdílnému vykreslení elips, což je problémem každé shlukové metody. Mřížku se proto doporučuje při opakovaných analýzách neměnit.

Obrázek 15 poskytuje názorný příklad využití této metody k identifikaci frekventovaných oblastí srpnových dopravních nehod. Slovo oblasti jsem zde použil záměrně, neboť nastavením větších hledacích radiusů, jako v případě této analýzy, došlo k vyhledání menšího počtu lokalit, zato však o větší rozloze a s větším počtem případů. Ostatní parametry jsem zadal následujícím způsobem, a to minimální počet bodů ve shluku 5 a vzor ulic nepravidelný (triangulární).

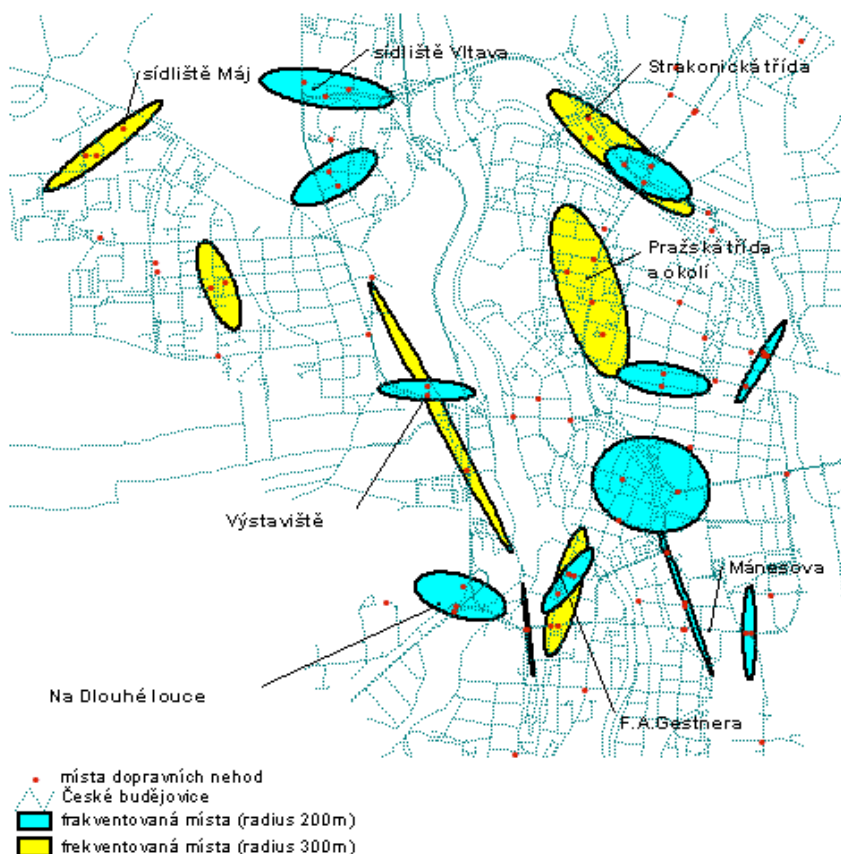
Obr. 15: Určení frekventovaných oblastí dopravních nehod metodou STAC



Analýza s radiusem 500m krásně odhaluje a ohraničuje tři nejpálčivější místa ve městě s osmi až devíti dopravními nehodami. Mánesova ulice (část) a její okolí, ulice Pekárenská s přilehlou čtvrtí činžovních domů a oblast kolem křižovatky Pražské a Strakonické třídy. Dvě z těchto míst jsou co do provozu velmi vytížená a tudíž je nadměrný výskyt dopravních nehod přirozený. Vysoká nehodovost ve víceméně residenční čtvrti kolem Pekárenské ulice však již nenabízí tak jednoznačné řešení. Jde možná o následek špatné přehlednosti zdejších křižovatek.

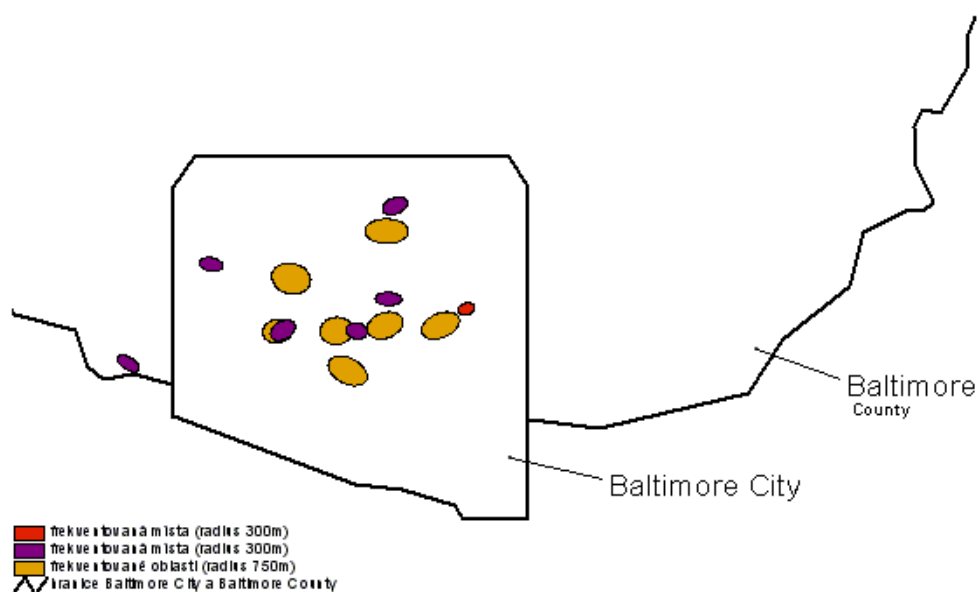
Za použití radiusu 750m vygeneroval program rovněž tři oblasti. Dvě z nich se částečně shodují a překrývají s předchozí analýzou, pokrývají však o poznání větší území s více případy. Elipsa o rozloze 0.7 čtverečního kilometru, pohlcující celou Mánesovu ulici, čítá dvacet dopravních nehod a prozrazuje tak pravděpodobně nejrizikovější ulici města.

Obr. 16: Určení frekventovaných míst dopravních nehod metodou STAC



Pro přehled o menších lokalitách, vyskytujících se v jednotlivých čtvrtích měst a okscích, se používá obecně malých hledacích poloměrů. Podívejme se tedy na výsledky analýz s menším poloměrem hledání, konkrétně 300m a 200m, které jsou vidět na obrázku 16. Další podstatnou změnou v nastavení bylo snížení minima bodů ve shluku na 3. Takovéto nastavení nám tedy odhaluje mnohem menší regiony. V některých případech, jako třeba v oblasti Mánesovy ulice, došlo jen k rozkouskování původní velké elipsy na jednotlivé mikroregiony. Některá místa byla na rozdíl od předchozí analýzy objevena zcela nově, jako např. křižovatka Strakonické ulice a Husovy ulice či křižovatka u výjezdu na Český Krumlov. Počet incidentů v elipsách je v rozmezí od 2 do 5 a jsou rozmístěny na ploše od 10 až po 240 metrů čtverečních.

Obr. 17: Identifikace frekventovaných míst loupeží ve městě Baltimore metodou STAC



Obrázek 17 je výsledkem aplikace metody STAC na loupeže ve městě Baltimore. Použil jsem tři druhy měření, pokaždé s jiným poloměrem hledání a nastavením minimálního počtu bodů ve shluku. První měření bylo s poloměrem 750m a minimem 15 činů ve shluku. S tímto nastavením vygenerovala metoda sedm míst, nacházejících se a odhalující z předchozí metody již známé lokality ve

středu a severu katastru města. Zajímavým zjištěním je fakt, že největší koncentrace činů 20 byla zjištěna v nejmenší elipse o rozloze 650 čtverečních metrů. Tento fakt odhaluje skutečně „horkou“ lokalitu s hustotou 30 činů na jeden čtvereční kilometr. Jedná se o nejzápadněji položenou žlutou, z části se překrývající elipsu. Výsledek druhého měření (radius 500, minimum bodů 10) vyhledal v tomto místě rovněž elipsu s největším množstvím deliktů. Konkrétně se jedná o 13 krádeží s hustotou 30 činů na čtvereční kilometr. Další lokality, vyhledané radiusem 500m, reprezentují skutečné mikroregiony krádeží, neboť jde ve čtyřech z pěti případů o zcela izolovaná místa nevelké rozlohy (všechna kolem 300 metrů čtverečních) s koncentrací až k 35 činům na kilometr čtvereční.

A konečně kalkulace s poloměrem hledání 300m, která odhalila jediné místo na východním okraji města s nejmenší rozlohou, avšak největší hustotou 48 činů na kilometr čtvereční. Toto vše tedy může pro policii představovat jasnou indicii o tom, kam zaměřit svou pozornost.

Na základě analýz, demonstrovaných v této kapitole, lze usuzovat o vhodnosti a praktičnosti použití jednotlivých metod. První předvedené metody, tedy mode a fuzzy mode, patří mezi velice jednoduché, navíc ani nevykazují grafické výstupy a uživatel si musí dle souřadnic data vizualizovat sám, což je při větším množství lokalit únavné. Proto je lepší rovnou použít metody grafické.

První z nich byla metoda hierarchických shluků. Ta mě o své užitečnosti přesvědčila pouze v případě, pokud má uživatel data pro sekundární složku a může tak uplatnit faktor rizika. Získat tato data je však zásadní problém, neboť oproti americkému systému sběru a uchování dat je ten český ještě v plenkách. Jako nejlépe aplikovatelná metoda se mi proto jeví metoda STAC, která svým širokým rejstříkem nastavitelných parametrů umožňuje uživateli získat na výstupu přesně taková data, která si přeje. Její vizuální stránka je rovněž více než příznivá.

4.3 Metody prostorového modelování

4.3.1 Jádrová (vnitřní) interpolace (Kernel density interpolation)

V této kapitole budeme diskutovat nástroje k provedení interpolace incidentů s využitím jádrového přístupu. Interpolace je technika, která zevšeobecňuje umístění incidentů na celou zkoumanou oblast. Zatímco metody prostorového rozdělení a „hot spots“ poskytují statistický přehled o zkoumaných datech samotných, interpolace generalizuje tato data na celý region a vytváří z bodů spojitý povrch. Poskytuje zejména odhady hustoty pro všechny části regionu a to v jakémkoliv místě. Odhad hustoty je proměnná intenzity, tzv. Z-hodnota, která je odhadována ve specifickém místě. Následně potom mohou být intenzity ve všech místech zobrazeny buď v podobě povrchových nebo vrstevnicových map.

Interpolačních technik existuje mnoho, ale většina z nich vyžaduje proměnnou odhadovanou na základě funkce umístění incidentů. Naproti tomu jádrová interpolace je technika vyhovující individuálnímu rozmístění bodů, tedy bez funkce rozmístění. Základem jádrové interpolace je odhad hustoty. Ten spočívá v umístění symetrického povrchu přes každý bod, vyhodnocení vzdálenosti od každého bodu k referenčnímu místu, stanovenému matematickou funkcí, a sečtení hodnot všech povrchů pro toto referenční místo. Tato procedura je pak opakována pro všechna referenční místa.

Poprvé byla tato metoda vynalezena na konci 50. let při odhadu hustoty histogramu. Bohužel však tato metoda způsobovala nejen jeden statistický problém, jenž se podařilo vyřešit až trojici vědců, kteří k odstranění chyb z metody použili několik *vnitřních (jádrových) funkcí*.

Klasická vnitřní distribuční funkce:

$$g(x_j) = \sum \left\{ [W_i * I_i] * \frac{1}{h^2 * 2\pi} * e^{-\left[\frac{d_{ij}^2}{2 * h^2}\right]} \right\}$$

kde d_{ij} je vzdálenost mezi polohou incidentu a jakýmkoliv referenčním bodem v okolí, h je standardní odchylna normální distribuce, W_i je váha v místě incidentu a I_i je intenzita v místě incidentu. Funkce plyne ve všech směrech do nekonečna a proto se vztahuje ke každému místu oblasti.

A dále byly navíc použity další čtyři alternativní vnitřní funkce, jimiž disponuje i program CrimeStat. Všechny tyto funkce mají omezený poloměr (na rozdíl od klasické distribuční funkce) a jsou vztaženy k limitované oblasti kolem každého bodu, definované tímto poloměrem. **Kvartická funkce** klesá postupně se vzdáleností, dokud nedosáhne poloměru. Vypadá následovně:

1. Vně specifikovaného poloměru: $g(x_i) = 0$

2. Uvnitř poloměru:

$$g(x_j) = \sum \left\{ [W_i * I_i] * \left[\frac{3}{h^2 * \pi} \right] * \left[1 - \frac{d_{ij}^2}{h^2} \right]^2 \right\}$$

kde d_{ij} je vzdálenost mezi místem incidentu a jakýmkoliv referenčním bodem v regionu, h je poloměr hledané oblasti, W_i je ukazatel váhy v místě incidentu a I_i je ukazatel intenzity v místě incidentu.

Triangulární funkce klesá rovnoměrně se vzdáleností, lineárním způsobem. Ve srovnání s kvartickou funkcí klesá rychleji. Má rovněž ohraničený poloměr a vztahuje se proto také na limitovanou oblast kolem každého bodu. Vypadá takto:

1. Vně poloměru: $g(x_i) = 0$

2. Uvnitř poloměru:

$$g(x_j) = \sum [K - K/h] * d_{ij}$$

kde K je konstanta z počátku nastavená programem na 0.25 a poté změněna, aby bylo zajištěno, že hustoty nebo pravděpodobnosti nabývají odpovídajících hodnot.

Negativně exponenciální funkce klesá velmi rychle se zvětšující se vzdáleností k ohraničenému poloměru.

Vypadá takto:

1. Vně poloměru: $g(x_i) = 0$

2. Uvnitř poloměru:

$$g(x_j) = \sum A * e^{-K * d_{ij}}$$

kde A je konstanta a K je exponent. V CrimeStat $K = 3$ a A zpočátku 1, poté opět upravena, aby byly zajištěny správné hodnoty hustoty a pravděpodobnosti.

A nakonec **uniformní funkce**, která přiřítá váhu všem bodům v kruhu rovnoměrně:

1. Vně poloměru: $g(x_i) = 0$

2. Uvnitř poloměru:

$$g(x_j) = \sum K$$

kde K je opět konstanta zpočátku o hodnotě 0.1, posléze však změněna, aby byly zajištěny správné hodnoty hustoty nebo pravděpodobnosti. Ve výčtu výstupů jsem tuto funkci nakonec ani nedemonstroval, neboť její výsledky byly téměř totožné s výstupy funkce klasické a zobrazovat je dvakrát mi nepřišlo nejvhodnější. Věnoval jsem proto více prostoru ostatním funkcím.

Uživatel tedy může v programu CrimeStat použít a vybrat si k interpolaci dat do mřížkové struktury pět výše zmíněných rozličných funkcí. Všechny tyto funkce mezi sebou navzájem produkují rozdíly ve tvaru interpolovaného povrchu. Klasická distribuce váží všechny body ve studované oblasti, ačkoliv bližší body jsou váženy více, než vzdálenější.

Další čtyři techniky používají, jak již bylo popsáno, kružnicí ohraničeného prostoru kolem buněk rastru. Uniformní funkce váží všechny body v kruhu rovnoměrně. Kvartická funkce váží bližší body uvnitř kruhu intenzivněji, než ty vzdálenější, její klesání je však postupné. Stejným způsobem váží body i triangulární funkce, ale její klesání je strmější a rychlejší. A konečně negativně exponenciální funkce, která váží uvnitř kružnice bližší body nesrovnatelně intenzivněji, než vzdálenější body v kružnici.

Otázka, jakou funkci použít, je tedy na uživateli a jeho rozhodnutí, do jaké míry chce mít blízké body vzhledem k vzdálenějším váženy. S větším rozdílem příkládaných vah mezi blízkými a vzdálenými body (negativně exponenciální, triangulární funkce) dojde k interpolaci s větší variabilitou povrchu mřížky, citlivé na detail. V opačném případě, u zbylých třech funkcí, dojde spíše k uhlazení drobných rozdílů a větší generalizaci.

Další veličinou, kterou musí uživatel před interpolováním dat nastavit, je **šířka pásma (intervalu) funkcí** a to zejména jakým typem intervalu mají být definovány. K dispozici jsou dvě možnosti. Pokud se uživatel rozhodne pro **fixní interval**, musí jeho rozsah určit ve velikostních jednotkách (metrech, kilometrech). Pro klasickou vnitřní funkci je šířka pásma standardní odchylkou klasického rozdělení. Pro ostatní čtyři použitelné funkce je šířka pásma

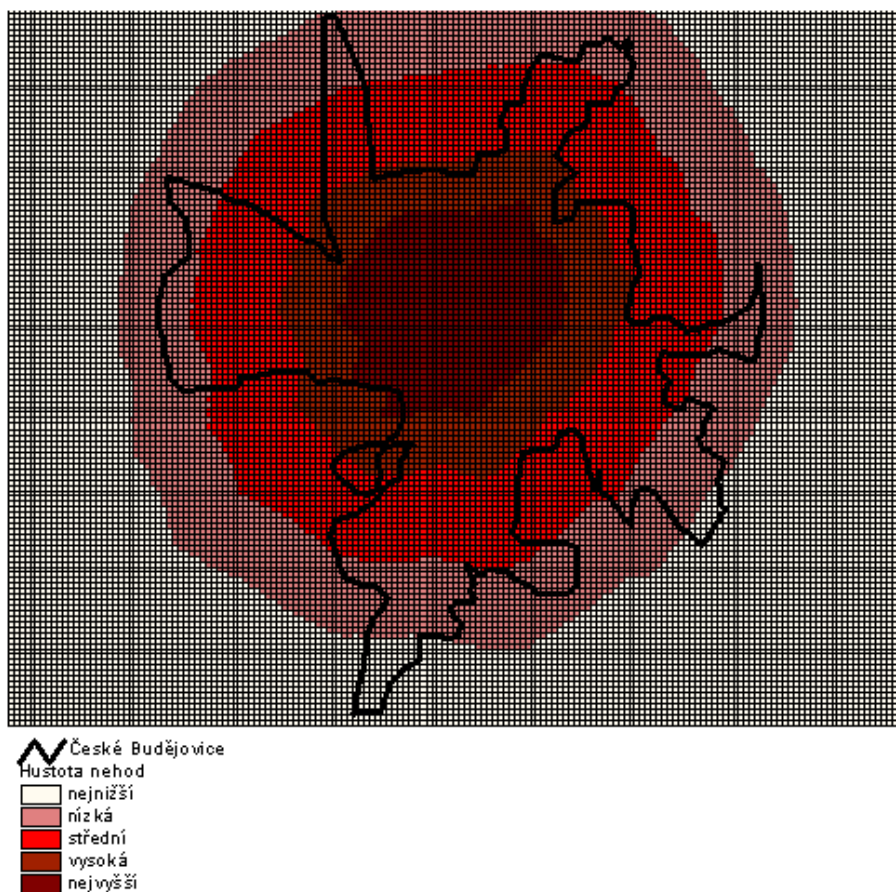
ohrazeným poloměrem funkcí s tím, že užší interval povede k jemnějšímu rozlišení vrcholů a klesání funkcí a širší interval k uhlazení distribuce a menší variabilitě mezi oblastmi.

Druhou variantou je volba **adaptivního intervalu**, který nastavuje šířku pásma tak, aby byl nalezen minimální počet incidentů, zadaný uživatelem. Toto nastavení skýtá výhodu v konstantní přesnosti měření odhadu hustoty pro celý region. Pokud chce uživatel dosáhnout rozmanitějších povrchů, zvolí v šířce pásma menší počet incidentů, pro zobrazení hladších povrchů zvolí vyšší počet bodů. Ideální nastavení není samozřejmě známo a proto jsem se o volbě intervalů rozhodl na základě velikosti souboru dat a výsledků provedených experimentů.

4.3.2 Výsledky analýz interpolovaných dat

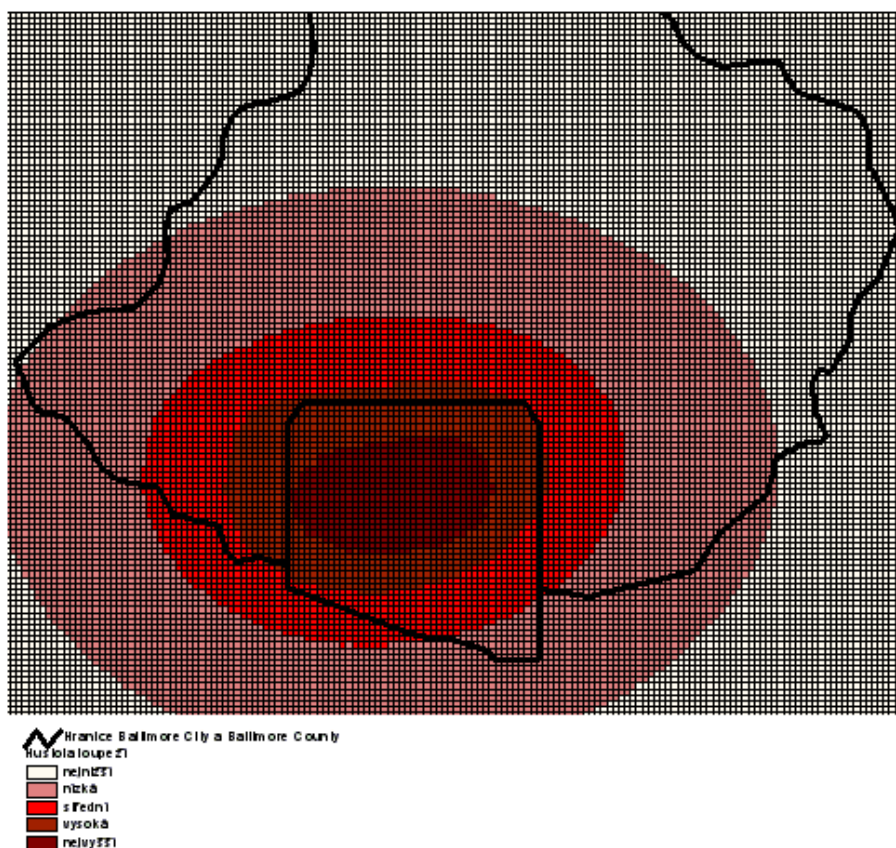
Pokud tedy budeme interpolovat data klasickou distribuční funkcí s adaptivním intervalem o minimálním počtu 40 bodů, získáme výsledky zobrazené na obrázku 18. Z těch je vidět, že klasická distribuční funkce opravdu váží všechny body ve vytvořené mřížce rovnoměrně, i když je patrné mírné zdůraznění váhy na body bližší, než vzdálenější. V každém případě poskytuje velice všeobecný přehled o hustotě incidentů a bude se hodit k analýze velkého množství dat na velkých územích. V tomto případě byla největší hustota nehod v Českých Budějovicích zaznamenána na velkém nevariabilním území, čítajícím centrum v okolí Pražské a Pekárenské ulice.

Obr. 18: Stanovení hustoty dopravních nehod v ČB jádrovou interpolací



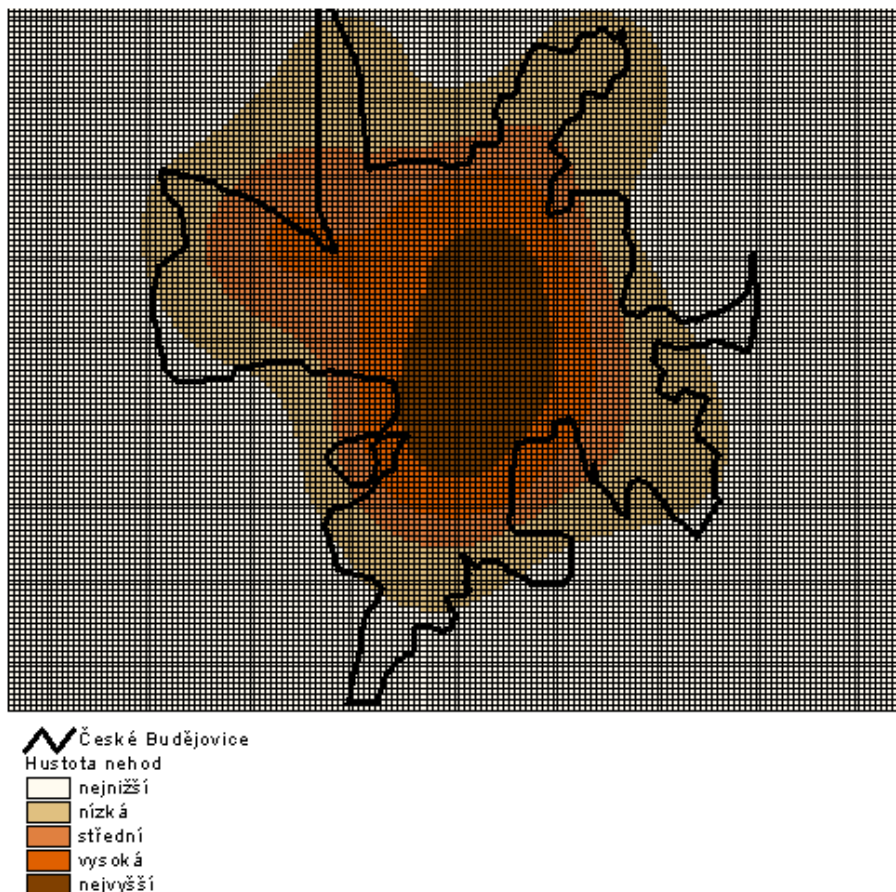
Na obrázku 19 města Baltimore vznikla největší hustota činů v nám již dobře známé části města. Oblast zaujímá podobný tvar, jako shluk druhého řádu z metody hierarchického skupinování a dá se tedy odhadnout jeho enormní rozloha až ke 40 kilometrům čtverečních, podtrhující všeobecnou aplikaci funkce. Bude proto jistě zajímavé sledovat nyní rozdíly mezi aplikacemi ostatních funkcí.

Obr. 19: Stanovení hustoty krádeží v Baltimore jádrovou interpolací



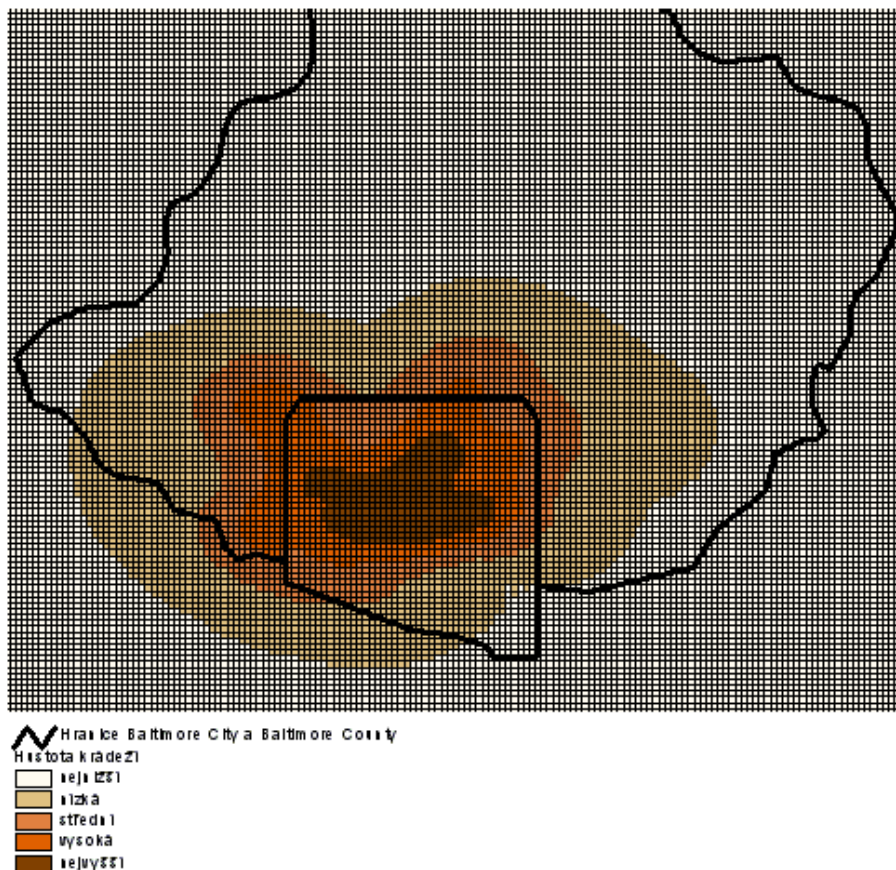
Obrázky 20 a 21 jsou výsledkem použití jádrové kvartické funkce, vážící již bližší body uvnitř ohraničeného intervalu intenzivněji, než vzdálenější, ale stále je na obrázcích patrné její postupné volné klesání, tudíž nevyvolává žádné extrémní změny a rozdělení interpolovaného povrchu je jemné. Znaky variability povrchu

Obr. 20: Stanovení hustoty dopravních nehod v ČB jádrovou interpolací (kvartická funkce)



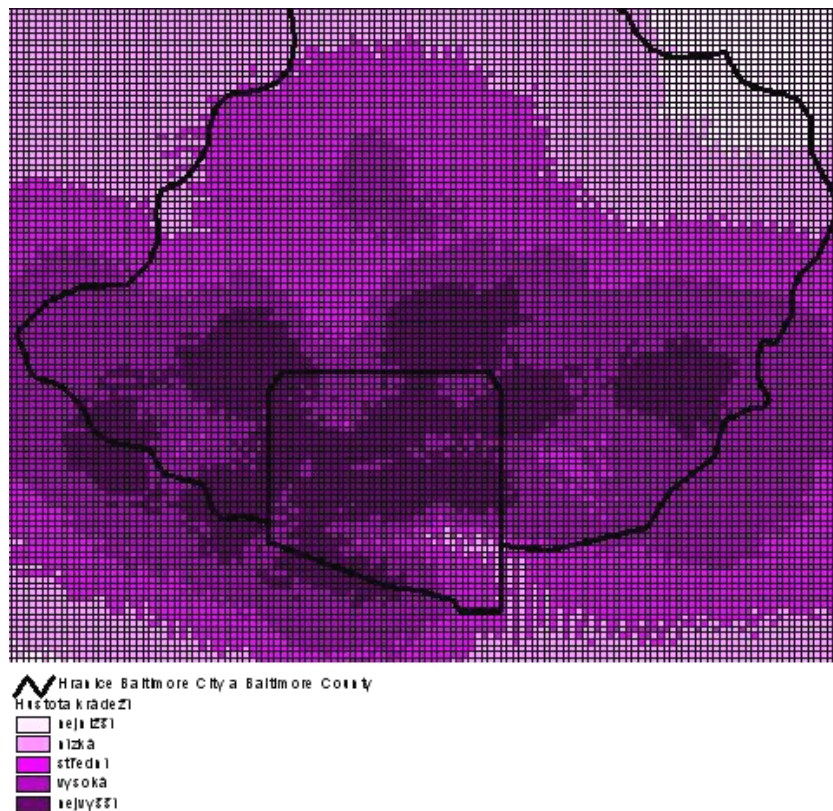
sice ještě nejsou nijak markantní, ale již jsou patrné, jako např. oddělující se útvar v oblasti čtvrti Čtyři Dvory v případě Českých Budějovic. O skutečných místech nehod nám však tato funkce s adaptivním intervalem mnoho nenapovídá, za což může nést velkou zodpovědnost nevelký soubor dat. Na datech krádeží z Baltimore jsou už však vidět trochu zásadnější změny a snímek celkem slušně demonstruje nejrizikovější části celé rozlehlé oblasti. Jako nástroj k odhadům hustoty velkých území, kde nás příliš nezajímají „hot spots“, ale právě informace obecnější, se zde kvartická funkce osvědčila. Interval byl v tomto případě vybrán rovněž adaptivní, s minimálním počtem incidentů 100, neboť zde se jedná o rozsáhlejší soubor dat a velké území. Toto nastavení bylo aplikováno na všechny analýzy krádeží města Baltimore.

Obr. 21: Stanovení hustoty krádeží v Baltimore jádrovou interpolací (kvartická funkce)



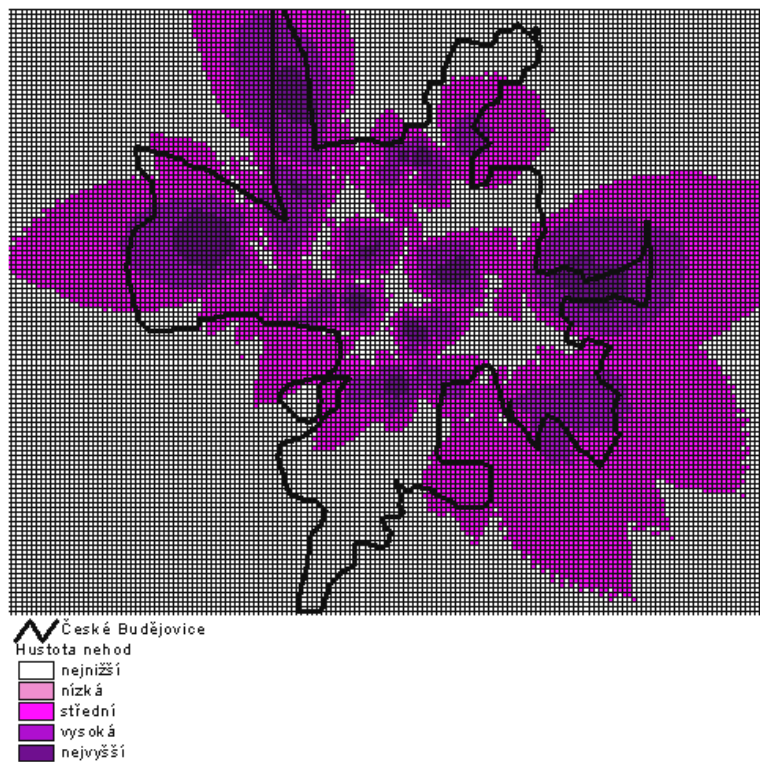
Na obrázku 22 je znázorněna interpolace lokalit loupeží města Baltimore triangulární funkcí, která nám jako první z funkcí umožňuje nahlédnout na rozmanitější povrch, ze kterého jsou již zřejmá i vysoce koncentrovaná místa incidentů. Je to dáno prudkým klesáním funkce směrem k hraničícímu intervalu. Tam, kde funkce kvartická klesala mírně, má funkce triangulární pokles strmější. Výsledkem je toto zobrazení, jehož koncentrovaná místa krádeží, znázorněná nejtmaší barvou, jsou mnohdy identická s výsledky frekventovaných míst shlukových analýz.

Obr. 22: Stanovení hustoty krádeží v Baltimore jádrovou interpolací (triangulární funkce)

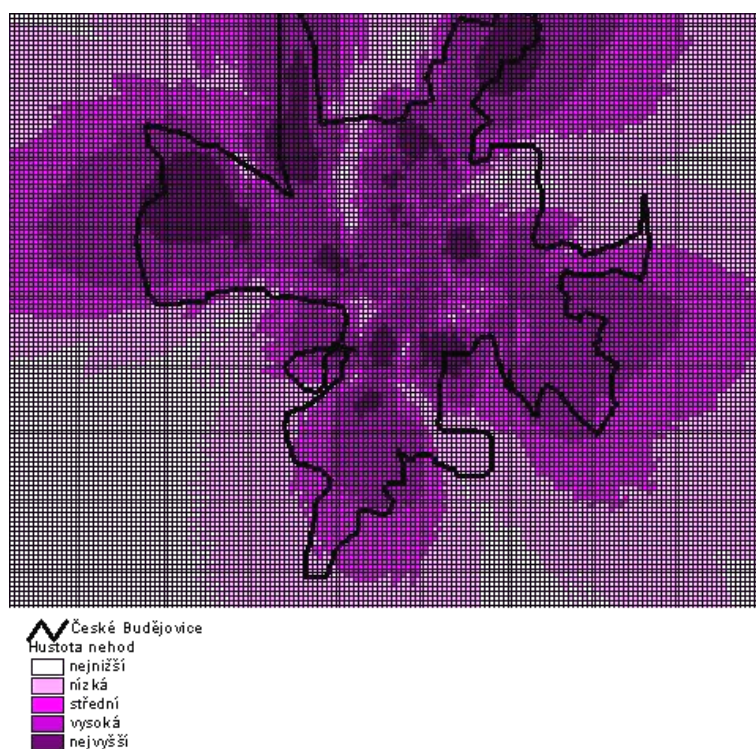


Obrázky 23 a 24, znázorňující hustotu dubnových a srpnových nehod v Českých Budějovicích, jsou v ukázce nerovnoměrnosti interpolovaného povrchu ještě názornější. Je to zřejmě rovněž proto, že byl kromě triangulární funkce v obou případech použit i fixní interval této funkce s hodnotou 1 km, který má sklon více rozlišovat vrcholky a klesání funkce. Nejtmavší barvou jsou na obrázku 19 vyznačena všechna nejčastější místa dubnových nehod čítající křižovatku Mánesova – Lidická, ulici Na Sadech, Pekárenskou ulici a Palackého náměstí, okolí křižovatky Na Dlouhé louce – Husova, ulici Karoliny Světlé, sídliště Máj a Kubatovu ulici. Všechna tato místa, s výjimkou Kubatovy ulice, jsou důležitými dopravními uzly města a tak je důvod vysoké nehodovosti v těchto místech přirozený a předvídatelný. Pokud tato místa porovnáme s místy vysoké hustoty nehod za měsíc srpen (obrázek 20), zjistíme, že se téměř všechna z nich shodují. Konkrétně jde o Mánesovu ulici, v tomto případě křižovatka s Čechovou ulicí a ulicí F. A. Gerstnera, což znovu dokazuje, že je Mánesova

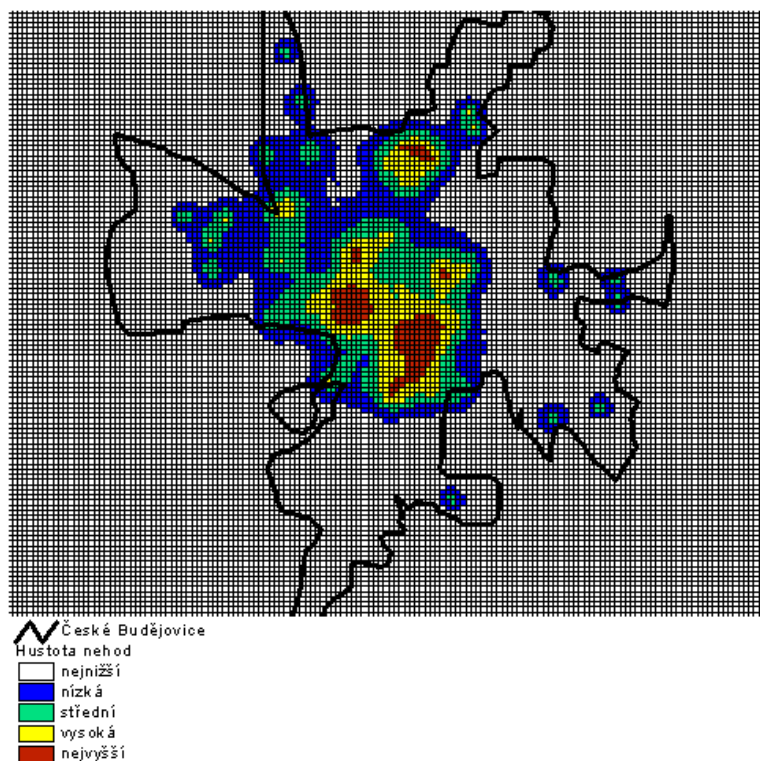
Obr. 23: Stanovení hustoty dubnových dopravních nehod v ČB jádrovou interpolací (triangulární funkce)



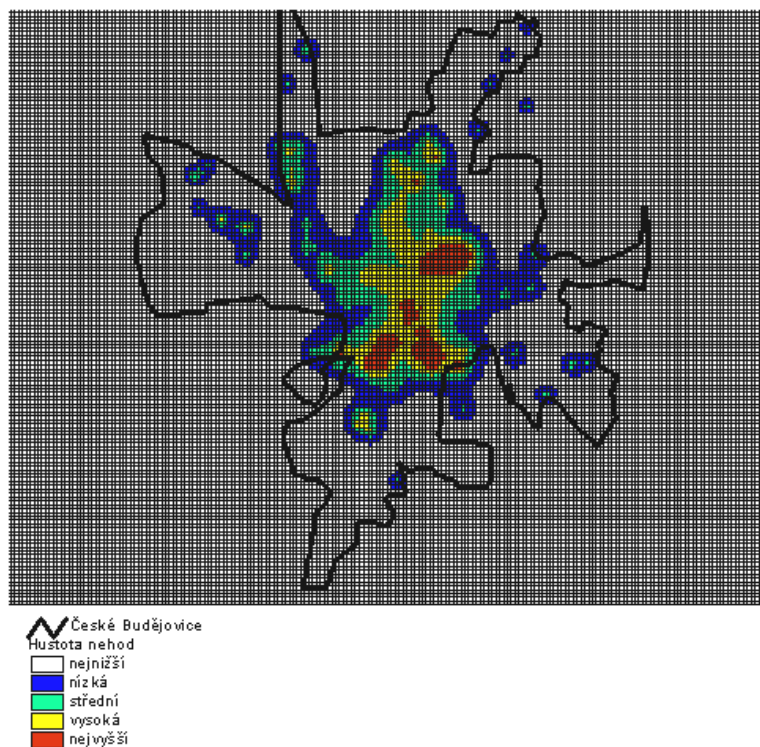
Obr. 24: Stanovení hustoty srpnových dopravních nehod v ČB jádrovou interpolací (triangulární funkce)



Obr. 25: Stanovení hustoty dubnových dopravních nehod v ČB jádrovou interpolací (negativně exponenciální funkce)



Obr. 26: Stanovení hustoty srpnových dopravních nehod v ČB jádrovou interpolací (negativně exponenciální funkce)



třída jednou z nejrizikovějších ulic v celém městě, což je vzhledem k její délce, dopravní vytíženosti a počtu křižovatek pochopitelné. Další shodná místa s vysokou koncentrací jsou pozorovatelná v oblasti křižovatek Pekárenská – Nádražní, Na Dlouhé louce – Husova a ulice Karoliny Světlé. Není jistě náhoda, že právě na dvě posledně jmenovaná místa byly umístěny měřicí radary. Jedinou lokalitou, která se při analýze srpnových nehod objevila poprvé, je křižovatka ulic Boženy Němcové – L. B. Schneidera u krajské nemocnice.

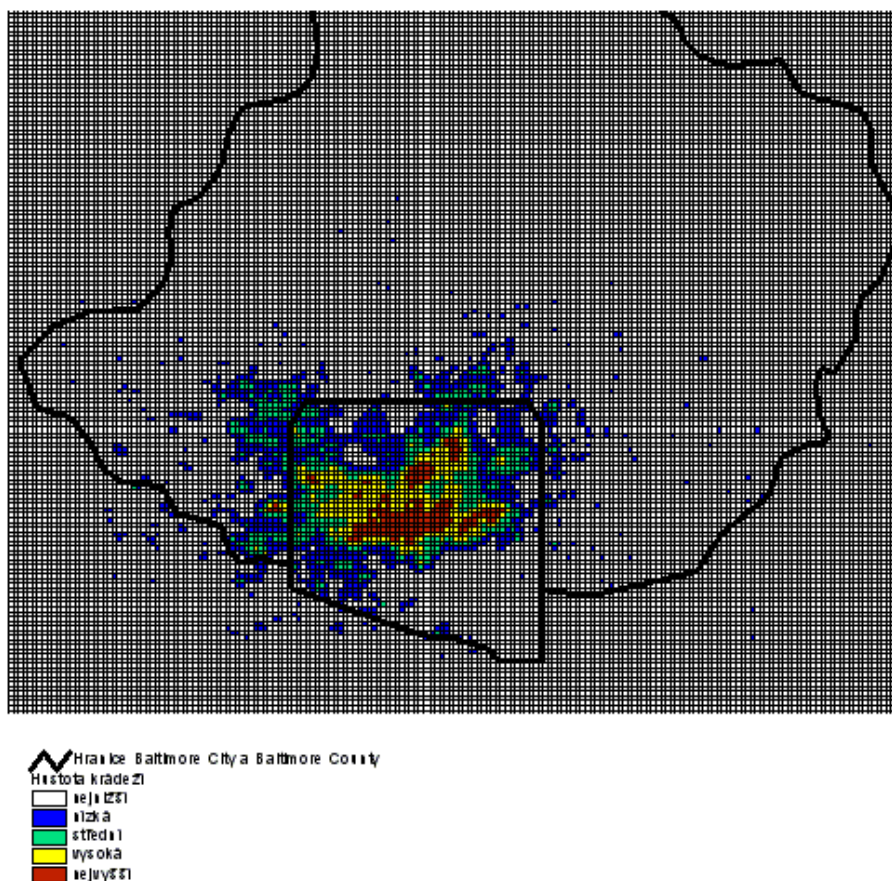
Vysokou hustotu nehod však na obou obrázcích vykazují i některá místa, kde k nehodám došlo pouze v minimálním množství, jako např. Pražská třída směrem na Nemanice a stejně tak velké plochy sídlišť Vltava a Máj v případě srpnových nehod a oblast Nové Vráto a sídliště Vltava v případě nehod dubnových. Tyto nedostatky jsou u použití této funkce s fixním intervalem přirozené a musí se tedy počítat s mírnou nepřesností měření.

Obrázky 25 a 26 zobrazují interpolace míst dubnových a srpnových dopravních nehod negativně exponenciální funkcí. Jejím použitím nám vznikla naprostá variabilita a nehomogenost interpolovaného povrchu mřížky, což přesně odpovídá povaze funkce. Ta přikládá bližším bodům nesrovnatelně větší váhu, než bodům vzdálenějším a se vzrůstající vzdáleností k ohraničenému poloměru prudce klesá. Proto je také povrch mřížky interpolován pouze částečně a některé incidenty jsou naprosto izolovány v podobě osamocených výstupků. Typickým znakem funkce jsou také rozlohou nevelká, avšak co do incidentů vysoce koncentrovaná místa. V případě dubnových dopravních nehod jsou krásně viditelná a jde opět o křižovatku ulic Mánesova – Čechova, v tomto případě tvořící rozsáhlejší územní celek spolu s ulicí Na Sadech. Dále si lze všimnout malých červených bodů na Kubatově ulici, Palackého náměstí a celé ulice Karoliny Světlé. Poslední inkriminovanou oblastí je Husova ulice v úseku od centra města ke křižovatce s ulicí Na Dlouhé louce.

Místům s největší koncentrací srpnových nehod dominuje opět Mánesova třída a její dvě velké křižovatky (s ulicemi F.A. Gerstnera a Čechovou), rozlehlá oblast kolem Pekárenské ulice a v podobě červených bodů křižovatky ulic Pražská – Strakonická a Pražská – K. Světlé. Jistou shodu lze tedy vidět i zde. Vizuální rozlišení obou souborů dat je citlivé, odhalující i zcela izolované případy a z hlediska efektivity se zdá být pro dané území nejpřehlednějším. Přesto došlo i zde k vytvoření vrcholků funkce (odhadování vysoké hustoty) v místech, která tomu ve skutečnosti neodpovídají. V tomto případě jde o červenou barvu

zvýrazněné centrum města při odhadu hustoty srpnových nehod na obrázku 26. V těchto místech totiž ve skutečnosti došlo pouze ke dvěma dopravním nehodám a tak je zde vysoká koncentrace pouze další nepřesností použité funkce.

Obr. 27: Stanovení hustoty krádeží v Baltimore jádrovou interpolací (negativně exponenciální funkce)

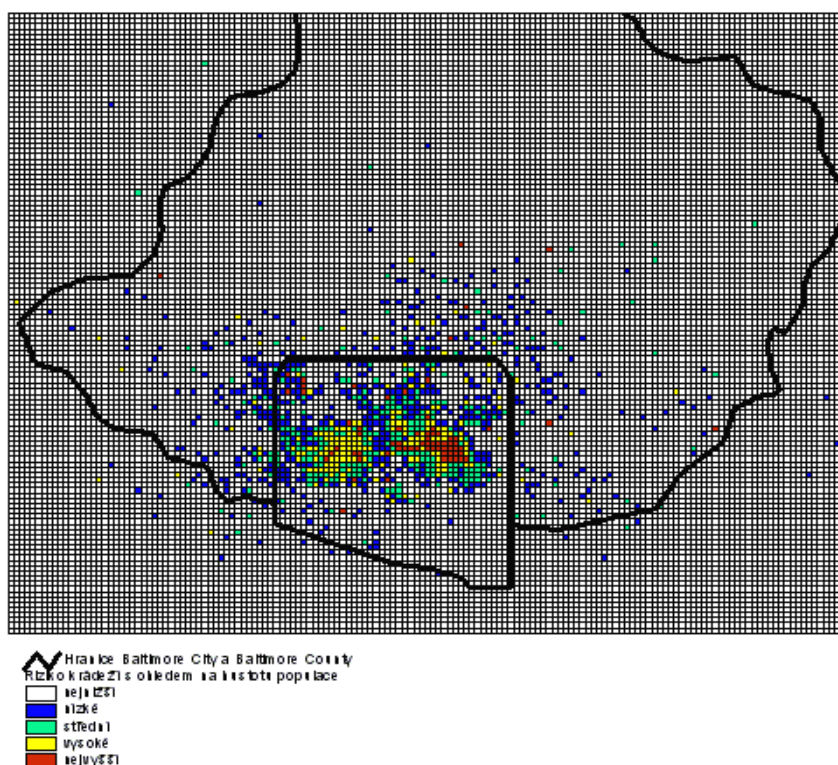


Data krádeží města Baltimore (obrázek 27), interpolovaná negativně exponenciální funkcí, již také nepřinášejí nijak překvapivé informace, neboť o lokalizaci nejfrekventovanějších míst a příčinách jejich vzniku (gentrifkace) jsme se už několikrát přesvědčili. Jsou umístěna ve středu města, které však není, což zde zdůrazňuji, centrem ekonomickým či komerčním, ale spíše populačním. Tak, jako dokázala pracovat metoda hierarchických shluků se sekundární složkou dat a využívat ji k odhadu rizikových oblastí podle rozdělení druhé proměnné, dokáže i program CrimeStat použít tuto sekundární složku dat a incidenty interpolovat

v mřížce položené přes sekundární proměnnou, coby základnu. Metoda se nazývá **duální jádrová hustota (Dual Kernel density)**. Jejím použitím na data z Baltimore získáme odhady hustoty krádeží vzhledem k hustotě populace podle sčítacích bloků. Jedná se tedy o dvě interpolace zároveň. Na obrázku 28 lze spatřit výsledky duální jádrové interpolace s klasickou distribuční funkcí, kde můžeme snadno porovnat změny mezi interpolací jedné proměnné a interpolace dvou proměnných, v tomto případě tedy navíc populace. Na předchozím případě jednoduché interpolace (obrázek 27) byla nejkonzentrovanejší místa pravidelně rozprostřena ve středu města a jeho blízkém okolí.

S použitím souboru dat sekundární složky však došlo k naprosté změně dat na výstupu. U předchozí spojitě koncentrace ve středu města dochází nyní k rozdělení na dvě samostatné části, orientované východním a západním směrem. Oba útvary jsou jasným výsledkem rozvržení populační základny na ploše města a nacházejí se přesně v místech residenčně-kulturních čtvrtí, kde je největší hustota obyvatel a stále tak i dostatek sociálně-kulturních aktivit, které nepřímo podporují páchaní trestné činnosti, zejména pak drobných krádeží. Východní část města je tak, dle této analýzy na obrázku, pro její obyvatele z hlediska rizika krádeží nejrizikovější.

Obr. 28: Stanovení rizika krádeží v Baltimore duální jádrovou interpolací (klasická distribuční funkce)



Tímto bych završil výčet interpolačních metod a na závěr si je dovolím stručně shrnout a zhodnotit. Z provedených analýz je jisté, že se užitečnost a i celková aplikovatelnost jednotlivých interpolačních funkcí odvíjí od velikosti souboru dat a velikosti území, ze kterého data pocházejí. Pro nevelké město jako jsou České Budějovice, bude s malým souborem dat tedy vždy lepší přiklonit se k funkcím, které se vyznačují variabilními povrchy (triangulární, negativně exponenciální), protože ty ve výsledku generují vizuálně zajímavější výstupy a jak jsme se mohli přesvědčit, výstupy byly nejenom poutavé, ale hlavně užitečné ve smyslu odhalení největších ohnisek dopravních nehod.

Zbylé funkce nám nic konkrétního neprozradily. Místo lokálních ložisek odhalily koncentrace oblastního charakteru. Jejich využití proto směřuje patrně na obrovská území, kde cílem analytiků je mít obecný přehled, spíše než konkrétní informace. Dokonce ani na mnohem větší ploše a s větším datovým souborem města Baltimore nenašly tyto funkce (kromě kvartické) uplatnění. Snad jen duální interpolace s použitím dat o rozložení populace vykazuje zajímavá zjištění a z praktického hlediska se jeví jako nejdůležitější, neboť zohledňuje zejména riziko. V českém prostředí jsou však bohužel tato data pro analytiku nedostupná, neboť zde se provádí sčítání obyvatel naprosto odlišným způsobem.

Interpolace jako shluková metoda prokázala nesmírné možnosti uplatnění a s množstvím přidružených funkcí a jejich nastavením poskytuje uživateli velice pružný a výkonný nástroj pro geostatistickou analýzu.

5. Závěr

Závěr mé práce bych rád věnoval shrnutí výsledků, kterých bylo analýzami dosaženo. Zpracování souborů dat dopravních nehod z období mezi 22. 4. – 27.4 2005 a 18.–31. 8. 2005 jasně prokázalo a identifikovalo nejkrizovější ulice, křižovatky a oblasti Českých Budějovic co do koncentrace nehodovosti:

1) Jedná se v první řadě o celou Mánesovu třídu a čtyři na ni přilehlé křižovatky, jmenovitě s ulicí F. A. Gerstnera, Lidickou třídou, Čechovou ulicí a Novohradskou ulicí. Vysoký počet nehod se zde objevil v obou pozorovaných měsících a nikoliv bezdůvodně. Jedná se totiž o jednu z nejkřiklavějších dopravních tepen města, vedoucí dopravu od Českého Krumlova směrem do centra a spojující tak svou délkou přes 1.5 Km západní část města s částí východní. Navíc, spojením s Lidickou třídou se stává důležitou dopravní trasou i pro celý jih města a jižní předměstí. Vysoká nehodovost zde proto pro nikoho nemůže být velkým překvapením, uvážíme-li rovněž velký počet křižovatek a rovný charakter ulice, který svádí k nedodržování maximální povolené rychlosti.

2) Další nejkřiklavější místo představuje křižovatka ulic Pekárenská a Nádražní, kde by mohl být častou příčinou havárií rovněž vysoký objem dopravy, dále nedání přednosti v jízdě a nepozornost po vypnutí světelné signalizace. Křižovatka je vytížena dopravou plynoucí jak po Nádražní třídě, tak po Pekárenské ulici, která pokračuje dále do průmyslové zóny, kde je objem zejména nákladní dopravy značný. Všeobecně je křižovatka nepřehledná, zejména díky dvěma železničním viaduktům ve směru od průmyslové zóny, a její bezpečnosti zřejmě rovněž neprospívá právě Nádražní třída, jež je označena za silnici hlavní a po níž jezdí řidiči velmi rychle.

3) Třetím místem je překvapivě relativně klidná rezidenční čtvrť činžovních domů v okolí Palackého náměstí, respektive křižovatky zdejších silnic s Pekárenskou ulicí, která tuto čtvrť protíná. Mnohé z těchto křižovatek se staly místem dopravních nehod. Zde může být důvodem jejich malá přehlednost, neboť jde o relativně úzké ulice, kde navíc prostor zužují parkující automobily, a řidiči se zde musí řídit především rozlišením hlavní a vedlejší silnice.

4) Dalším místem s vysokým objemem dopravních nehod je Strakonická třída, což jistě není vzhledem k délce a dopravní vytiženosti této komunikace překvapující. Dopravní nehody se zde vyskytovaly na mnoha místech po celé délce, avšak největší koncentrace byly zaznamenány ve zhruba pětisetmetrovém radiusu kolem křižovatky s Pražskou třídou, jmenovitě v úseku výskytu velkých nákupních center (Baumax, Interspar, OBI atd.), na křižovatce u čerpací stanice Esso a na křižovatce ulic Pražská – Karoliny Světlé u pivovaru Budvar. Tato křižovatka (a celá ulice K. Světlé) byla v obou sledovaných obdobích jednou ze třech nejfrekventovanějších lokalit. Není proto překvapením, že právě zde byl umístěn jeden z radarů, zachycujících přestupky řidičů.

Důvody nehodovosti v oblasti mohou být následující. Jednak je to obrovský objem dopravy v celé oblasti, přes kterou proudí obousměrně veškerá doprava na trase Praha – Č. Krumlov – Rakousko, což tvoří z Č. Budějovic významné tranzitní město. Druhým důvodem je přilehlá obchodně – průmyslová zóna s mnoha výrobními podniky, firmami a supermarkety, která tuto, již tak vytiženou část města, dále zatěžuje zejména nákladní dopravou. Nakonec nelze nezmínit nedaleký výskyt největšího českobudějovického sídliště Vltava, kde reziduje několik tisíc obyvatel města, využívajících komunikace v oblasti denně k cestě do zaměstnání.

Pro zajímavost bych ještě rád připomněl, že se drtivá většina dopravních nehod z období mezi 22.4 – 28.4. 2005 obešla bez smrtelných úrazů a vážných zranění. Z počtu 48 dopravních nehod byla zaznamenána pouze čtyři lehká zranění (z čehož jedno po kolizi dvou cyklistů) a dalším jistě nadmíru pozitivním faktem je pak skutečnost, že přítomnost alkoholu byla prokázána pouze v jednom jediném případě ze všech.

Taková jsou tedy krizová místa nehod ve městě a bylo by jistě zajímavé sledovat, zda-li se trend nehodovosti na těchto místech bude opakovat i v dalších obdobích či nikoliv. To už je však na složkách policie rozhodnout se, zda tuto praktickou pomůcku bude využívat. Mým posláním a úkolem ji bylo v této práci představit a zejména obeznámit s její existencí, která v naší zemi není samozřejmostí, ba trůfám si tvrdit, že je výjimkou. Důvodem je zřejmě i nedostatečný sběr a uchování dat v digitální podobě, který je na západ od našich hranic samozřejmou záležitostí a program se tak stává pro mnoho analytiků, pracujících v zahraničních policejních odděleních, rychle, snadno a efektivně použitelným nástrojem pro práci s kriminalistickými daty.

Nezbývá nám tedy než si přát, aby se moderní způsoby zpracování dat implementovaly co nejrychleji i v České Republice a usnadnily tak používání aplikací, které kvalitní digitalizované údaje vyžadují. Program CrimeStat jednoznačně potvrdil, že je dnes prostorová analýza s využitím GIS v mnoha sférách lidské činnosti opodstatněná a nabývá nemalé důležitosti. V případě mapování výskytu, četnosti a pohybu kriminality její důležitost a užitečnost ještě vzrůstá. Cílem policie by proto dle mého názoru měla být snaha těchto moderních, technicky vyspělých nástrojů, co nejvíce ke svému prospěchu využívat.

6. Seznam použité literatury

1. Levine, N.: CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations (v 3.0), Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington, DC, May, 2002
2. Tuček, J.: Geografické informační systémy, Computer Press, Praha, 1998
3. Klufová, R.: Geografické informační systémy (cvičení), ZF JČU, České Budějovice, 2000
4. Klufová, R.: GIS – učební texty ZF JČU, České Budějovice, 2000
5. Tollingerová, D.: Geografické informační systémy, VŠ Báňská Ostrava, 1996
6. Davis, D.E.: GIS pro každého, Computer Press, Praha, 2000
7. Sýkora, L.: Teoretické přístupy a vybrané problémy v současné geografii, KSGRR, PřF UK, Praha, 1993

Použité prameny:

1. <http://www.icpsr.umich.edu/NACJD/crimestat.html>
2. <http://www.esri.com>
3. <http://www2.zf.jcu.cz/~klufova/>
4. <http://www.vsb.cz>
5. <http://www.czechradar.cz>

OBSAH

1. Úvod	1
2. Literární rešerže	2
2.1 Geografické informační systémy – GIS.....	2
2.2 Mezníky vývoje GIS	3
2.3 GIS a jiné počítačové systémy.....	4
2.4 Využití geografických informačních systémů	5
2.5 Činnosti v rámci projektu GIS	7
2.5.1 Stanovení cílů projektu	7
2.5.2 Budování databáze	8
2.5.3 Restrukturalizace a manipulace s údaji.....	10
2.5.4 Vykonání analýz a syntéz	11
2.5.5 Prezentace výstupů	11
2.6. Prostorová analýza dat	12
2.6.1 Definice prostorových analýz	12
2.6.2 Cíle prostorových analýz	13
2.6.3 Statistické prostorové analýzy	14
3. Metodika	19
3.1 Program CrimeStat	19
3.2 Získání potřebných dat a jejich úprava	19
3.3 Vkládání dat do programu CrimeStat	21
4. Propojení GIS a programu CrimeStat	26
4.1 Prostorové rozdělení	26
4.1.1 Grafické výstupy prostorově popisných analýz.....	30
4.2 Analýza frekventovaných míst (Hot Spots).....	38
4.2.1 Mode a Fuzzy Mode	39
4.2.2 Hierarchické shlukování metodou nejbližšího souseda (nearest neighbor hierarchical clustering)	42
4.2.3 Prostorová časová analýza kriminality (Spatial and Temporal Analysis of Crime - STAC)	47
4.3 Metody prostorového modelování	53
4.3.1 Jádrová (vnitřní) interpolace (Kernel density interpolation)	53
4.3.2 Výsledky analýz interpolovaných dat.....	57
5. Závěr	68
6. Seznam použité literatury	71

