

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

DIPLOMOVÁ PRÁCE

Brno, 2021

Bc. Dominik Walach



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV RADIOELEKTRONIKY

DEPARTMENT OF RADIO ELECTRONICS

ODHAD POKRYTÍ SIGNÁLEM MOBILNÍ SÍŤE PROSTŘEDNICTVÍM REGRESE

ESTIMATING MOBILE COVERAGE BY MEANS OF REGRESSION

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. Dominik Walach

VEDOUCÍ PRÁCE

SUPERVISOR

doc. Ing. Jiří Blumenstein, Ph.D.

BRNO 2021

Diplomová práce

magisterský navazující studijní program **Elektronika a komunikační technologie**

Ústav radioelektroniky

Student: Bc. Dominik Walach

ID: 195468

Ročník: 2

Akademický rok: 2020/21

NÁZEV TÉMATU:

Odhad pokrytí signálem mobilní sítě prostřednictvím regrese

POKYNY PRO VYPRACOVÁNÍ:

Seznamte se s volně dostupnými datovými sadami obsahujícími měření pokrytí v mobilních sítích. Prostudujte charakter měřených dat, typickou plošnou hustotu měření, popište soubor měřených parametrů. Navrhněte vlastní měřicí kampaň a proveďte měření pokrytí ve venkovním prostředí tak, abyste získali sít' měřených bodů s možností škálování prostorového rozlišení. Proveďte rešerši algoritmů aplikovatelných na měřená data tak, aby byl získán odhad pokrytí radiovým signálem v místech, kde měřená data nejsou k dispozici.

Vytvořte model, ve kterém budou srovnány jednotlivé algoritmy a určena jejich použitelnost pro odhad pokrytí sítě v lokalitách, kde reálná měření chybějí. Uvažujte také věrohodnost získaných odhadů s cílem optimalizovat lokality doplňujících měření.

DOPORUČENÁ LITERATURA:

[1] FAHRMEIR, Ludwig, Thomas KNEIB, Stefan LANG a Brian MARX. Regression [online]. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013 [cit. 2020-05-19]. DOI: 10.1007/978-3-642-34333-9. ISBN 978-3-642-34332-2.

Termín zadání: 8.2.2021

Termín odevzdání: 20.5.2021

Vedoucí práce: doc. Ing. Jiří Blumenstein, Ph.D.

prof. Ing. Tomáš Kratochvíl, Ph.D.
předseda rady studijního programu

UPOZORNĚNÍ:

Autor diplomové práce nesmí při vytváření diplomové práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

ABSTRAKT

Tato diplomová práce se zabývá využitím volně dostupných datových sad k odhadu síly signálu v lokalitách, kde měření chybí. Cílem bylo za pomoci interpolačních a regresních metod co nejpřesněji odhadnout možné pokrytí signálem v dané lokalitě. Nejdříve byly analyzované datové sady, na jejichž základě byl vytvořen testovací model různých metod. Metody s nejlepšími výsledky byly implementovány na reálných datových sadách obsahujících tisíce měření. V závěru je okomentována problematika přesných odhadů a jsou specifikovány okolnosti pro snížení odchylky naměřené a vypočítané hodnoty.

KLÍČOVÁ SLOVA

LTE, RSRP, regrese, interpolace, Open Data, RF měření, Python, analýza dat

ABSTRACT

This diploma thesis deals with the use of freely available data sets to estimate the signal strength in localities where measurements are missing. This work aimed to use interpolation and regression methods to estimate as accurately as possible the possible signal coverage in a given locality. First, data sets were analyzed, based on results, a test model for various methods has been created. The methods with the best results have been implemented on real data sets containing thousands of measurements. At the end of the work, accurate estimates are commented and circumstances to reduce the deviation of the measured and calculated values are specified.

KEYWORDS

LTE, RSRP, regression, interpolation, Open Data, RF measurement, Python, data analysis

WALACH, Dominik. *Odhad pokrytí signálem mobilní sítě prostřednictvím regrese*. Brno, 2021, 61 s. Diplomová práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav radioelektroniky. Vedoucí práce: doc. Ing. Jiří Blumenstein, Ph.D.

PROHLÁŠENÍ

Prohlašuji, že svou diplomovou práci na téma „Odhad pokrytí signálem mobilní sítě prostřednictvím regrese“ jsem vypracoval samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené diplomové práce dále prohlašuji, že v souvislosti s vytvořením této diplomové práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

Brno

.....

podpis autora

PODĚKOVÁNÍ

Rád bych poděkoval vedoucím diplomové práce panu doc. Ing. Martinovi Slaninovi, Ph.D. a panu doc. Ing. Jřímu Blumensteinovi Ph.D. za odborné vedení a konzultace při psaní mé diplomové práce.

Obsah

Úvod	11
1 Volně dostupné datové sady	12
1.1 Stupně otevřenosti datových sad	12
1.2 Výběr datových sad	13
1.3 Popis parametrů datové sady	14
2 Plošná hustota měření	16
2.1 Příprava dat	16
2.2 Haversinův vzorec	17
2.3 Analýza datasetu	18
2.4 Plošná hustota ve městě	20
3 Vlastní měřicí kampaň	22
3.1 Měřicí zařízení a ukládání výsledků	22
3.2 Měření v terénu	23
3.2.1 Časový vývoj	25
4 Regrese	27
4.1 Druhy regrese	27
5 Interpolace	30
5.1 Metoda nejbližšího souseda	31
5.2 Bilineární interpolace	32
5.3 Bikubická interpolace	33
5.4 Testovací metodika	34
5.5 Porovnání deterministických metod	35
5.6 Inverzní vážení vzdálenosti	36
5.6.1 Použití IDW na reálných datech	38
5.6.2 Přizpůsobení IDW metody	40
5.6.3 Omezení dosahu IDW	42
5.7 Kriging	44
5.7.1 Modely variogramu	46
5.7.2 Použití krigingu na reálných datech	47
6 Srovnání dostupných metod	50
6.1 Rozdělení podle procentuálních vah	50
6.1.1 Omezení nadbytečnosti vzorků	51

6.2	Otestování na vybraných bodech	52
6.3	Společný model	54
6.3.1	Otestování datasetu na menší ploše	54
	Závěr	57
	Literatura	58
	Seznam symbolů, veličin a zkratk	61

Seznam obrázků

1.1	Rozložení počtu měření v různých zemích v září 2019.	14
2.1	Mapa Rakouska se změřenými body v období leden až červen 2020. . .	19
2.2	Závislost četnosti měření Rakouska pro plochu 10 a 1 km ²	19
2.3	Mapy rozložení změřených bodů ve Vídni.	20
2.4	Histogram hustoty měření pro mřížku 1 x 1 km a 100 x 100 m.	21
3.1	Přehled měřicího systému, na základě [2].	23
3.2	Závislost měřené úrovně <i>RSRP</i> na čase pro stabilní držení telefonu (nalevo), s telefonem je manipulováno (napravo).	24
3.3	Mapa pilotního měření v městské části Brno-Ponava.	24
3.4	Mapa pilotního měření v městské části Brno-Ponava s LTE vysílači. . .	25
3.5	Porovnání rozdílů prvního a druhého měření v časovém období 4 měsíců. .	26
4.1	Graf nevhodného použití lineární regrese (nalevo) v porovnání se správným výsledkem (napravo).	28
5.1	Porovnání jednorozměrné (nalevo) a dvourozměrné interpolace (na- pravo).	30
5.2	Metoda nejbližšího souseda.	32
5.3	Bilineární interpolace.	33
5.4	Bikubická interpolace.	34
5.5	Porovnání odchylek odhadu a rozptylu deterministických metod.	35
5.6	Porovnání rozptylu interpolačních metod v závislosti na počtu od- straněných bodů.	36
5.7	Snižování relevantnosti změřeného bodu se vzrůstající vzdáleností. . .	37
5.8	Porovnání parametrů <i>p</i> na reálných datech.	38
5.9	Srovnání odhadu a rozptylu pro odlišné nastavení silového parametru <i>p</i>	39
5.10	Porovnání vlivu odlišného počtu iterací odhadu na výsledné přesnosti. .	39
5.11	Porovnání různých hodnot parametru <i>PLE</i>	41
5.12	Graf rozložení pokrytí signálem pro <i>PLE=3</i>	42
5.13	Porovnání odchylek odhadu a rozptylu původního IDW.	43
5.14	Model variogramu, převzato z [26].	45
5.15	Porovnání odlišných variogramů pro brněnský dataset.	48
5.16	Srovnání šíření signálů pro různé modely variogramů	49
5.17	Porovnání odchylek odhadu a rozptylu OK a UK.	49
6.1	Porovnání odchylek odhadu a rozptylu všech metod na datech z Vídně. .	50
6.2	Porovnání odchylek odhadu a rozptylu exponenciálního OK s odliš- ným počtem vstupních proměnných	51
6.3	Srovnání všech metod pro dataset Brno Lužánky	52

6.4	Srovnání všech metod pro vídeňský dataset	53
6.5	Porovnání odchylek odhadu ve společném modelu ve Vídni	54
6.6	Mapa části Vídně obsahující s početnou plošnou hustotou měření . .	55
6.7	Porovnání metod v části Vídně	55
6.8	Porovnání různých variogramů pro část Vídně.	56

Seznam tabulek

1.1	Evropské regulační orgány využívají RMBT [3].	13
2.1	Tabulka změny 1. stupně zeměpisné šířky a délky na geodetickém úhlu šířky	17
5.1	Hodnoty <i>PLE</i> pro různá prostředí [22]	40
5.2	Parametry variogramu krigingu pro brněnský dataset	47

Úvod

Lidé si čím dál tím více zvykají na pohodlí, které jim poskytuje datové připojení v mobilu. Operátoři pokrývají okolo 99% plochy České republiky, a není tedy problém se kdykoliv odkudkoliv připojit. Připojení je dostupné pomocí technologie pro vysokorychlostní přenos signálu – Long Term Evolution (LTE), blížící se svými vlastnostmi vlastněm pro 4G. V dnešní době je možné pohodlně a rychle dohledat informace o pokrytí vysokorychlostním 4G signálem v dané lokalitě. Tyto informace zveřejňují jak samotní operátoři (O2, T-Mobile a Vodafone), tak i Český telekomunikační úřad (ČTÚ). Jejich stránky však neobsahují informace o síle signálu a o reálných rychlostech připojení, které jsou pro koncového uživatele nejdůležitější.

Na trhu je velké množství různých měřících aplikací, které jsou schopné za několik sekund poskytnout poměrně přesnou informaci o parametrech sítě. Velká řada těchto aplikací nabízí výsledky měření volně přístupné veřejnosti. Tato diplomová práce se zabývá možnostmi využití zmíněných datových sad k odhadu pokrytí signálem LTE v místech, kde nemáme nebo nejsme schopni tato data získat.

K tomuto odhadu se dá využít značný počet různých metod. V průběhu let byla vyvinutá široká škála přístupů k odhadování pokrytí pomocí modelů šíření. Tyto modely jsou užitečné pro výpočet útlumu, ztráty šířením a dalších parametrů. Sofistikovanější metody jsou schopné pracovat s modely budov a lépe tak poskytnout odhad síly signálu v konkrétním místě. V této práci se budeme zabývat využitím různých interpolačních metod pro dosažení nejpřesnějších výsledků odhadu síly signálu Reference Signal Receive Power *RSRP*. Ke stažení, zpracování a vyhodnocení datových sad je využité prostředí Google Colaboratory a programovací jazyk Python.

Práce je rozdělena do šesti kapitol. V první kapitole jsou popsány vlastnosti volně dostupných datových sad, je popsáno jejich rozdělení a parametry, které obsahují. Druhá kapitola se věnuje bližší analýze vstupních dat, jako jsou plošná hustota měření, počet datových bodů a přepočty souřadnicového systému zeměpisné délky a šířky na metry. Ve třetí kapitole je popsána vlastní měřící kampaň, která vznikla analýzou předchozích kapitol. Ve čtvrté kapitole jsou popsány regresní algoritmy a jejich použitelnost v této práci. Pátá kapitola se věnuje představení vhodných interpolačních metod, které jsou otestované na vlastní datové sadě. Jsou diskutované odchylky naměřené a odhadované síly signálu a rozptyl odhadovaných hodnot *RSRP*. Jsou zobrazené modely šíření. V poslední, šesté kapitole jsou nejlepší metody otestovány na reálných datových sadách obsahujících tisíce měření.

1 Volně dostupné datové sady

V této kapitole diplomové práce budou představené volně dostupné distribuční datové sady. Datovou sadou se rozumí soubor informací, které spolu navzájem souvisí, jsou úplné, snadno dostupné, obsahují minimální omezení užití a je možné je strojově zpracovat.

Slovem distribuční je myšleno, že při jejich zveřejňování na síti musí být zvolena datová sada volně ke stažení. Obecně platí, že tyto datové sady by mělo být možné stáhnout jako jeden celek. V případě, že by daná datová sada byla příliš velká, je ji možné rozdělit, avšak musí se přesně definovat, jak byla rozdělená, a musí být uvedené všechny provedené změny. Upřednostňuje se dělení na co nejmenší počet. Tyto sady musí vždy obsahovat stejný obsah, avšak můžou se lišit v distribučním formátu. Jako nejrozšířenější formáty jsou díky své jednoduchosti a použitelnosti označovány formáty XLS a CSV [1].

XLS formát je přehlednější, a proto je používán převážně analytiky v tabulkovém editoru, jako je například Excel pro Windows či Libre Office v linuxových distribucích.

CSV se více využívá, pokud chceme staženou databázi použít pro následné strojové zpracování. Tento formát je více využíván právě programátory [1].

1.1 Stupně otevřenosti datových sad

Distribuční sady mohou být publikovány v různých formátech neboli stupních otevřenosti. Ty se obvykle označují počtem hvězdiček na stránkách poskytovatele dat. Jejich možné rozložení je popsáno níže.

- **1. stupeň**, data jsou uzavřena v dokumentu, není snadné je dále zpracovat, například .pdf.
- **2. stupeň**, je možné pracovat s daty v proprietárním softwaru, či převést na jiný formát, například .xls.
- **3. stupeň**, jedná se o zmíněné .csv soubory, se kterými budeme primárně pracovat. Tato data je možné zpracovávat bez nutnosti vlastnění proprietárního softwaru.
- **4. stupeň**, data jsou uložena přímo ve webu, je možné se na ně přímo odkazovat, například grafový formát RDF.
- **5. stupeň**, možnost nalinkování dat na jiná a tím zaručit další propojení, LOD formát.

1.2 Výběr datových sad

V současné době je dostupných mnoho distribučních sad od různých poskytovatelů mobilních služeb. Pro tuto práci byla vybrána data založená na volně dostupném testovacím systému Open-RMBT [2]. Tento systém umožňuje provádění testů a ukládání jejich výsledků na web, kde je možné je strojově zpracovat. Strukturu tohoto systému je možné rozdělit do dvou bloků:

- **Měřicí zařízení**, neboli terminál, který má dostatečný výpočetní výkon na zpracování výsledků bez jejich negativního ovlivnění. Může se jednat například o mobilní telefon nebo stolní počítač.
- **Měřicí server**, slouží jako protistrana pro měření konkrétních parametrů, jako je například rychlost stahování, nahrávání a odezvy. Opět platí, že výkon serveru a jeho kapacita úložiště musí být dostačující, aby nedocházelo ke znehodnocování výsledků.

Tento systém byl vybrán z důvodu jeho rozšířenosti a dostupnosti velkého počtu dosažených měření v Evropě. Umožňuje jeho uživatelům kromě již zmíněných parametrů také změřit sílu signálu rádiové sítě. V tabulce níže jsou zobrazené jednotlivé státy s institucemi, které si přizpůsobily tento nástroj podle svých požadavků. Jednotlivé nástroje se liší především v parametrech, které ukládají a zároveň poskytují osobám a organizacím, jež mají o tato měření zájem.

Tab. 1.1: Evropské regulační orgány využívají RMBT [3].

Stát	Regulátor	Nástroj
Rakousko	RTR	RTR-Netztest ¹
Chorvatsko	HAKOM	HAKOMETAR Plus ²
Česká republika	ČTÚ	NetMetr ³
Lucembursko	ILR	checkmynet ⁴
Srbsko	RATEL	RATEL NetTest ⁵
Slovensko	RÚ	RUMobilTest ⁶
Slovinsko	AKOS	AKOS Test Net ⁷

¹<https://www.netztest.at/en/OpenData>

²<https://hakometarplus.hakom.hr/opendata>

³<https://www.netmetr.cz/en/open-data.html>

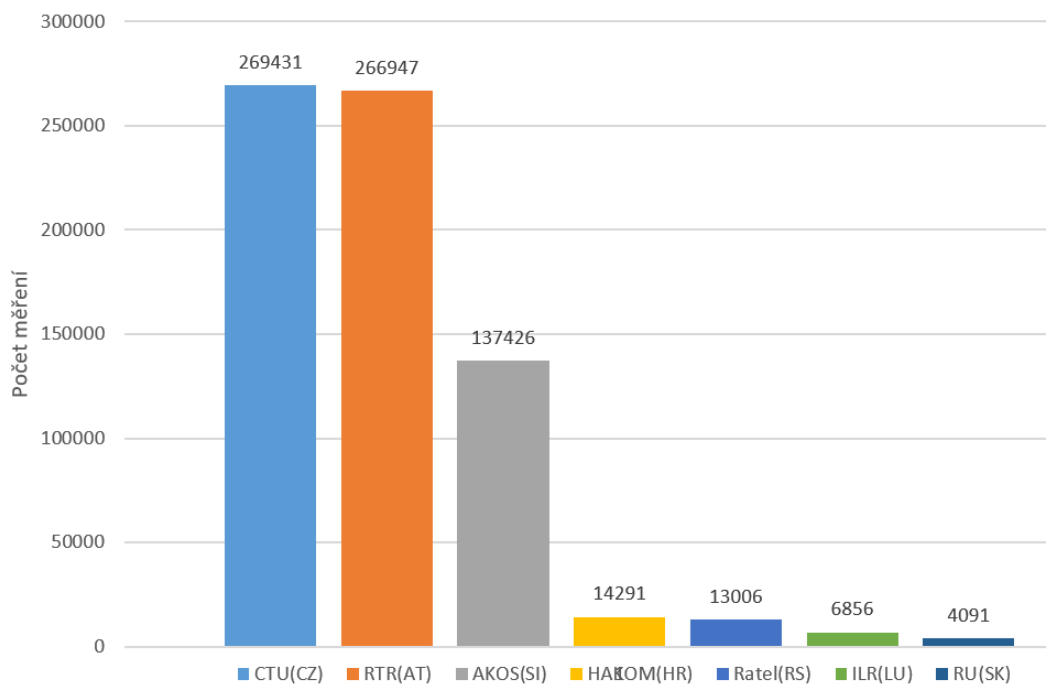
⁴<https://checkmynet.lu/opendata>

⁵<https://nettest.ratel.rs/en/opendata>

⁶<https://www.meracinternetu.sk/en/opendata>

⁷<https://www.akostest.net/en/opendata.html>

Obr. 1.1 znázorňuje četnost měření provedených jednotlivými společnostmi v jednom měsíci roku 2019. Je z nich patrné, že nejvíce měření pochází z českého ČTÚ a Rakouského regulačního úřadu pro vysílání a telekomunikace (RTR). Počet a kvalita dostupných dat jsou klíčové pro zajištění co nejlepších podmínek, aby mohl vzniknout co nejpřesnější odhad. Z tohoto důvodu je kladen důraz výhradně na zpracování datových sad těchto dvou organizací.



Obr. 1.1: Rozložení počtu měření v různých zemích v září 2019.

Obě měřicí aplikace jsou založené na stejném systému, mohou se lišit v parametrech, které zveřejňují. ČTÚ ukládá z měření 47 různých parametrů a u RTR to je až 49 parametrů. V této práci byly vytvořeny 2 skripty, které slouží pro automatizaci stahování dat zmíněných dvou poskytovatelů s nejpočetnějšími databázemi. Byl vybrán referenční vzorek, který odpovídá celkovému počtu získaných dat za období prvních 6 měsíců v roce 2020. Po jejich stažení byly datové sady navzájem porovnány, co se týče počtu měření, a byl vypočítán průměrný počet vzorků na km^2 pro obě země. Pro Rakousko tato hodnota činila $32,545$ měření/ km^2 a pro Česko $24,124$ měření/ km^2 . Na základě těchto výsledků bylo vybráno Rakousko a konkrétně jeho hlavní město Vídeň, které obsahuje největší soustředění měření z celého datasetu.

1.3 Popis parametrů datové sady

V této části jsou představeny nejdůležitější parametry datových sad. Protože máme zájem především o odhad síly signálu LTE, není nutné analyzovat či filtrovat všech

49 různých parametrů. Proto byly vybrány nejdůležitější parametry, které jsou dále popsány:

- **network_type**, určuje typ telekomunikační technologie, pro cíle této práce se volí z nabízených možností LTE;
- **lat**, zeměpisná šířka, udává se s přesností minimálně na 7 desetinných míst;
- **long**, zeměpisná délka, , díky zeměpisné šířce a délce jsme schopni s vysokou přesností určit polohu měření;
- **loc_src**, určuje - jaký je zdroj informací o poloze, např. *gps* nebo *network*. Z vlastního měření bylo zjištěno, že měření se zapnutou *gps*, která určuje polohu, *j* je v databázi označena jako *network*;
- **loc_accuracy**, odhad přesnosti v metrech. Pro *gps* se uvádí hodnota od 4 do 6 metrů u *networku* je nepřesnost větší, a to až do 20 metrů;
- **lte_rsrp**, *RSRP*, síla LTE signálu v dBm. Jedná se o průměrný přijímaný výkon jednoho zdrojového elementu. Rozsah použitelného výkonu se pohybuje od 50 dBm, který se nachází v blízkosti LTE vysílače až k 120 dBm, který je na okraji pokrytí signálu;
- **country_location**, pokud bychom měli k dispozici datovou sadu poskytovatele, který působí i v jiných zemích, je zde nutné zadat tento filtrační parametr – AT pro Rakousko, nebo CZ pro Česko;
- **network_name**, velmi důležitý parametr, určuje poskytovatele datové služby. Musí být vybrán pouze jeden poskytovatel, aby byla nastavená návaznost měření. Byl vybrán poskytovatel Magenta Telekom, který obsadil první příčku v počtu měření, druhý poskytovatel s polovičním počtem byl A1 Telekom Austria a poté Three;
- **lte_rsrq**, Reference Signal Received Quality (RSRQ), kvalita LTE signálu v dBm. Zohledňuje počet použitých zdrojových bloků N , použitých na stejné šířce frekvenčního pásma.

$$RSRQ = (N \cdot RSRP) / RSSI \quad (1.1)$$

Hodnota N se pohybuje v rozmezí od 6 do 100 bloků (20 MHz šířka pásma). Received Signal Strength Indicator (RSSI) měří průměrný celkový přijímaný výkon pouze v Orthogonal Frequency Division Multiplexing (OFDM) symbollech. Pro více informací ohledně architektury LTE, viz [4].

2 Plošná hustota měření

2.1 Příprava dat

Přesná lokalizace změřených bodů je důležitá pro následné odhady pokrytí signálu mezi jednotlivými měřeními. Existuje několik způsobů, jak polohu určit, například určování polohy pomocí internetového prohlížeče, nebo GPS. To, který systém je možné využít, záleží na tom, jestli mobilní zařízení využívají datové připojení LTE a GPS současně, nebo jenom datové připojení. Oba způsoby se liší v odchylce přesnosti v určení polohy klienta.

Měřicí aplikace ukládají údaje o poloze ve formátu zeměpisné šířky a délky ve stupních. Tento souřadnicový systém není vyhovující pro analýzu dat, protože nemá takovou vypovídající hodnotu jako charakterizace dat v metrech a kilometrech. Z tohoto důvodu byly využité algoritmy, co tento přepočítání umožňují. První slouží pro přepočítání konkrétních souřadnic na alternativu v kilometrech a druhý na co nejpřesnější výpočet vzdálenosti mezi jednotlivými body datové sady.

Pro charakteristiku Země v navigačních službách se jako vztažné těleso využívá geoid – fyzikální model povrchu Země, kde je hladina oceánu reprezentována jeho střední hodnotou. Nicméně je tento matematický popis nevhodný pro GPS systémy. Využívá se tedy ve zjednodušené formě jako aproximace prvního stupně – koule, nebo druhého stupně – elipsoid. My používáme geografický referenční systém World Geodetic System 1984 (WGS 84), ten definuje souřadnicový systém a referenční elipsoid pro geodézii a navigaci. Referenční elipsoid obsahuje geodetické datum s počátkem ve středu Země. Druhá část se skládá ze systému zeměpisných souřadnic, tedy zeměpisné šířky a délky ve stupních [5]. Tyto údaje je přínosné si přepočítat na metrické jednotky pro snadnější určení vzdálenosti mezi jednotlivými body, než při ponechání zeměpisné délky a šířky, jako orientačních jednotek.

Pro zjednodušení výpočtu budeme uvažovat, že Země je koule, toto řešení pro účely analýzy bude dostatečné. Případná odchylka bude v celém modelu konzistentní.

Zeměpisná šířka je jednodušší na přepočítání, protože vzdálenost mezi rovníkem a severním pólem se pohybuje okolo 10 000 kilometrů. Podle WGS84 to je 10 001,965729 km [6]. V závislosti na konkrétní poloze, která je vyjádřena ve stupních, je možné přisuzovat hodnotu jednoho stupně od 110,574 km až do 111,694 km. Pro přesnější odhad naší zeměpisné šířky (střední Evropa) byl zvolen přepočítání 111,132 km na 1. stupeň. Celé rozložení je zobrazeno v Tab. 2.1.

Tab. 2.1: Tabulka změny 1. stupně zeměpisné šířky a délky na geodetickém úhlu šířky

ϕ	$\Delta 1 \text{ lat}$	$\Delta 1 \text{ long}$
0°	110.574 km	111.320 km
15°	110.649 km	107.550 km
30°	110.852 km	96.486 km
45°	111.132 km	78.847km
60°	111.412 km	55.8 km
75°	111.618 km	28.902 km
90°	111.694 km	0 km

Jak si lze povšimnout z tabulky, u přepočtu zeměpisné délky je už nutný komplexnější přístup. Jak se vzdalujeme od rovníku, vzdálenosti mezi jednotlivými liniemi se zmenšují a téměř zanikají u pólů. Pokud předpokládáme, že Země je dokonalá koule, pak vzdálenost mezi liniemi délky je úměrná kosinu zeměpisné šířky. Délka rovníku je 40 075 km, je ji nutné vydělit 360, aby se přepočet vztahoval na 1. stupeň. Výsledný vzorec tedy vypadá následovně:

$$40075/360 * \cos(\text{zeměpisná šířka}) \quad (2.1)$$

Tyto vzorce byly implementované v prostředí Google Colaboratory, napsané v jazyku Python. Souřadnice zvolené plochy, ať už se jedná o celou zemi, nebo pouze o město, jsou přepočítané na metry a dále rozdělené do mřížek, které odpovídají nastavení uživatele. Jedná se o mřížky od 10 km² až po 100 m². Všechna data jsou rozdělená do patřičných sektorů a jsou spočítány četnosti, které jsou vyneseny v grafech v části 2.3.

2.2 Haversinův vzorec

Jak už bylo naznačeno v předchozí kapitole 2.1, tento vzorec se používá u výpočtu vzdálenosti mezi jednotlivými body v databázi.

Jedná se o jeden z nejpoužívanějších vzorců při práci se vzdáleností dvou bodů na kouli. Využívá se pro výpočet ortodromy, která určuje nejkratší vzdálenost na sférickém objektu.

Zeměpisná šířka je reprezentována symbolem ϕ a je definována jako úhel mezi přímkou v určitém místě a rovníkovou rovinou.

Zeměpisnou délku znázorňuje úhel, který svírá rovinu místního a nultého poledníku, ten je označený symbolem λ .

Před použitím Haversinova vzorce je nutné nejprve převést souřadnice na radiány. Poté je možné dosadit do vzorce 2.2.

$$a = \sin^2\left(\frac{\Delta\phi}{2}\right) + \cos(\phi_1) \cdot \cos(\phi_2) \cdot \sin^2\left(\frac{\Delta\lambda}{2}\right) \quad (2.2)$$

kde $\Delta\phi$ je rozdíl zeměpisných šířek $\phi_2 - \phi_1$ v radiánech, $\Delta\lambda$ je rozdíl zeměpisných délek $\lambda_2 - \lambda_1$ v radiánech.

Dalším krokem je výpočet průsečíku osy c .

$$c = 2 \cdot \operatorname{atan2}\left(\sqrt{a}, \sqrt{1-a}\right) \quad (2.3)$$

Poslední krokem je výpočet finální vzdálenosti d dvou souřadnic.

$$d = R \cdot c \quad (2.4)$$

kde R označuje poloměr zeměkoule v km (6378.1).

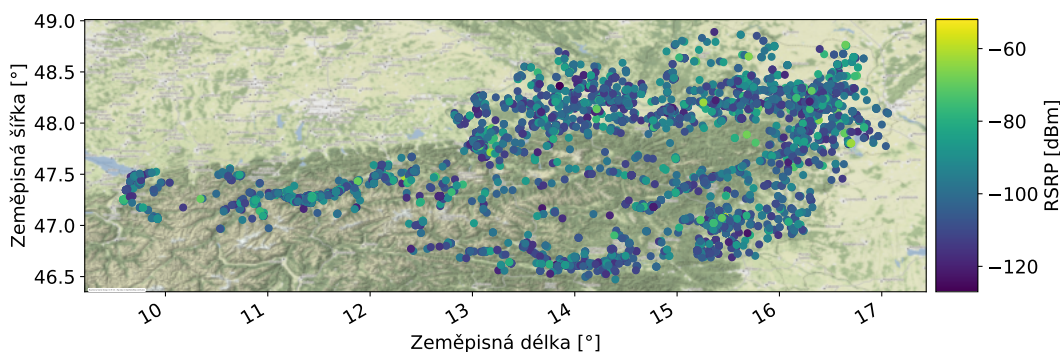
Možná nepřesnost zde opět vzniká v zjednodušeném předpokladu, že Země je koule [7][8].

2.3 Analýza datasetu

V této části se již dostaneme ke konkrétnímu zpracování stažených souborů. Ke strojovému zpracování byl vybrán zmíněný programovací jazyk Python. Jeho předností je jeho univerzálnost, kdy je možné stahovat, zpracovávat a zobrazovat výsledky v jedné aplikaci. Těší se širokému rozšíření a podporou mnoha modulů. Při řešení projektu byla vytvořena sada skriptů, jejichž funkčnost bude popsána v následujících odstavcích. Veškeré skripty jsou dostupné v elektronické příloze této práce.

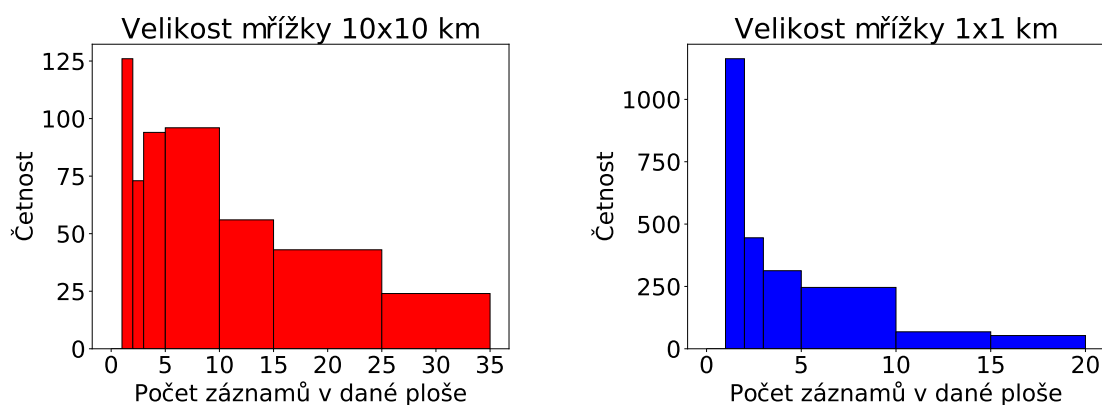
Pro lepší pochopení histogramů v dalších částech práce je na Obr, 2.1 zobrazená mapa Rakouska z vyznačenými změřenými body.

Vybraná data z konkrétní stránky ČTÚ nebo RTR jsou automaticky stažená po konkrétních měsících a letech, jež si zvolíme. Samozřejmě platí, že čím více datasetů stáhneme, tím více máme vzorků, ale zvyšuje se i výpočetní náročnost. Je tedy nutné zvolit vhodný kompromis. Pokud není uvedeno jinak, jsou zvolená data za prvních 6 měsíců z roku 2020. Dále jsou tyto datové sady stažené v zip souborech, ty se extrahují do formátu .csv a postupně se nahrají do dvourozměrného datového rámce, který kromě samotných číselných hodnot obsahuje označené osy. V takto vytvořeném datovém rámci je možné inicializovat jednotlivé indexy a lehce filtrovat



Obr. 2.1: Mapa Rakouska se změřenými body v období leden až červen 2020.

data, která chceme rozlišit. Poté dojde z přepočtu zeměpisné délky a šířky na odpovídající hodnoty v metrech podle vzorců v kapitole 2.1. Jsou vytvořené mřížky podle nastavení. Pro výpočet plošné hustoty pro celý stát jsou nejvíce vypovídající mřížky o ploše 10 km^2 a 1 km^2 . Poté jsou mřížky očíslované a je spočítán počet četnosti jednotlivých měření v každé mřížce. Z těchto dat jsou následně vytvořené histogramy.



Obr. 2.2: Závislost četnosti měření Rakouska pro plochu 10 a 1 km^2 .

Šířky intervalů pro velikost mřížky $10 \times 10 \text{ km}$ jsou rozdělené následovně: 0,1,2,3,5,10,15,25,35. Pro mřížku $1 \times 1 \text{ km}$ se jedná o 0,1,2,3,5,10,15,20.

Z histogramu pro mřížku o ploše 10 km^2 si lze povšimnout, že rozložení intervalových sloupců do četnosti 10 je majoritní. Jedná se o zhruba 70 % všech měření. Protože celá plocha Rakouska obsahuje většinu ploch, ve kterých neprobíhala měření, stačí, že se na této ploše vyskytnou 1 nebo 2 měření, která statisticky vytvoří největší sloupec. Histogramy byly prozkoumané i v rozsahu větších intervalů, než které jsou uvedené na Obr. 2.2. Konkrétně se tam vyskytovala data, která obsahovala i několik set měření na plochu $10 \times 10 \text{ km}$ – tato měření budou představená pro menší plochy.

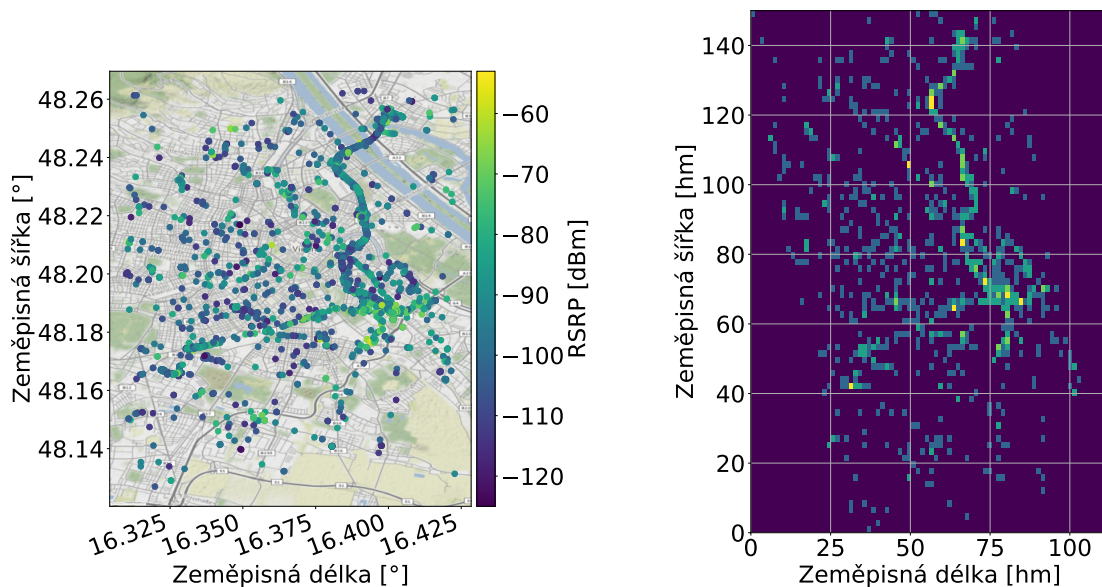
Nyní se předchozí plocha $10 \times 10 \text{ km}$ rozdělí na 100 menších o bloků o plochách

1 x 1 km, graf tedy obsahuje větší četnost vzorků, než v předchozím histogramu. Zde se již projevuje nedostatečná hustota měření, kde se u více než 60 % měření vyskytuje pouze 1 měření na kilometr čtvereční plochy. Jako druhý nejpočetnější sloupec se řadí s 27% dvě měření na danou plochu.

Bylo rovněž provedeno sčítání hustoty vzorků i pro plochy menší, tedy s kvadranty o velikosti 100 x 100 m a 10 x 10 m. Nejsou zde však zobrazené, protože při tak velkém výběru plochy – celého státu nemají žádnou vypovídací hodnotu.

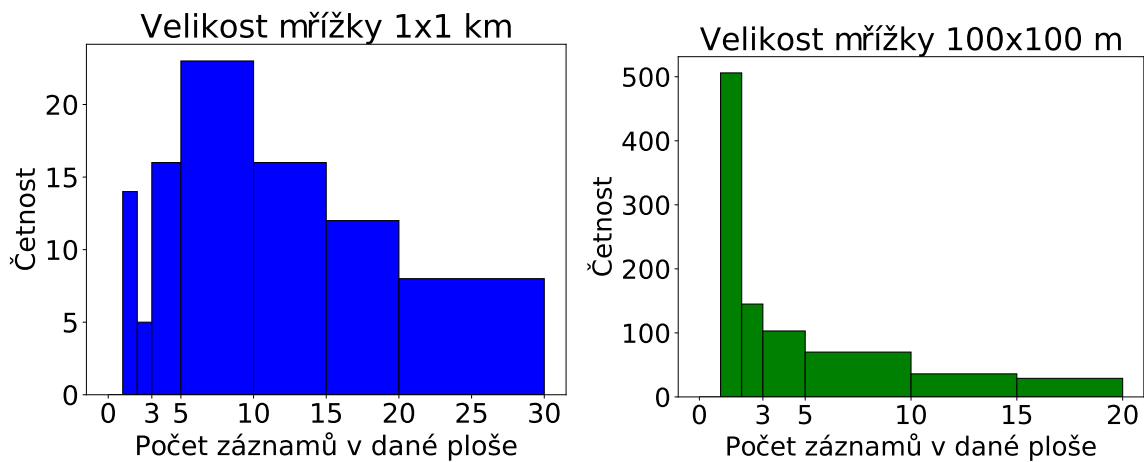
2.4 Plošná hustota ve městě

Protože primárním cílem této práce není určit odhad úrovně signálu LTE v celém státě, bude zkoumaná plocha omezená pouze na jedno město, či jeho část. Stejný rozbor plošné hustoty pomocí histogramu je nyní provedený pro Vídeň. Ta byla vybrána, protože se jedná o relativně hustě obydlené město, kde se vyskytuje větší četnost měření, jež zlepšuje kvalitu odhadu.



Obr. 2.3: Mapy rozložení změřených bodů ve Vídni.

V Obr. 2.3 je nalevo zobrazená mapa Vídne s vyznačenou úrovní *RSRP* jako v předchozím případě s celým Rakouskem. Napravo je zobrazené rozložení počtu měření do dvourozměrné heatmapy. Tato mapa je vytvořená s využitím dat při sčítání počtu měření v mřížce 100 x 100 m, protože poskytuje nejlepší prostorovou rozlišitelnost. Je zde patrná jasná korelace mezi plochou okolo hlavní silnice, kde se vyskytuje nejvíce měření. V heatmapě napravo zobrazují větší četnost jasnější až žluté barvy. Tento graf nám může lépe přiblížit, jakou hustotu můžeme v určitých částech čekat.



Obr. 2.4: Histogram hustoty měření pro mřížku 1 x 1 km a 100 x 100 m.

V histogramu mřížky 1 x 1 km je nejvíce měření v sloupci obsahujícím od 5 do 10 měření. Medián takto změřených hodnot se pohybuje okolo 10 měření na kilometr čtvereční. Na základě těchto dat je v další kapitole navrženo vlastní měření, které obsahuje dostatek bodů, aby byla zajištěná možnost změny prostorového rozlišení. Vytvoření vhodného predikčního modelu na základě těchto dat je ověřeno v dalších kapitolách. Šířky intervalů jsou rozdělené podle následujících hodnot: 0, 3, 5, 10, 15, 20, 25, 30.

Velikost mřížky 100 m² se ukázala být jako příliš jemná pro náhodné rozložení dat. Její využití by mělo větší smysl pro kontrolovaný projekt, kde by bylo možné provádět měření na předem určené ploše. Šířky intervalu v Obr. 2.4 jsou 0,3,5,10,15,20.

3 Vlastní měřicí kampaň

V kapitole 2.4 byla představená plošná hustota pro Vídeň i s mediánem 9 měření na km². Z výsledků je patrné nerovnoměrné rozložení pro celou plochu. Vyskytují se tam často místa, která obsahují pouze minimální počet známých bodů. Taktéž se zde vyskytují lokální maxima obsahující stovky bodů, které byly nashromážděné v průběhu 6 měsíců na malé ploše. Pro zpracování měřených dat je vhodné vytvořit menší datovou sadu umožňující relativně rychlé průběhy simulací pro odhad síly signálu *RSRP* v místech kde měření chybí. Vlastní měření má za cíl poskytnout takové datové sady, které odpovídají reálné hustotě měření a musí rovněž obsahovat dostatečný počet vzorků, aby byla zajištěná možnost škálování prostorového rozlišení. Pomocí vlastního měření jsme schopni ovlivnit parametry, které není možné zjistit z datových sad. U poskytnutých vzorků není možné dohledat, jestli měření probíhalo v budově, nebo ve venkovním prostředí.

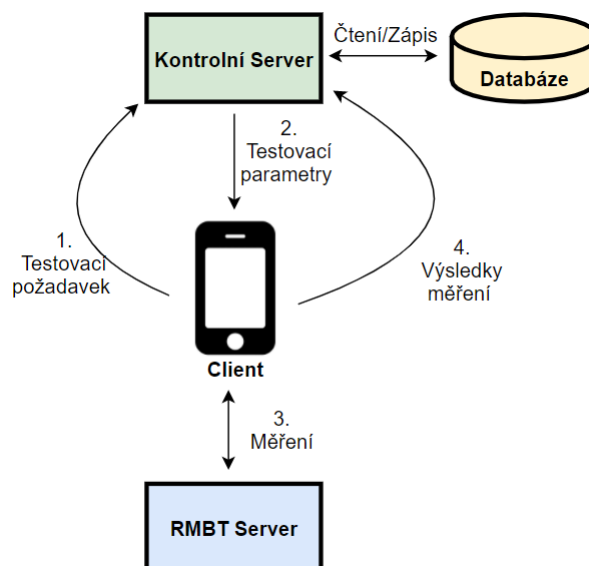
3.1 Měřicí zařízení a ukládání výsledků

K měření byl využíván školní telefon, který byl použitý v [9]. Jedná se o telefon značky Samsung model A10. Ten byl vybrán na základě rozsáhlé studie, ve které dosahoval nejlepší kombinace ceny, verze operačního systému Android a výkonu antén pro příjem a vysílání.

Aby bylo možné dosáhnout přesného měření maximální propustnosti, využívá se vícenásobné paralelní přenášení datových toků přes TCP připojení po přesně určený čas. Data určená pro tuto zatěžovací zkoušku jsou náhodně vygenerovaná s vysokou úrovní entropie. Jakákoliv komprese dat během přenosu je zakázána. Protože při podobných testech může dojít k zablokování dat například bránami firewall a proxy server, jsou data přenášená prostřednictvím Transport Layer Security (TLS) připojení [2].

Na obrázku 3.1 je znázorněno, jakým způsobem probíhá testování. Nejprve se uživatel připojí pomocí přiděleného rozlišovacího kódu na kontrolní server –linuxová distribuce. Instalace je nejspolehlivější na operačních systémech založených na Debianu 9 a novějších. Podle přiděleného rozlišovacího kódu zařízení dojde k rozpoznání uživatele. Tento kód je generuje s kratší délkou, aby nepřidával zpoždění do celkového systému. V druhém kroku pošle kontrolní server nastavení ohledně parametrů měření, jež obsahují například výběr konkrétního dostupného RMBT serveru a přesný čas začátku měření. Začátek je určen podle aktuální vytíženosti serveru, zpoždění je zpravidla zanedbatelné. Jakmile se otestují všechny aspekty spojení, dojde k odeslání výsledků na kontrolní server a k uložení do databáze [2], k níž je potom možné

se připojit z prohlížeče a po zadání synchronizačního kódu z aplikace v mobilním zařízení je možné stáhnout všechna data ve formátu .csv. Pokud by byl nutný sběr většího počtu měření, stojí za zvážení vytvoření vlastního měřicího serveru. Poskytovatelé založení na Open-RMBT systému jsou schopni detekovat zvýšený počet probíhajících měření z jednoho zařízení a tomuto zařízení ubírat prostředky, či mu nastavit Timeout v připojení.



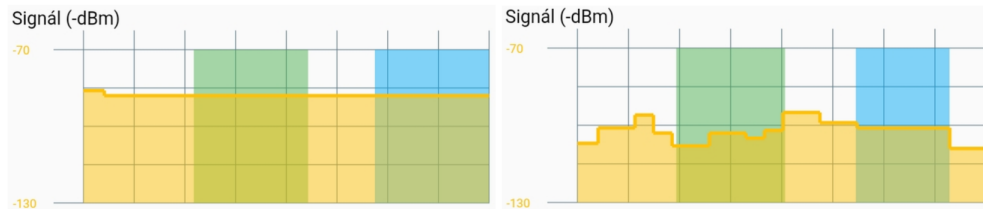
Obr. 3.1: Přehled měřicího systému, na základě [2].

3.2 Měření v terénu

Medián měření na km^2 pro reálná data odpovídá 9. Mapa Brna byla analyzována a byla vybrána taková část, která přibližně odpovídá této ploše. V této kapitole je popsáno získávání reálných vzorků z městské části Brno-Ponava a v okolí parku Lužánky. Tyto lokality byly zvolené, protože obsahují jak otevřená prostranství, tak i poměrně symetrické rozložení, co se týče prolínání bytové zástavby a cest. Nejefektivnější aproximace lze dosáhnout z datových bodů, které jsou rozmístěny v rovnoměrné mřížce. V dostupných datových sadách jsou měřené body rozmístěny náhodně a většinou daleko od sebe. Často se vyskytuje i pravý opak, kdy jsou lokality, ve kterých se často testuje rychlost připojení a je tam těchto vzorků přebytek. Obě tyto možnosti představují zanesení značné nepřesnosti k metodě odhadu. Při provádění měření v terénu byl kladen důraz na tyto parametry.

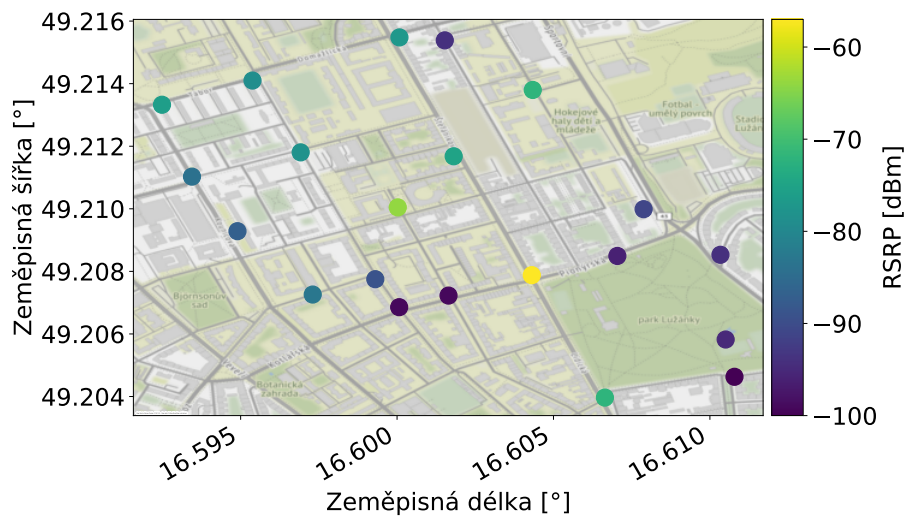
Jedno kompletní měření trvá v rozmezí od 40 do 50 vteřin. Tato doba se liší od konkrétního nastavení testu a doby pro dosažení maximální rychlosti stahování a nahrávání. Telefon byl v této době držen rovnoběžně se zemí, aby poloha telefonu

odpovídala nejčastějšímu používání pro prohlížení internetového obsahu. Bylo ověřeno, jak se liší hodnota $RSRP$, pokud se manipuluje s telefonem okolo vlastní osy. Tento jev je patrný na Obr. 3.2, kde nalevo je znázorněn časový vývoj $RSRP$, pokud je telefon držen ve stabilní poloze. Napravo bylo s telefonem pohybováno kolem vlastní osy.



Obr. 3.2: Závislost měřené úrovně $RSRP$ na čase pro stabilní držení telefonu (nalevo), s telefonem je manipulováno (napravo).

Na Obr. 3.3 jsou znázorněny konkrétní měření v ulicích Brna. Tmavší barvy - purpurová a modrá označují slabou kvalitu signálu $RSRP$, jasně zelená až žlutá označují velmi dobrý příjem.

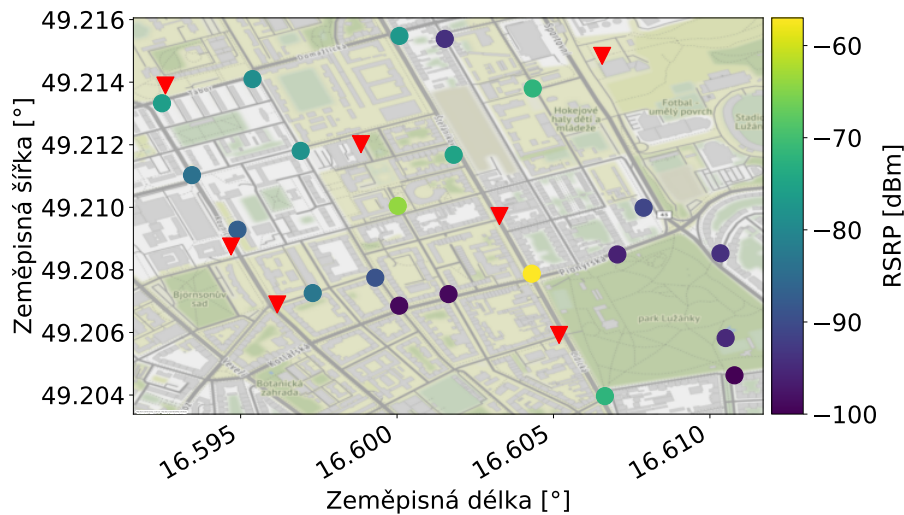


Obr. 3.3: Mapa pilotního měření v městské části Brno-Ponava.

Úrovně síly signálu LTE se pohybovaly od -100 dBm až do -57 dBm v síti T-Mobile. Průměrná hodnota $RSRP$ se pohybuje okolo -82 dBm. Pro porovnání s měřeními uskutečněnými ve Vídni je tato hodnota o 18 dB vyšší, dá se tedy usuzovat, že měřená část Brna má lepší pokrytí než Vídeň. Z měření je patrné, že nejslabší výkony LTE byly naměřené ve dvou lokalitách. První je park Lužánky, kde je patrný útlum způsobený stromy a větší vzdáleností od vysílačů. Přijímaný výkon se zde pohyboval od 100 dBm do 90 dBm. Druhou typovou lokalitou je území, kde

se nachází hustá zástavba s úzkými ulicemi – jedná se o dva body na ulici Kotlářská (levá strana mapy dolní sektor). V ostatních lokalitách, kde je více otevřeného prostoru, byla úroveň signálu *RSRP* naměřená ve výborné (více jak -70 dBm) nebo dobré (-70 až -85 dBm) kvalitě.

Určování polohy fungovalo velmi přesně, výjimkou jednoho měření, kde byl výsledný bod zobrazen s odchylkou přibližně 50 metrů. Tato odchylka mohla být způsobena odrazem GPS signálu družice od okolních stromů.



Obr. 3.4: Mapa pilotního měření v městské části Brno-Ponava s LTE vysílači.

Experimentálně byly vyhledány LTE vysílače v dané lokalitě. K tomu posloužil gsmweb [10], který obsahuje přesné polohové informace i s fotografiemi Base Transceiver Station (BTS). Ty jsou znázorněny jako červené trojúhelníky v Obr. 3.4. Jedná se pouze o orientační zobrazení, pro lepší představu jak jednotlivé úrovně síly signálu závisí na vzdálenosti od vysílače. Jak je patrné z Obr. 3.4, neplatí, že čím menší je vzdálenost mezi uživatelem a BTS, tím je signál silnější. To je dáno především vyzařovací charakteristikou těchto antén a jejich nasměrováním a odrazem signálu od budov.

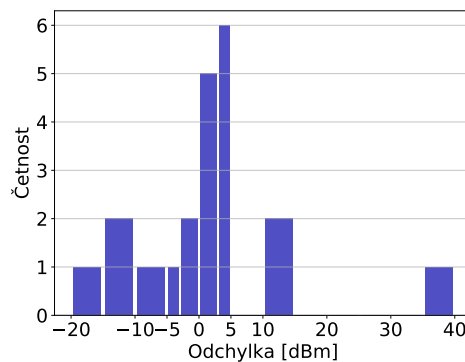
3.2.1 Časový vývoj

Po 4 měsících bylo provedeno druhé měření pro ověření časové závislosti/přesnosti prvního měření. Sběr dat probíhal ve stejný den v týdnu (pátek), ve stejné době (13:00 až 15:00) a na totožných místech (odchylka se pohybuje v jednotkách metrů). Toto období je relativně dlouhé a je možné, že došlo k úpravě parametrů sítě. Je nutné brát tuto skutečnost v úvahu při porovnání větších výkyvů úrovně *RSRP* u některých měření. Soubor je k dispozici v příloze: *Brno_Luzanky2.xls*.

Výsledky jsou prezentovány pomocí histogramu, který znázorňuje rozdíl hodnot změřených v prvním měření v porovnání s měřením druhým. Hodnoty v záporné části grafu značí, že při druhém měření byl naměřen silnější signál. Tomu odpovídá 33 % měření. Mnohem početnější je kladná část odchylky, která značí, že byl naměřený slabší signál v druhém měření.

Největší četnost odchylky se pohybuje v rozmezí od 5 do 5 dBm. Zde se nachází 66 % změřených vzorků. Pokud tento interval ještě rozšíříme od 10 do 10, zjistíme, že v této části se nachází 85 % všech hodnot. Rovněž byly změřené hodnoty, které se významně lišily od předchozího měření (rozdíl 20 a 39 dBm), ty tvořily okolo 10 % všech vzorků.

Pomocí druhého sběru dat bylo zjištěno, že hodnoty v první fázi měly obecně vyšší úroveň signálu. Je také velká šance odchylky měření v rozmezí 5 dBm *RSRP*. Tato úroveň může být použita jako referenční pro odhad síly signálu, kdy i interpolační metoda blízká se přesnosti 5 dBm úrovně *RSRP* může podávat relativně spolehlivé výsledky v reálném prostředí. Intervaly histogramu mají následující šířky: 20, 15, 10, 5, 3, 0, 3, 5, 10, 15, 20, 25, 30, 35, 40.



Obr. 3.5: Porovnání rozdílů prvního a druhého měření v časovém období 4 měsíců.

4 Regrese

Regresi využíváme k nalezení funkce, která nejlépe popisuje vztah mezi dvěma skupinami proměnných. Jako zjednodušený příklad využití této metody může sloužit odhad ceny nemovitosti na základě lokality, počtu pokojů, stáří atd.

Skupina nezávislých proměnných, které vysvětlují náhodnou proměnnou, se nazývá regresory či vysvětlující proměnné. Ve zmíněném příkladě budovy by se jednalo o její vlastnosti.

Závislá proměnná, kterou chceme odhadnout pomocí regrese se označuje jako odezva nebo vysvětlovaná proměnná. V motivačním příkladě se jedná o cenu nemovitosti.

Jednoduchá regrese pracuje s jedním regresorem, který určuje vysvětlovanou proměnnou. U vícenásobné regrese je využito regresorů více.

Pro určení nejlepší kombinace regresorů pro odhad odezvy se využívá metoda nejmenších čtverců, kde nejmenší odchylka označuje nejlepší kombinaci. Poté je možné vytvořit model zkoumané veličiny Y , kterou chceme definovat pomocí kombinace vstupních parametrů X [11].

Na začátku regresivní či interpolační analýzy se doporučuje rozdělit dostupné datové sady na dvě skupiny: trénovací a testovací.

Trénovací datový set se využije na vytvoření modelu, podle něhož zjistíme nejlepší přístup k řešení.

Testovací hodnoty se poté vloží do vytvořeného modelu a je srovnána odchylka mezi pravdivými hodnotami testovací sady a hodnotami, které byly odhadnuté trénovacím modelem.

Je možné zvolit různé poměry trénovacích a testovacích dat, obvykle se využívá $> 80 \%$ pro trénovací sekvence. Pro větší úspěšnost je možné tento poměr zvyšovat.

Takto vzniklý model lze použít k předpovědi závislé proměnné. Potom lze predikované hodnoty porovnat s původními a určit jejich přesnost, případně odchylku, střední hodnotu nebo rozptyl. Lze použít různé přesnosti měření jako R-kvadrát, střední průměrnou chybu, Pearsonův korelační koeficient a další [12].

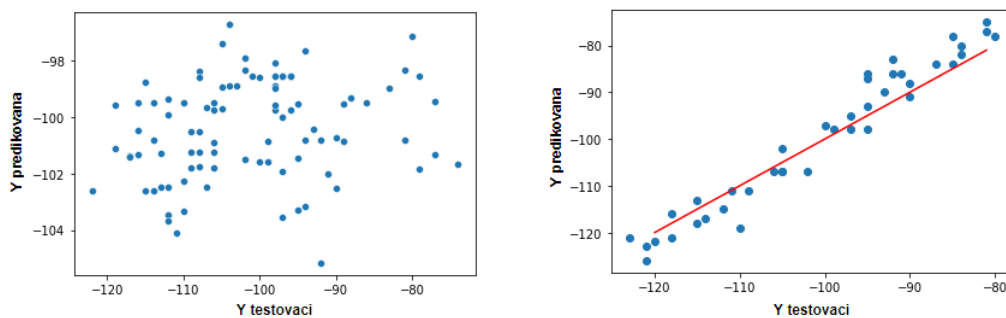
4.1 Druhy regrese

Metody regresivní analýzy lze rozdělit na 3 druhy:

- Lineární regrese;
- Víceprvková lineární regrese;
- Nelineární Regrese.

Lineární regrese je užitečná pro predikování vzájemné závislosti dvou a více proměnných za předpokladu, že mezi těmito veličinami existuje matematická závislost. Použití lineární regrese na datové sady není možné. Regresory by v tomto případě musely být jednotlivé souřadnice a ty mezi sebou nevykazují žádnou korelaci [13].

Z experimentálního hlediska byla tato regresní metoda natrénovaná na datech z Vídně. Na Obr. 4.1 je na levé straně znázorněný výstup použité lineární regrese na data zeměpisné šířky a délky se silou signálu *RSRP*. Na ose x se nachází výsledky testovací sekvence Y , jedná se o data, která byla získána z měření (nebyly použité v trénovací sekvenci). Ose y reprezentují predikované hodnoty *RSRP*. Z grafu lze vyčíst, že jednotlivé předpovědi nebyly příliš přesné a data jsou náhodně rozházené po celé zkoumané oblasti.



Obr. 4.1: Graf nevhodného použití lineární regrese (nalevo) v porovnání se správným výsledkem (napravo).

Na pravé straně se nachází náhodně zvolená data, které jsou následně aproximované přímkou metodou nejmenších čtverců. Tento graf dokazuje, jak by mělo vypadat rozložení bodů, aby se jednalo o lineární regresi. Takto vytvořenou přímku je možné popsat podle následující rovnice 4.1:

$$Y = a + bX + \epsilon \quad (4.1)$$

kde Y je závislá proměnná, X je nezávislá proměnná, a je konstantní koeficient, který je nutné dopočítat - udává posunutí přímky po ose y , b je směrnice přímky a ϵ je chybový člen.

Protože byly použité dvě hodnoty jako vstupní proměnné, jedná se o víceprvkovou lineární regresi. Tyto popisující parametry musí splňovat minimální korelaci mezi sebou. Tento popis odpovídá situaci v levém grafu Obr. 4.1.

Nelineární regrese, jak již název napovídá, je jakýkoliv model regrese, který nespĺňuje pravidla pro lineární model. Výsledná funkce zde má více možností/tvarů, které mohou popisovat stejné body. Také nelineární rovnice nejsou složené pouze

z násobení a sčítání. Nicméně ani tato obměna nezajišťuje použitelnost regrese na náhodná data, která obsahují datové sady [14].

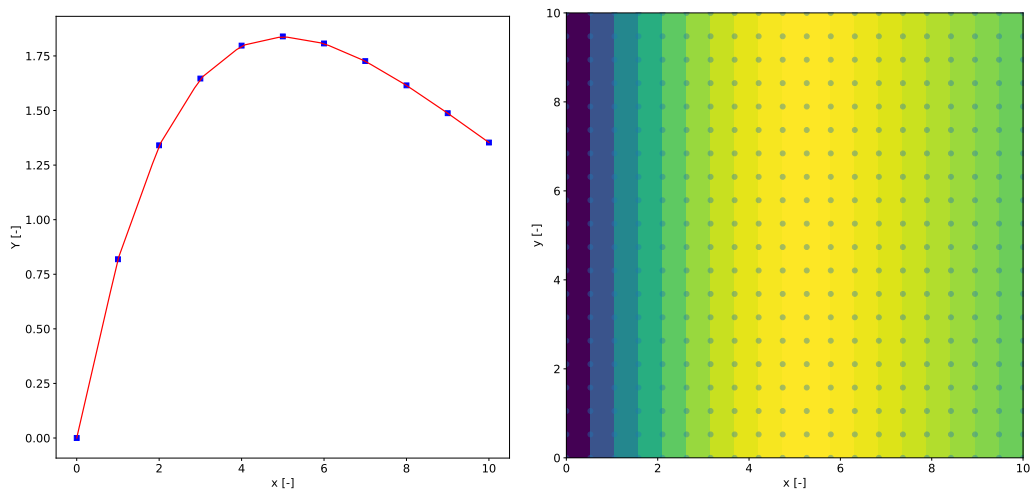
Po bližší analýze regresních metod bylo rozhodnuto, že tento způsob odhadu parametru *RSRP* není vhodný pro tuto práci. Nicméně poznatky z kapitoly o rozdělení datové sady na trénovací a testovací sekvence jsou využity v následujících kapitolách.

5 Interpolace

Interpolace se od regrese liší v tom, že nehledá přímkou, případně polynom, který popisuje nejmenší vzdálenost metodou nejmenších čtverců. Namísto toho hledá funkci, která je schopná procházet přes všechny body ve výběru. Ta nám umožňuje odhadnout hodnoty, které se nacházejí na zmíněné funkci a jež nebyly obsažené ve vstupních proměnných. Lépe je tato skutečnost zobrazená na Obr. 5.1, Tyto 2 grafy znázorňují stejnou funkci odpovídající rovnici 5.1.

$$Y = e^{-x/5} \cdot x \quad (5.1)$$

V obou případech jsou vstupní data znázorněna modrými body a zvolená interpolace je lineární. V prvním grafu tuto interpolaci značí červená křivka. U druhého případu je to barevný rozsah, který se mění od nejnižších hodnot Y vyobrazených fialovou barvou přes zelenou až po žlutou odpovídající nejvyšším hodnotám, jichž křivka může nabývat.



Obr. 5.1: Porovnání jednorozměrné (nalevo) a dvourozměrné interpolace (napravo).

Protože uvažujeme prostorovou interpolaci, je nutné zvolit vhodnou dvourozměrnou interpolační metodu. Velmi častým interpolačním principem je využití prostorové autokorelace. Ta určuje, že hodnoty v bližších lokalitách budou mít větší váhu pro odhad než ty vzdálenější [15].

Použití vhodného algoritmu záleží na mnoha faktorech, především na účelu, objemu vstupních dat nebo požadované přesnosti. Protože změřená data se vykreslují ve dvourozměrném prostoru, je nutné zvolit funkci, která je umí zpracovat. Algoritmy lze rozdělit do 2 skupin.

- Přesné (Deterministické) interpolační metody

- **Vyrovňovací (Stochastické) interpolační metody**

U deterministických algoritmů, jak už název napovídá, se při interpolaci počítá s maximální vahou polohy, tedy 1.0. Mezi tyto metody se řadí splajny (bilineární a bikubická interpolace), IDW bez vyrovňovacího faktoru, Kriging, Thiessenovy polygony a Nearest Neighbor.

U vyrovňovacích (stochastických) algoritmů má vždy hodnota datových bodů hodnotu menší než 1.0, vzniká tak jemnější spojení mezi jednotlivými body a větší vyhlazení lokálních nerovností. K těmto metodám patří minimum curvature, Kriging s nugget efektem, IDW s vyrovňovacím faktorem a polynomická [16].

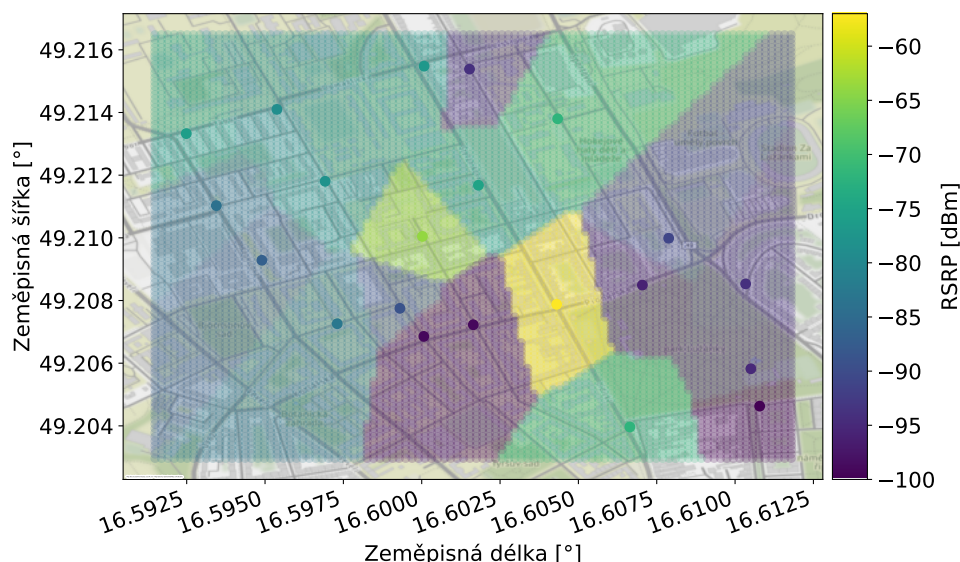
V dalších částech práce jsou nejvhodnější zmíněné metody více vysvětlené a budou použité na reálných datech znázorněných na Obr. 3.3.

5.1 Metoda nejbližšího souseda

Jedná se o jednu z nejjednodušších interpolačních metod. Princip spočívá v nalezení bodů z množiny všech subjektů, které mají nejmenší vzdálenosti od námi sledovaného vzorku. Sledovaný objekt se zařadí k několika nejbližším sousedům. Při této prostorové analýze vznikají Thiessenovy polygony, tyto obrazce je možné vidět na Obr. 5.2. Každý tento polygon reprezentuje jednu změřenou hodnotu. Nevýhodou této metody jsou okrajové polygony, jež mají teoreticky nekonečnou plochu a je nutné určit hranice zkoumaného území. Thiessenovy polygony slouží jako základ dalších interpolačních metod. Další nevýhodou této metody je nulová citlivost na hodnoty, které se nacházejí ve větší vzdálenosti od sledovaného subjektu a samotná metoda není vhodná pro analýzu spojitě se měnících jevů. Tato metoda se nejčastěji využívá v meteorologii a klimatologii, kde meteorologické stanice pokrývají pouze určitou oblast, v níž mají nejspolehlivější výsledky [17]. K vytvoření této a dalších predikcí byla konkrétně využita knihovna pro vědecké a technické výpočty *SciPy*. Pro vytvoření interpolace pomocí metody nejbližšího souseda byla konkrétně využita třída `scipy.interpolate.NearestNDInterpolator`¹. Tato třída dokáže rychle najít nejbližší sousedy libovolného bodu v N-rozměrovém prostoru, jehož rozměry jsou specifikované na vstupu funkce.

Z výsledné interpolace na Obr. 5.2 je patrné, že Thiessenovy polygony jsou vytvořené pomocí 4 nejbližších bodů. Hodnota změřeného bodu predikuje stejnou hodnotu až do poloviny vzdálenosti k dalšímu bodu, tato operace se opakuje čtyřikrát pro každý bod.

¹<https://docs.scipy.org/doc/scipy/reference/generated/scipy.interpolate.NearestNDInterpolator.html>



Obr. 5.2: Metoda nejbližšího souseda.

5.2 Bilineární interpolace

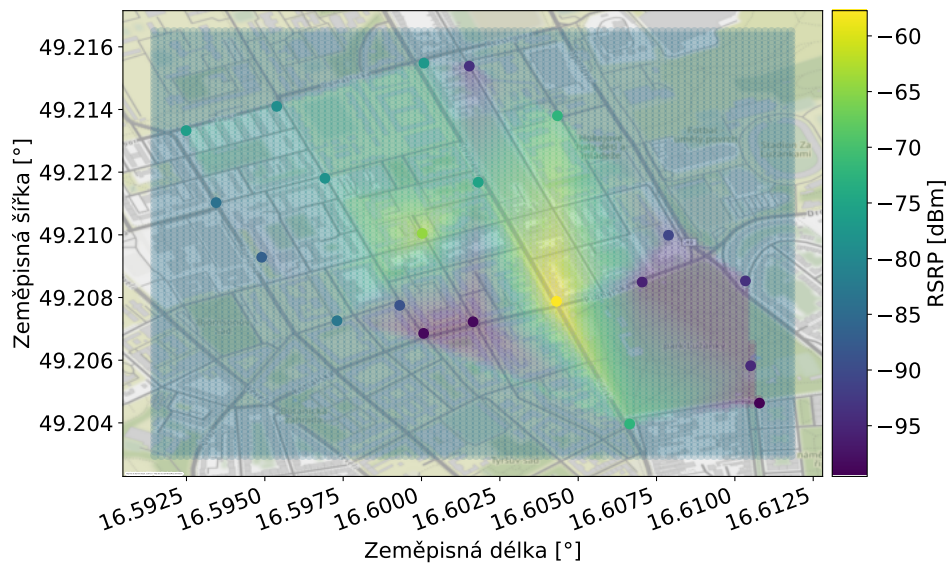
Použití lineární interpolace ve dvourozměrném prostoru označujeme jako bilineární interpolaci. Odhad neznámé hodnoty, která se vyskytuje mezi dvěma známými vzorky, se vypočte pomocí postupné lineární interpolace osy x a poté osy y . Příklad výpočtu hodnoty bodu P se souřadnicemi x a y je popsán v rovnici: 5.2.

$$\begin{aligned}
 P = & \frac{(x_2 - x)(y_2 - y)}{(x_2 - x_1)(y_2 - y_1)} Q_{11} + \frac{(x - x_1)(y_2 - y)}{(x_2 - x_1)(y_2 - y_1)} Q_{21} \\
 & + \frac{(x_2 - x)(y - y_1)}{(x_2 - x_1)(y_2 - y_1)} Q_{12} + \frac{(x - x_1)(y - y_1)}{(x_2 - x_1)(y_2 - y_1)} Q_{22}
 \end{aligned} \tag{5.2}$$

kde x_1, y_1, x_2, y_2 jsou souřadnice nejbližších bodů, které budou použité k interpolaci, $Q_{11}, Q_{21}, Q_{12}, Q_{22}$ jsou příslušné hodnoty těchto bodů. Výsledné grafické znázornění této metody vyobrazuje Obr. 5.3.

Zde je na první pohled patrné, že plocha je značně omezená oproti metodě nejbližšího souseda, která pokrývala celou oblast. To je zapříčiněno tím, že interpolace vzniká pouze mezi známými body. Predikovaná plocha tedy je omezená přímkami krajních bodů. Okolní ploše byla přiřazena průměrná hodnota všech vzorků v databázi, aby nedocházelo k prázdným odhadům/chybám při testování. Je vhodné si povšimnout viditelnějších přímek, které spojují body s více rozdílnými úrovněmi *RSRP*.

Stejně jako v předchozí metodě, je i zde využita knihovna *SciPy*. Vstupní data jsou nejprve triangulována pomocí metody *Qhull* ². Poté je na každém trojúhelníku provedena lineární barycentrická interpolace. Barycentrická interpolace používá tři



Obr. 5.3: Bilineární interpolace.

nejbližší body - vrcholy trojúhelníku a podle jejich vzdálenosti je vypočítaná relativní váha na odhadovaný bod vně trojúhelníku [18].

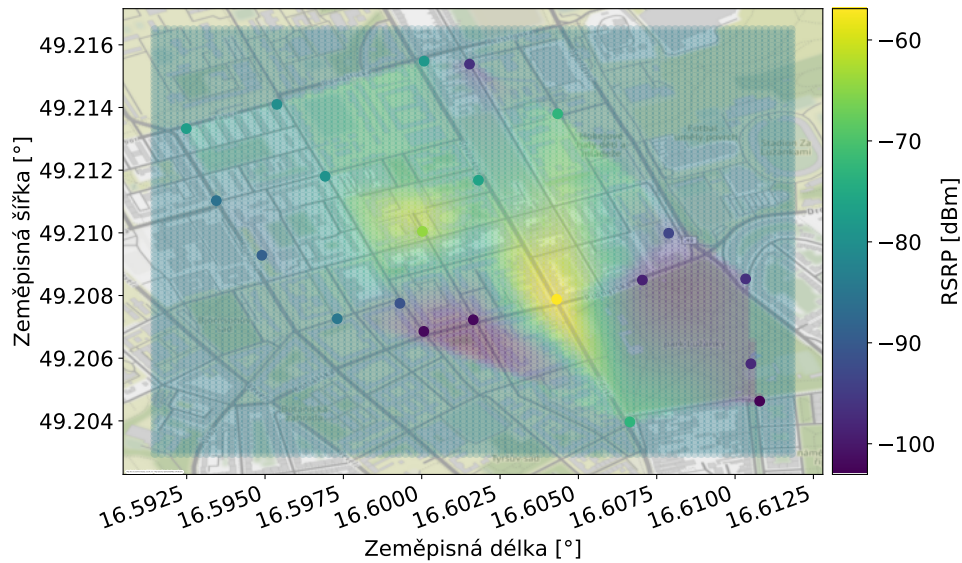
5.3 Bikubická interpolace

Výpočet se provádí pomocí kubického polynomu, ten představuje hlavní rozdíl od bilineární interpolace, jež používala přímku. Jedná se o polynom třetího stupně, který lépe dokáže vyhladit odhady bilineární interpolace, proto se jedná o nejpoužívanější volbu, která obsahuje méně interpolačních artefaktů [19].

I zde dochází k omezení predikované plochy pouze ke známým hraničním bodům. Patrnější rozdíl mezi lineárním a kubickým modelem je v úrovni vyhlazení, kde kubický model dosahuje lepších výsledků. U větších datových sad je nutné mít na paměti, že lepší vyhlazení s sebou nese i větší složitost na výpočet.

Postup je podobný jako u bilineární interpolace. Opět jsou vstupní data triangulovaná pomocí Qhull. Každý trojúhelník je pomocí Clough-Tocherové metody rozdělený na 3 menší. Pro každý tento menší trojúhelník je vytvořena Bézierova křivka. Tato parametrická křivka umožňuje vytvoření dostatečně interaktivního nastavení, které vytvoří vyhlazenější vyobrazení odhadu [20].

²<http://www.qhull.org>



Obr. 5.4: Bikubická interpolace.

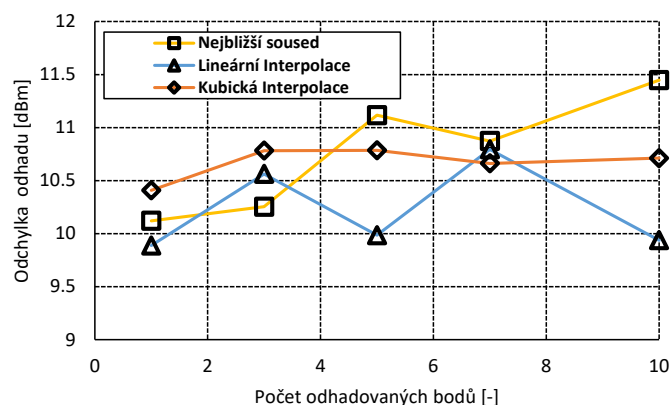
5.4 Testovací metodika

V minulých kapitolách byly popsány interpolační metody k dosažení co nejlepší predikce. Jak ovšem víme, která je nejlepší, a podle čeho se to bude posuzovat? K tomuto účelu byl vytvořený skript, který náhodně vybere známý bod či více bodů a vyřadí je ze skupiny vzorků, jež jsou určeny pro vytvoření modelu. U menších datových sad, jako je např. *Brno_Luzanky*, je přijatelnější využívat přesný počet odstraněných bodů. Kdyby se zde používalo procentní rozdělení, může se stát, že nebude odstraněn žádný bod a dojde k chybě v programu. U sad, které obsahují stovky až tisíce měření, se používá určité procento odstraněných známých vzorků. Dá se předpokládat, že čím méně bodů bude použito pro stanovení odhadu neznámého vzorku, tím nepřesnější tento odhad bude. Počet odstraněných bodů pro datovou sadu z Brna byl zvolený tak, aby odpovídal plošné hustotě měření získané z kapitoly 2.4. Z 21 známých bodů tedy odstraníme maximálně polovinu (10 vzorků), tím se dostaneme k horní hranici odpovídající četnosti reálných dat. Zbylé vzorky jsou určeny k vytvoření modelu predikce, původní souřadnice jsou následně vloženy do tohoto modelu a je vypočítán odhad úrovně *RSRP*. Tato predikce je porovná s původní hodnotou změřenou v terénu. V absolutních hodnotách se sečtou všechny rozdíly a vydělí se počtem predikovaných vzorků. Tím získáme výchylku, která charakterizuje chybovost jednoho měření. Pro přesnější predikci a eliminaci náhodného vlivu malého počtu testů je tato metoda použita 100krát a poté je znovu zprůměrována počtem opakování. V každém průběhu programu je vypočtena střední hodnota, odchylka a rozptyl odhadu pro určení charakteru dat. Všechna takto získaná data

jsou uložena do textových souborů, které jsou součástí příloh.

5.5 Porovnání deterministických metod

V této části práce jsou porovnány dosavadní výsledky interpolačních metod: nejbližšího souseda, lineární a kubické interpolace. Obr. 5.6 znázorňuje rozdíl naměřené a odhadované hodnoty *RSRP* na svislé ose a počtu odstraněných/odhadovaných bodů na vodorovné ose. Odchylka odhadu se pohybovala nejčastěji mezi 10 a 11 dBm. Celkově nejlepších výsledků dosahovala lineární interpolace, která má v součtu nejmenší odchylku napříč různými počty vstupních bodů. Avšak je patrné, že nezáleží na počtu odstraněných bodů a rozložení vypadá nahodile, kdy nezáleží na tom, jestli je k odhadu použito 20 bodů, nebo 11. Odchylka je v obou případech těsně pod 10 dBm. Kubická interpolace v tomto ohledu podává více konzistentní výsledky, avšak s horším celkovým odhadem o 0.4 dBm od lineární interpolace. Metoda nejbližších sousedů přímo souvisí se sníženým počtem vstupních hodnot. Zde se projevuje nevýhoda metody, kdy pro celou oblast Thiessenova polygonu je přiřazena jedna hodnota nejbližšího bodu. Se snižujícím se počtem polygonů se zvyšuje i pravděpodobnost větší nepřesnosti.



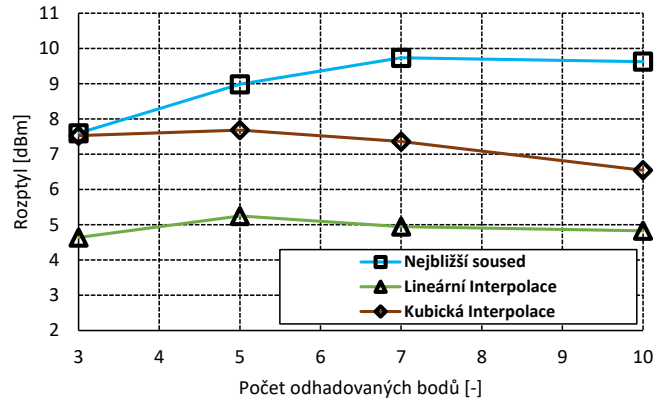
Obr. 5.5: Porovnání odchylek odhadu a rozptylu deterministických metod.

Dalším parametrem, jenž nám pomůže stanovit charakter predikovaných dat je rozptyl. Jeho analýza (viz rovnice 5.4) je založená na porovnání predikované hodnoty *RSRP* a střední hodnoty (viz rovnice 5.3). Pro případ, kdy je dopočítáván pouze jeden odstraněný bod, je rozptyl rovný 0, proto není v Obr. 5.6 uveden.

$$\bar{X} = \frac{\sum X_i}{n} \quad (5.3)$$

$$\sqrt{\sigma} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (5.4)$$

kde \bar{X} je střední hodnota, X_i jsou jednotlivé hodnoty *RSRP* určené k odhadu, n je počet predikovaných bodů, σ označuje rozptyl.



Obr. 5.6: Porovnání rozptylu interpolačních metod v závislosti na počtu odstraněných bodů.

Z grafu 5.6 vyplývá, že nejmenší rozptyl má lineární interpolace, která měla i nejmenší odchylku. Vykazuje rovněž konstantní hodnotu, co se týče snižování počtu vstupních bodů. Jeví se tedy jako nejlepší možnost k odhadu síly signálu z deterministických metod. Kubická interpolace nabízí opět více konzistentní výsledky, které mají tendenci se zlepšovat s menším počtem vstupních informací. Tato metoda se může zdát také jako použitelná k odhadu z datových sad, avšak je nutné brát v úvahu, že výpočet bude $4\times$ náročnější vzhledem k využití kubického splajnu. U datových sad, které by obsahovaly několik stovek či tisíc vzorků, klesá úroveň využití kubické interpolace. Výsledek odchylky a rozptylu jsou na sobě přímo závislé a metoda nejmenších čtverců má i u tohoto porovnání nejhorší výsledky.

5.6 Inverzní vážení vzdálenosti

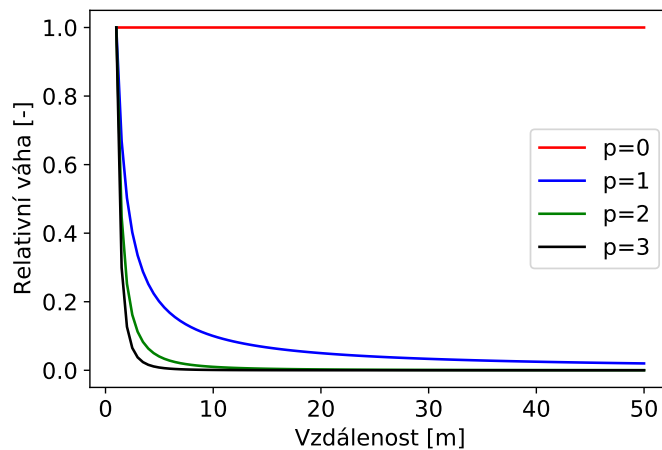
Anglicky Inverse distance weighing (IDW) je multivariační interpolační metoda. Jejím hlavním principem je, že známé body, které se nacházejí blíže k bodu, které chceme odhadnout, mají větší váhu než ty vzdálenější. Neznáme hodnoty *RSRP*, jsou vypočítané váženým průměrem všech hodnot v datasetu.

Rovnice 5.5 popisuje výpočet interpolační IDW funkce.

$$u(x) = \frac{\sum_{i=1}^n \frac{1}{d(x,x_i)^p} \cdot u_i}{\sum_{i=1}^n \frac{1}{d(x,x_i)^p}} \quad (5.5)$$

kde d je vzdálenost, x je neznámý bod, x_i je n -tý známý bod, p je mocnina, u_i je hodnota $RSRP$ známého bodu a u je odhadovaná hodnota pro danou souřadnici.

Pro výpočet vzdálenosti známého bodu x_i od námi odhadovaného x je opět použit Haversinův vzorec. Tento vzorec je implementován/naprogramován v několika zjednodušených krocích. Nejdříve je vybrán jeden neznámý bod a k jeho souřadnicím jsou vypočteny všechny vzdálenosti známých měření. Poté jsou všechny tyto parametry vážené podle zvoleného silového parametru p . Následně jsou všechny tyto váhy sečtené do jedné proměnné, nazveme ji *suminf*, neboli jmenovatel rovnice 5.5. Všechny váhy získané v prvním kroku jsou vynásobené se známými hodnotami $RSRP$ a jsou uloženy do proměnné *sumsup*, čitatele zmíněné rovnice. Posledním krokem je vydělení *sumsup* a *suminf* pro získání váženého odhadu.



Obr. 5.7: Snižování relevantnosti změřeného bodu se vzrůstající vzdáleností.

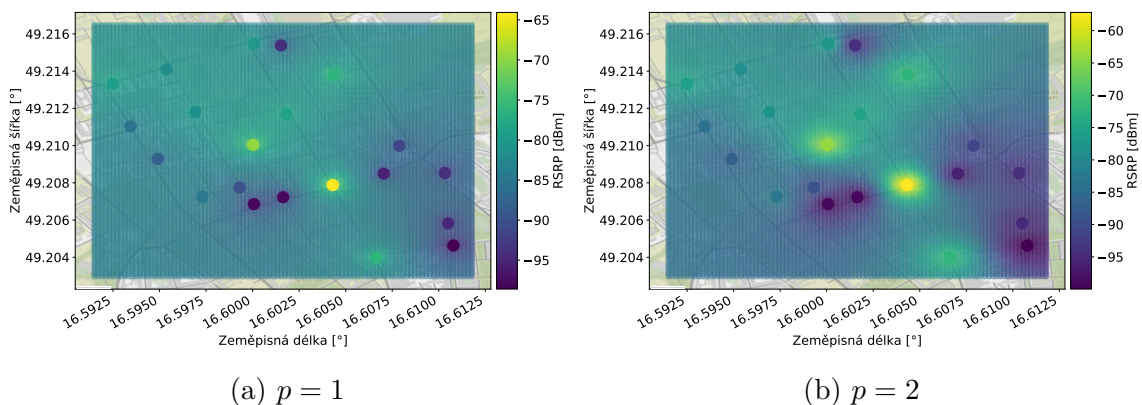
Jediný parametr, který lze v této metodě ovlivnit, je silový parametr p . Definuje závislost, jakou mají známé body ve větší vzdálenosti. Pokud je tento parametr nízkou hodnotou (0 nebo 1), je menší i váha vzdálenějších bodů a dochází k převládání průměrné hodnoty celého datasetu. Nastavení větší hodnoty silového parametru (2 a 3) zvyšuje vzdálenost relevantnosti každého známého bodu na své okolí. V praxi se nejčastěji využívají hodnoty 1 a 2. Geostatističtí analytici používají nejčastěji parametr $p = 1$. Parametr $p = 2$ se používá jako výchozí nastavení podle [21]. Experimentálně je otestovaná možnost kdy se parametr p rovná 3.

Pro omezení velkého vlivu vstupních vzorků na své okolí je možné přičíst vyhlazovací parametr σ k $d_{i,0}$. Tento parametr změnil tuto přesnou metodu na aproximující. Dojde ke zkreslení vzorků obsahujících větší naměřené hodnoty než u těch, které

byly změřené v dané oblasti. V dalších částech práce bylo ověřeno, že výpočetní výkon pro datové sady, o něž máme zájem, je dostatečný a není nutné tento parametr přidávat do rovnice pro výpočet IDW. Určitá filtrace vstupních parametrů byla implementovaná u další interpolační metody – Krigingu, kde již dochází k podstatně větším výpočetním nárokům.

5.6.1 Použití IDW na reálných datech

Metoda popsaná v minulé kapitole je implementovaná na datech z vlastního měření z kapitoly 3.2. Výsledné zobrazení predikce intenzity síly signálu je zobrazeno v Obr. 5.8. Částečně průhledné body označují měřená místa.



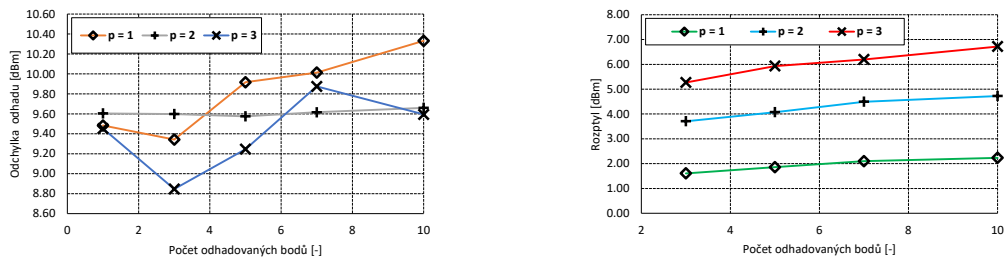
Obr. 5.8: Porovnání parametrů p na reálných datech.

Z prvního vyobrazení predikce pro $p = 1$ (za a) je patrné, že změřené body nemají tak široký dosah na své okolí. Vykazují spíše lokální charakter, který má dosah pouze několik metrů. Největší zastoupení má v celé ploše průměrná hodnota 83,85 dBm. Tento jev má za následek relativně velkou váhu všech bodů ve výběru. Proto má většina bodů větší váhu pouze v jednotkách metrů a poté dochází ke zprůměrování všech vzorků. Celkový rozsah se pohybuje od 65 dBm do 98 dBm.

V případě zvýšení parametru p na 2 (za b) lze pozorovat zvětšení dosahu dostupných měření. Dochází k většímu potlačení vlivu průměru všech dostupných dat na výkon signálu *RSRP*. Celkový rozsah se mírně zvýšil a pohybuje se v rozsahu od -60 dBm do -98 dBm.

Výsledky porovnání odhadu a rozptylu pro nejpoužívanější hodnoty parametru p jsou přehledně zobrazené na Obr. 5.9.

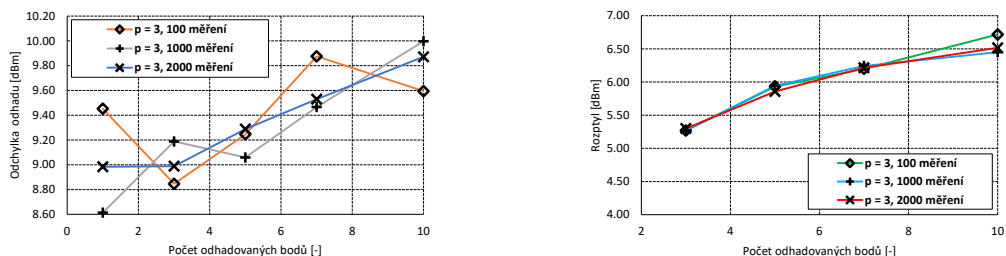
Z výsledků testů lze u grafu vyčíst, že IDW prezentuje lepší odhad než deterministické metody. U deterministických metod byla průměrná hodnota odchylky v rozmezí 10,5 dBm. U IDW se jedná o zlepšení okolo 1 dBm na přibližně 9,5 dBm.



Obr. 5.9: Srovnání odhadu a rozptylu pro odlišné nastavení silového parametru p .

Nejhorší výsledek, co se týče odhadu, reprezentuje křivka náležící $p = 1$. Jedná se o model, u něhož převládá průměrná hodnota datasetu. Tuto skutečnost znázorňuje i nejnižší hodnota rozptylu. Tím, že je nejčastěji odhadovaná průměrná hodnota, musí být i rozptyl nejmenší ze všech metod. Úsečka popisující silový parametr $p = 2$ disponuje téměř konstantní odchylkou cca 9,6 dBm. Je patrné, že u této metody nezáleží na počtu vstupních proměnných určených pro vytvoření modelu. Rozptyl se pohybuje v okolí střední hodnoty 4,2 dBm. Jedná se o konzistentní interpolační metodu. Pro silový parametr $p = 3$ je typický stejně nekonzistentní odhad jako tomu bylo u první křivky. Na rozdíl od ní, se výsledný odhad pohybuje v lepší přesnosti. Pouze když se počet odhadovaných bodů rovnal 7, vykazoval horší výsledky než křivka $p = 2$. Celkový rozsah naměřené a vypočítané hodnoty $RSRP$ se pohybuje od 8,8 dBm do 9,8 dBm. Z grafu rozptylu je patrný největší rozptyl hodnot okolo 6 dBm. Tato hodnota označuje skutečnost, že statisticky vykazuje nejlepší výsledky navzdory občasným nepřesným odhadům.

Bylo experimentálně zjištěno jestli je 100 průběhů simulací dostatečně vysoký počet. Byly navíc otestovány hodnoty 1000 a 2000 opakováními. Výsledky jsou vyneseny v grafech na Obr. 5.10.



Obr. 5.10: Porovnání vlivu odlišného počtu iterací odhadu na výsledné přesnosti.

Z výsledku lze usoudit, že dochází k určitému rozptylu hodnot, jenž nemusí být

způsoben omezeným počtem průběhů testování, ale náhodným charakterem odhadu. Tento jev lze nejlépe vidět, když byl odstraněn pouze 1 bod z 21 známých vzorků. Zde se odlišné počty měření liší o 0,8 dBm. Během odlišného počtu odhadovaných bodů si lze povšimnout blízké korelace mezi 1000 a 2000 opakováními. Lze tedy očekávat určitou neopřenost, která je ovšem stále přítomná. V porovnání rozptylu není takřka žádná odchylka kromě situace, kdy je odstraněná polovina vstupních bodů. Křivka reprezentující 100 měření se liší o zhruba 0,2 dBm ve své predikci.

5.6.2 Přizpůsobení IDW metody

Pro výpočet vhodné váhy byla v kapitole 5.6.1 použita mocnina vzdálenosti. V této části práce bude do určité míry upravená metoda získávání odhadu síly signálu *RSRP* aby odpovídala technickým specifikacím šíření LTE signálu v městském prostředí. Důležitým parametrem ovlivňujícím šíření elektromagnetické vlny v různých prostředích je Path Loss Exponent (PLE). Ten definuje, jakým způsobem se bude měnit síla signálu v závislosti na vzdálenosti v různých přenosových prostředích. Zahhruje v sobě odrazy od okolních budov, vícenásobné šíření a další jevy ovlivňující přenos signálu. Odpovídající hodnoty *PLE* v závislosti na prostředí jsou znázorněné v Tab. 5.1.

Tab. 5.1: Hodnoty *PLE* pro různá prostředí [22]

Prostředí	Path Loss Exponent
Volný prostor	2
Městská zástavba	2.7 - 3.5
Městská zástavba - bez přímé viditelnosti vysílač a přijímač	3 - 5
Vnitřní prostory - přímá viditelnost	1.6 - 1.8
Vnitřní prostory - bez přímé viditelnosti	4 - 6
Vnitřní prostory továren - bez přímé viditelnosti	2 - 3

Jak již bylo naznačeno, je nutné upravit původní IDW metodu tak aby lépe vystihovala šíření elektromagnetické vlny. Z tohoto důvodu byla použita Friisova rovnice pro výpočet útlumu se vzdáleností, viz 5.6 [23].

$$P_r = P_t \frac{G_r G_t}{L} \left(\frac{\lambda}{4\pi} \right)^2 \left(\frac{1}{d} \right)^\gamma \quad (5.6)$$

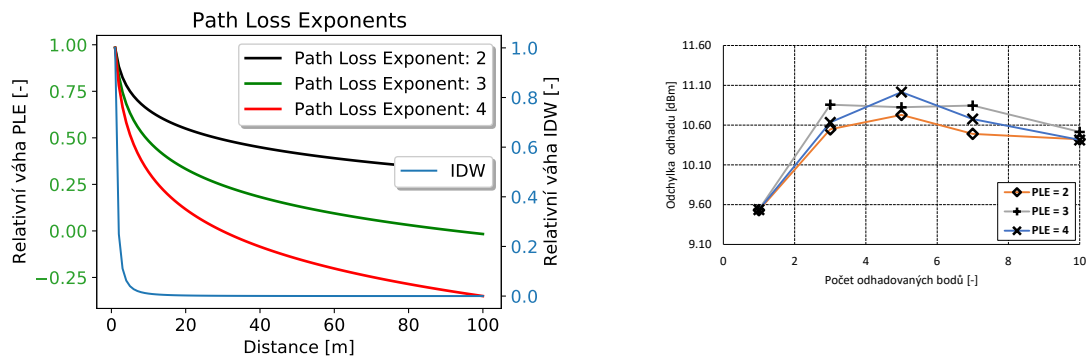
kde P_r a P_t jsou přijatý a vysílaný výkon [dBm], G_r a G_t zesílení na přijímací a vysílací straně [dB], L je útlum systému [dB], d je vzdálenost mezi dvěma body (vysílač-přijímač) [m], $\lambda = c/f$, kde f je nosná frekvence LTE [Hz], γ je zmíněné *PLE* [-]

V praxi není možné přesně určit všechny tyto parametry pro každé prostředí. Můžeme se setkat v terénu s různými typy vysílačů BTS s odlišnou konfigurací. Typicky se maximální výkon pohybuje v rozmezí od -43 do -48 dBm. Útlum systému záleží na použitých kabelech, místě zapojení zesilovačů v rádio frekvenčním řetězci a vysílací frekvenci. Tento útlum se pohybuje v rozmezí od 1 - 6 dB [24]. Mobilní telefony se také liší ve svých parametrech, které ovlivňují jak silný signál dokáží přijmout a poté ho dále upravit pro dosažení co nejlepšího objektivního zážitku pro uživatele.

V této práci bude použita zjednodušená rovnice 5.7. V případě, že známe všechny parametry z rovnice 5.6 a chceme dosáhnout co nejpřesnějšího modelu šíření, je možné zmíněnou zjednodušenou rovnici přepsat do podoby, která nám vyhovuje. Tímto je zachovaná univerzálnost této práce, která se dokáže přizpůsobit budoucím požadavkům. Pro výzkumné účely je možné tuto rovnici zjednodušit jako funkci vzdálenosti a *PLE* podle [23].

$$Pr = 19 - 10\gamma \log_{10}(d) \quad (5.7)$$

Z rovnice 5.7 lze odvodit, že rozhodujícím parametrem při výpočtu je *PLE*. Podle hodnot z Tab. 5.1 byly vybrány nejvhodnější proměnné - 2, 3, 4 odpovídající městskému prostředí. Byl upravený zjednodušený vzorec *Pr* aby odpovídal rozložení relativní váhy. Výsledky jsou prezentovány v Obr. 5.11 (b).



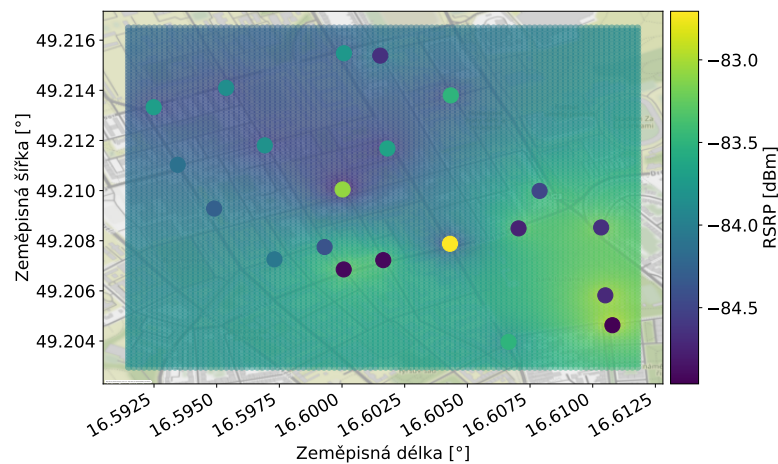
(a) Přizpůsobení váhy hodnot *PLE*, aby odpovídaly IDW ($p = 2$) (b) Výsledky srovnání odhadů pro různé hodnoty *PLE*

Obr. 5.11: Porovnání různých hodnot parametru *PLE*.

Jak je patrné, jednotlivé grafy se mezi sebou příliš neliší. Změny jsou minimální a je zde vidět podobný trend, kdy odchylka je nejnižší při odstranění pouze jednoho bodu ze vstupních dat a jeho následného dopočítání. Průměr za celé měření se pohybuje relativně vysoko okolo 10.6 dBm. Hodnoty rozptylu zde nejsou prezentované, je

ovšem možné přesně dohledat v příloze. Výsledky rozptylu jsou pro všechna měření v rozmezí od 0 do 0,4 dBm. Tak malý rozptyl je způsobený pomalou konvergencí k 0, jak jsme mohli vidět na Obr. 5.7. V odhadovaných bodech převládá průměrná hodnota, která ve velkém počtu měření reprezentuje odchylku okolo 10 dBm.

Pro následné simulace byl vybrán $PLE=3$, který by měl nejlépe vystihovat měřené prostředí. Jak je na první pohled patrné z Obr. 5.12, na rozdíl od Obr. 5.8 zde došlo k omezení celkového rozsahu $RSRP$, nový rozsah je pouze od -82 do -85 dBm. To znamená, že ani body, které jsme změřili například s úrovní -70 dBm, nemají podle toho algoritmu správnou odhadovanou úroveň. Napříč celým spektrem je vidět patrné průměrování, které se pohybuje okolo hodnoty -84 dBm.



Obr. 5.12: Graf rozložení pokrytí signálem pro $PLE=3$.

Tento jev je způsobený pomalou změnou váhy jednotlivých vzorků, která odpovídá křivce změny útlumu v prostředí.

5.6.3 Omezení dosahu IDW

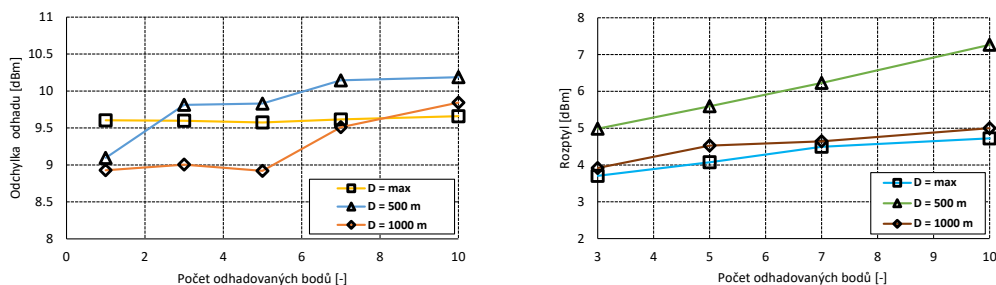
Z praxe již víme, že v reálném prostředí má každý změřený vzorek pouze určitou vzdálenost do které je relevantní. Musíme do určité míry předpokládat výskyt BTS v určitém dosahu od změřeného bodu. Cílem této práce není započítávat do odhadu signálů i známé BTS, které by pomohly získat přesnější modely šíření pro zvolení lepší regresní metody. Na základě těchto předpokladů je nutné zavést omezení, které se často používá jak u IDW, tak i u Krigingu. Nastavení tohoto omezení má 2 možnosti:

1. Omezení počtu bodů, které by se použily pro výpočet odhadované úrovně $RSRP$.
2. Omezení vzdálenosti, do které by okolní vzorky přispívaly patřičnou váhou.

První možnost by se dala použít, pokud bychom měli vzorky rovnoměrně naměřené a pokud by měly mezi sebou ideálně konstantní vzdálenosti či by tvořily pravidelnou mřížku. Tomuto požadavku je těžko vyhovět i v kontrolovaném prostředí, kde jsou vzorky speciálně sbírané za tímto účelem. Použití této omezovací metody na data, u kterých neznáme rozložení, by způsobovalo značnou nepřesnost a nedalo by se použít pro důvěryhodný odhad.

Z tohoto důvodu je upřednostňovaná druhá metoda, níž je možné si nastavit hraniční vzdálenost, kterou budeme označovat jako *dosah*. *Dosah* popisuje maximální vzdálenost změřeného bodu, do kterého poskytuje relevantní odhad. Pokud bude *dosah* příliš malý, odhad bude vytvořený na základě malého počtu bodů. Tato skutečnost může zapříčinit vznik lokálních nepřesností, kde v dosahu bude pouze jeden, nebo žádný bod. Pokud bude *dosah* naopak příliš velký, může nastat podobný případ jako na Obr. 5.12. Zde měly všechny body v měřicím výběru značný vliv na celou plochu a došlo k celoplošnému zprůměrování.

Je tedy nutné zvolit vhodný *dosah*. *Dosah* můžeme odhadnout, jak už bylo naznačeno, pomocí vzdálenosti jednotlivých BTS v dané lokalitě. Byl provedený výpočet vzdáleností mezi všemi BTS. Vzdálenost mezi nejbližšími stanicemi se pohybovala v rozmezí od 200 do 600 metrů. Nejsilnější příjem signálu nastává pokud je zajištěná přímá viditelnost mezi BTS a UE. S touto variantou je také nutné počítat pro případ řidší zástavby.



Obr. 5.13: Porovnání odchylek odhadu a rozptylu původního IDW.

Byly vybrány 3 dosahy 500 m, 1 km a bez omezení. Silový parametr byl nastavený na 2. Rozdíl je v tomto nastavení pouze v přidání parametru *dmax*, který označuje maximální vzdálenost ve které se může naházet známý bod od neznámého aby tvořil část predikce. Pokud nastane případ, že by v dané lokalitě nebyl žádný bod, ze kterého by bylo možné vytvořit predikci, jak je tomuto bodu přiřazena průměrná hodnota celého datasetu. Z výsledných měření je patrné, že nejlepšího výsledku bylo docíleno, když byl dosah omezený na 1 km. Rozdíl je patrný především pro menší

počet odhadovaných bodů, kde se odchylka pohybuje okolo 9 dBm. Jedná se zatím o nejlepší odhad. Rovněž i rozptyl se pohybuje na nízké hodnotě v rozmezí 4,5 dBm. Pokud odstraníme okolo 30 % bodů datasetu, tedy 7 vzorků, odchylka se zvyšuje na 9,5 dBm a přibližujeme se druhé křivce, která nemá žádné omezení vzdálenosti. Pohybuje se konstantě lehce nad hodnotou 9,6 dBm. Jako nejhorší metoda z těchto tří výsledků je splajn značící omezení vzdáleností na 500 m. Z analýzy vzdáleností BTS vysílačů by tato metoda měla podávat nejlepší výsledky.

Omezení vzdálenosti bylo implementováno i pro metodu z kapitoly 5.6.2. Pro tyto výsledky byla typická vysoká odchylka (11 dBm) a stejně i značný rozptyl (6 dBm). Pro tyto nedostatečné výsledky zde nejsou uvedené grafy. Výsledky jsou obsažené v příloze.

5.7 Kriging

Kriging je metoda, která je hojně využívána v geostatistice, k predikci hodnoty náhodné proměnné v určitém regionu. Stejně jako ostatní modely využívá různé úrovně vlivu jednotlivých měření. Tuto metodu je možné použít i pro odhad síly signálu *RSRP*.

Důležitými předpoklady pro kriging jsou stacionarita a izotropie. Pro stacionaritu je důležité, aby se ve výběru dat nenacházela místa, kde dochází ke skokové změně úrovně *RSRP* v krátkých úsecích. Z analyzovaných dat je patrné dobré a rovnoměrné pokrytí signálem LTE, kde se tento jev nevyskytuje. U izotropie je důležité, aby výsledný variogram zůstal neměnný v celém úseku vybraném k analýze [25].

Na rozdíl od IDW je kriging založený na prostorové korelaci (závislosti), kterou lze použít pro vysvětlení variací v úrovních *RSRP*. Váhy nezávisí pouze na vzdálenostech mezi měřenými body, ale také na jejich uspořádání v prostoru. Čím více bodů budou obsahovat vstupní data, tím více bude růst výpočetní náročnost. Tuto překážku lze částečně eliminovat seskupováním blízkých skupin hodnot. Kriging je vícestupňový proces, který zahrnuje průzkum statistické analýzy dat a modelování variogramu.

Variogram je možné popsat pomocí tří parametrů:

- *Nugget*, znázorňuje posunutí variogramu v počátku, neboli představuje variabilitu dat. Čím výše je *nugget* umístěný na ose *y*, tím jsou vstupní data zanesená s vyšší nepřesností. Je možné aby se nacházelo více bodů v těsné blízkosti od bodu, který chceme určit, *nugget* je zobrazený jako jistá aproximace.
- Druhým parametrem je *Sill*, určující *práh* variogramu.

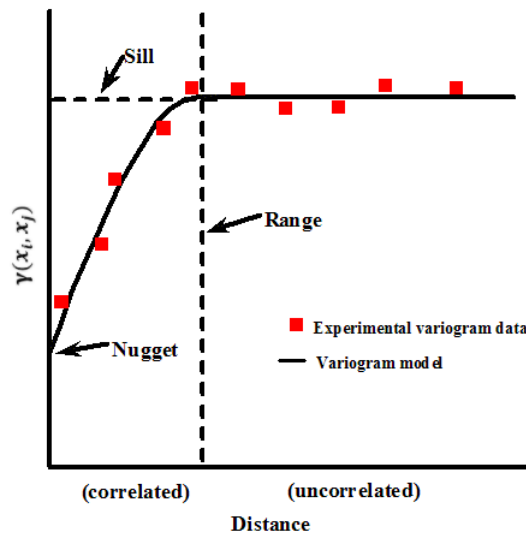
- Posledním parametrem je *Range*, česky *rozsah*. Je přímo závislý na *prahu* variogramu a určuje maximální vzdálenost dvou bodů, kdy je odhad ještě relevantní. Pokud je tato vzdálenost překročena a *práh* je na maximální hodnotě, jedná se o nekorelovanou závislost.

Variogram znázorňuje závislost γ na vzdálenosti dvou bodů. Proměnná γ reprezentuje rozdíl datových bodů y_i, y_j na souřadnicích x_i, x_j - čím větší je vzdálenost dvou bodů, tím menší vliv korelace je očekáván, viz rovnice 5.8.

$$\gamma(x_i, x_j) = 0.5(y_i - y_j)^2 \quad (5.8)$$

kde x_i, x_j jsou vzdálenosti dvou bodů (vypočítané z jejich souřadnic) a $y_i - y_j$ reprezentují jednotlivé úrovně bodů *RSRP*. V dalších rovnicích se používá zjednodušený zápis souřadnic, kde x_i, x_j je nahrazeno body x_0 .

Variogram navíc oproti ostatním metodám obsahuje určitou nepřesnost, či interval pravděpodobnosti, ve kterém se predikované hodnoty mohou nacházet. Pro účely této práce je tento interval zprůměrován. Křivka variogramu má tvar podobný logaritmu. Příklad variogramu i s popisem dat je zobrazen na Obr. 5.14.



Obr. 5.14: Model variogramu, převzato z [26].

Váhy pro každý interpolovaný bod se počítají podle prostorové struktury interpolovaného místa ve vztahu ke všem dostupným bodům. Tyto váhy jsou odvozeny z variogramu a dále se aplikují jednotlivé vzorky podle rovnice 5.9 [27].

$$\hat{Z}(x_0) = \sum_{i=1}^N \lambda_i z(x_i) \quad (5.9)$$

kde $\hat{Z}(x_0)$ je odhad krigingu na pozici (x_0) , λ_i určuje neznámou váhu pro vzorek (x_i) . Suma všech vah musí být rovná 1.

Soustava rovnic pro výpočet měřených bodů při základním krigingu:

$$\begin{bmatrix} K_{11} & K_{12} & \cdots & K_{1m} & 1 \\ K_{21} & K_{22} & \cdots & K_{2m} & 1 \\ \vdots & & & & \\ K_{n1} & K_{n2} & \cdots & K_{nm} & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} K_{10} \\ K_{20} \\ \vdots \\ K_{n0} \\ 1 \end{bmatrix}$$

kde K_{nm} je matice modelových hodnot variogramu všech párů měřených bodů, λ_i je vektor vah měřených bodů μ je Langerův multiplikátor, K_{n0} je kovariance (statistická míra lineární závislosti dvou veličin) mezi měřeným vzorkem n a odhadovaným bodem.

5.7.1 Modely variogramu

Zvolení vhodného modelu variogramu je důležitá část analýzy. Největší vliv má tento model na korelovanou část variogramu - od *nugettu* po *práh* variogramu. V této práci je přezkoumáno několik odlišných modelů, které jsou popsány dále. Podle tzv. 'fittingu', byly vybrány ty nejvhodnější, které svou křivkou nejlépe popisují rozložení dat.

- Gaussův model

$$p \cdot \left(1 - e^{-\frac{d^2}{\left(\frac{4}{3}r\right)^2}}\right) + n \quad (5.10)$$

- Exponenciální model

$$p \cdot \left(1 - e^{-\frac{d}{r/3}}\right) + n \quad (5.11)$$

- Sférický model

$$\begin{cases} p \cdot \left(\frac{3d}{2r} - \frac{d^3}{2r^3} + n\right) & d \leq r \\ p + n & d > r \end{cases} \quad (5.12)$$

- Lineární model

$$s \cdot d + n \quad (5.13)$$

- Mocninný model

$$s \cdot d^2 + n \quad (5.14)$$

kde d označuje vzdálenost mezi body, pro něž chceme určit variogram, p označuje částečný práh (*práh - nugget*), r určuje rozsah, n je *nugget*, s je škálovací faktor nebo sklon variogramu, e exponent pro mocninný model.

5.7.2 Použití krigingu na reálných datech

Stejně jako u předchozích interpolačních metod i tato je nejdříve otestována na vlastním datasetu. V jazyku Python je dostupných několik knihoven, které jsou schopné tyto operace provádět. Byla vybrána sada nástrojů PyKrige³, která podporuje vytváření 2D a 3D modelů. Kromě již popsaných modelů variogramu je možné nastavením specifických parametrů definovat vlastní modely variogramů. PyKrige také umožňuje vygenerování variogramů s nejlépe odpovídající funkcí.

Testování s jistými obměnami probíhalo stejně, jako v předchozích metodách. PyKrige obsahuje několik tříd, které umožňují přímo generovat konkrétní možnosti krigingu. V této práci byly otestovány 2 nejvíce používané, a to základní krigování (Ordinary Kriging-OK) a univerzální krigování (Universal Kriging-UK). Základní rozdíl mezi OK a UK je ve výpočtu průměrné hodnoty. Pro OK se vypočítá průměrná hodnota z celého datasetu jako v předchozích metodách. UK pracuje s nesplněním podmínky prostorové stacionarity. Průměrná hodnota je určena za pomoci dvou komponent trendu (driftu) a reziduí. Trend určuje průměrnou hodnotu v daném místě, po jeho výpočtu je možné získat hodnotu rezidua odečtením hodnoty trendu v daném místě od skutečné hodnoty. Tyto třídy požadují několik vstupních parametrů k vytvoření modelu. Nejdůležitějšími parametry jsou souřadnice bodu a jeho datová hodnota. Tyto hodnoty musí být v samostatných polích. Dalším parametrem je zvolení *variogram_model* z možností v kapitole 5.7.1. Důležitým parametrem je *verbose*, který vypíše do konzole parametry charakterizující vlastnosti funkce. Každý model má jiné vstupní parametry, viz Tab. 5.2. Posledním parametrem je *nlags*, ten označuje počet průměrných intervalů pro variogramy. Tyto intervaly jsou podobné jako v první kapitole, kde byly použité histogramy. Tento parametr označuje počet bodů vyobrazených ve variogramu, viz Obr. 5.15. Větší počet nám umožní lépe posoudit ‚fitting‘ funkce.

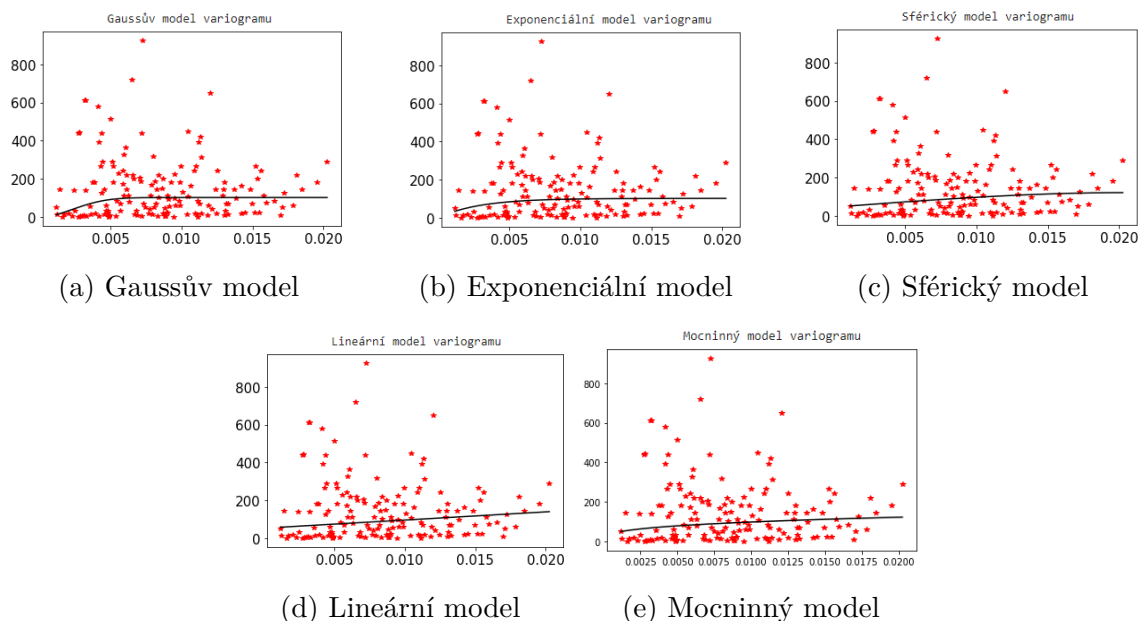
Výsledky automatického nastavení ‚fittingu‘ pro náš dataset jsou prezentovány v Tab. 5.2.

Tab. 5.2: Parametry variogramu krigingu pro brněnský dataset

Model variogramu	Slope	Scale	Exponent	Partial Sill	Full sill	Range	Nugget
Gausův model	-	-	-	101.7453	101.793	0.0058	0.0477
Exponenciální model	-	-	-	101.967	101.968	0.00865	0.001
Sférický model	-	-	-	75.83	121.198	0.0202	45.3684
Lineární model	4312.534	-	-	-	-	-	52.8254
Mocninový model	-	423.373	0.3179	-	-	-	5.35e-9

Jak už bylo zmíněno, tak analýza modelů variogramu a zvolení funkce uzpůsobení

¹<https://geostat-framework.readthedocs.io/projects/pykrige/en/stable/index.html>

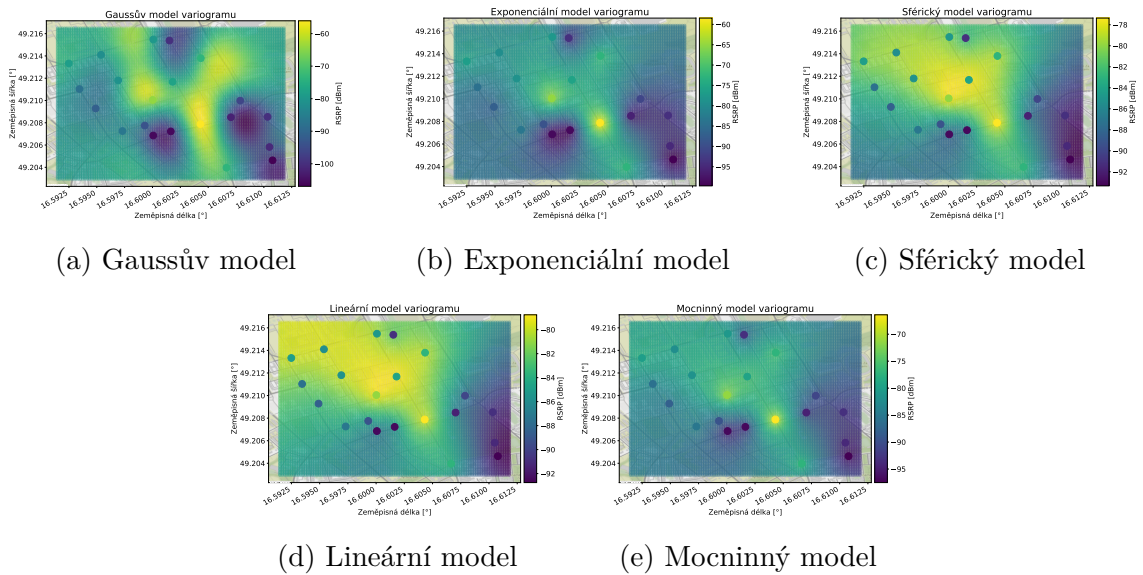


Obr. 5.15: Porovnání odlišných variogramů pro brněnský dataset.

je důležitá pro co nejpřesnější odhad. Tuto funkci umožňují třídy nástrojů z PyKrige automaticky přizpůsobit. V ideálním případě by měla křivka vypadat jako na Obr. 5.14. Na příkladě datasetu z Brna je patrné, že se zde nachází poměrně velký rozptyl hodnot. Není možné s dostatečnou přesností vytvořit takový splajn, který by popsal všechny body v grafu. Nicméně je nutné otestovat jestli i přes tuto nepřesnost bude metoda podávat lepší výsledky než ostatní testované.

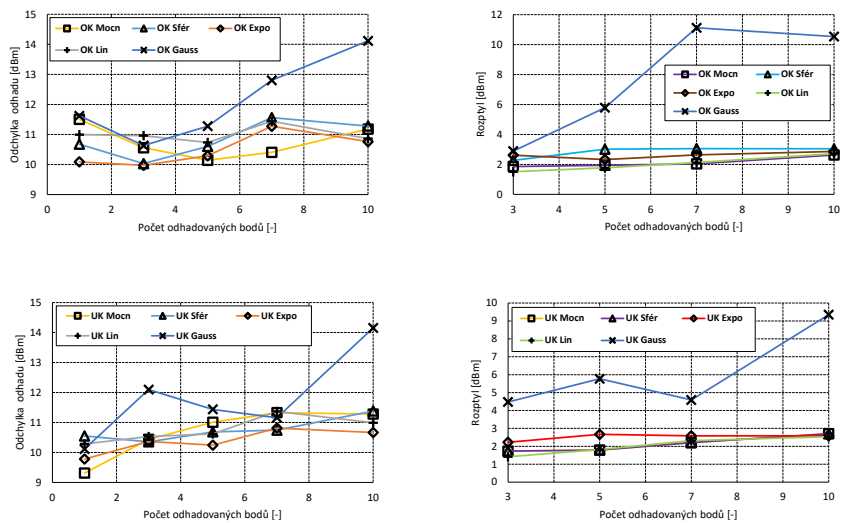
Obr. 5.16 znázorňuje jednotlivé modely. Hned na první pohled na rozsah *RSRP* je možné odvodit, které metody zavádí příliš velké průměrování jako u upraveného IDW v minulé kapitole. Mezi tyto metody se řadí lineární, mocninný a sférický model. Jejich rozsah odhadu nepokrývá všechny možné změřené hodnoty. Gaussův a exponenciální model mají větší pravděpodobnost podání lepšího výsledku u predikce *RSRP*. U Gaussova modelu je patrné převládání vysokých změřených hodnot, lze tedy očekávat větší rozptyl při testování. Exponenciální rozložení se podobá IDW při silovém parametru $p = 2$. Dosah jednotlivých bodů je malý ale na rozdíl od mocninného modelu, se kterým si je velmi podobný, poskytuje plný rozsah hodnot datasetu. V dalším testu budou jednotlivé metody otestovány a porovnány mezi sebou.

Na první pohled je z dat testování OK patrná odlišnost Gaussova modelu. Ten u všech měření vykazoval nejhorší výsledky, hlavně při zvyšujícím se počtu odhadovaných bodů. Všechny ostatní modely vykazovaly přibližně stejné hodnoty napříč různými vstupními parametry. Odchylka se pohybovala od 10 do 11 dBm. Jedná se o horší výsledky než v předchozích méně složitých metodách. Při pohledu na graf



Obr. 5.16: Srovnání šíření signálů pro různé modely variogramů

rozptylu je patrné, že v celém odhadu datasetu převládá střední hodnota. Rozptyl má skoro konstantní hodnotu 2 dBm pro změnu vstupních parametrů.



Obr. 5.17: Porovnání odchylek odhadu a rozptylu OK a UK.

Testy byly provedené i u UK, který by měl lépe pracovat s lokálními výchytkami. Výsledky jsou velmi podobné jako u OK. Gaussův model znovu potvrzuje nejhorší odhad, kvůli přiřazování větší váze větším hodnotám *RSRP*. Průměrně se odhad zhoršil o 0,5 dBm oproti OK. Rozptyl byl už u předchozí metody na nízké úrovni i u UK se drží kolem hranice 2 dBm.

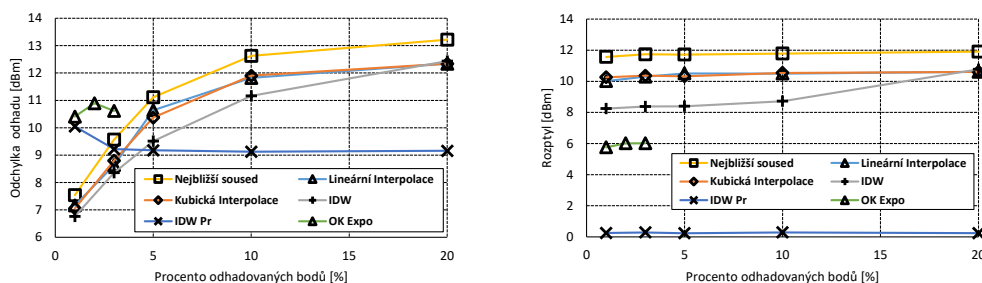
6 Srovnání dostupných metod

Metody s nejlepšími výsledky, které byly popsány v kapitole 5, jsou nyní otestovány na reálných datech. Dataset z Brna byl vytvořený po analýze plošné hustoty měření ve Vídni. Dá se očekávat, že výsledky by se neměly elementárně lišit. Byly zvoleny různé testovací metodiky, které nám umožní přezkoumat dostupné data z různých úhlů.

6.1 Rozdělení podle procentuálních vah

V prvním případě byla změněná testovací metodika aby byla splněná pravidla pro vytvoření trénovacího a testovacího modelu v kapitole 4. Počet bodů, které byly určeny k odhadu je určován procentuálně podle počtu vstupních dat datasetu.

Dataset Vídně obsahuje 5280 měření, tento počet nedělal jednodušším algoritmům žádné problémy a vytvoření odhadu netrvalo dlouhou dobu. Neobjevila se ani žádná chyba v implementaci. To platí pro metody nejbližšího souseda, lineární, kubickou a IDW. U krigingu bylo možné rozdělit dataset pouze na 3 % testovací sekvence, zatímco ostatní metody zvládly až 20 %. Z tohoto důvodu následující grafy obsahují pouze menší vzorek těchto měření.



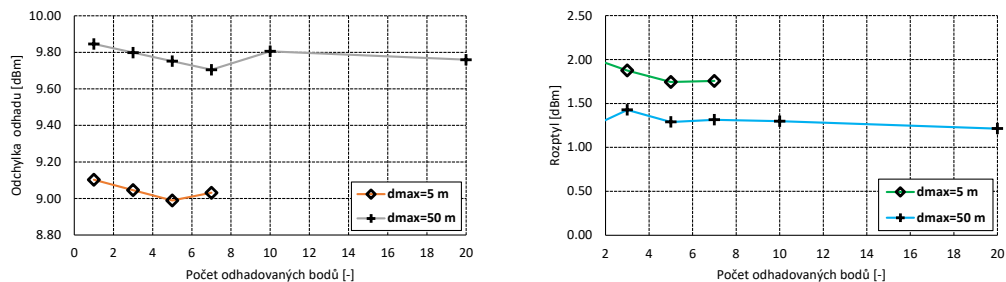
Obr. 6.1: Porovnání odchylek odhadu a rozptylu všech metod na datech z Vídně.

Na tomto dostatečně velkém datasetu je vidět teoretický předpoklad, který na malém počtu měření v Brně nebyl jasně pozorován. Jedná se o snižování přesnosti odhadu se zvyšujícím se počtem bodů, které chceme predikovat, respektive snižování počtu vstupních dat. Rozdíl mezi změřenou a odhadovanou hodnotou tvoří pro 1 % odstraněných dat okolo 7 dBm. Jedná se doposud o nejmenší vypočítanou odchylku od změřené hodnoty. Všechny metody kromě základního krigingu a upravené IDW metody exponenciálně zhoršují svůj odhad až do 5 % odhadovaných bodů. Tam se odchylka pohybuje v rozmezí od 9 do 11 dBm a dále roste. U 20 % dosahuje rozdílu

až 13 dBm pro metodu nejbližšího souseda. Pro větší počet odhadovaných bodů od 5 % vytváří nejlepší výsledky upravená IDW Pr metoda podle rovnice 5.7. Tato rovnice ve velké míře využívá pouze průměr celého datasetu a s tím souvisí i její takřka nulový rozptyl. Ostatní algoritmy mají relativně velký rozptyl, který se příliš nemění podél osy x . Rozptyl se v celém zkoumaném souboru pohybuje od 8 dBm pro IDW až po 12 dBm pro metodu nejbližšího souseda.

6.1.1 Omezení nadbytečnosti vzorků

V minulé kapitole byla zdůrazněna omezená použitelnost krigingu, která pro vytvoření většího výpočetního modelu vypotřebovala veškeré hardwarové prostředky. Proto byl vytvořen skript, který zvýší použitelnost této metody i pro výpočet většího množství bodů. Toho bylo docíleno eliminací míst s velkým počtem měření na malou plochu.



Obr. 6.2: Porovnání odchylek odhadu a rozptylu exponenciálního OK s odlišným počtem vstupních proměnných

Původní soubor dat obsahuje 5280 měření pro plochu celé Vídně. Protože doba během níž byla data snímána, obsahuje data za 6 měsíců jsou zde místa, která se často měří touto aplikací, zde dochází ke shlukování měření. Pro metody, které vypočítávají vzdálenosti 1 bodu ke všem ostatním, představuje každý nepotřebný vzorek navíc značné množství výpočetního výkonu. Aby bylo možné tato data dále zpracovávat je nutné toto číslo omezit. Z tohoto důvodu byl vytvořen skript, který v určitém okolí, které si zvolíme, odstraní všechny další vzorky. Byla testována různá nastavení od 5 do 50 metrů. Je nutné tuto hodnotu zvolit s rozvahou, protože pokud se zvolí až příliš malá, zůstane zachováno značné množství vzorků a výpočet se zrychlí pouze nepatrně. Pokud se tato hodnota zvolí až příliš vysoká, dojde k odstranění konkrétního intervalu vzdáleností, které poté budou chybět u modelování variogramu krigingu. Tento nedostatek počtu konkrétních měření vede k nemožnosti

vytvoření modelu a k nefunkčnosti odhadu. Tímto omezením jsme schopní zredukovat při nastavení parametru na 50 m až 73 % vzorků. Při snížení vzdálenosti na 5 m dojde k redukci původních měření o 27 %. Vybrané omezené datasety jsou zobrazené na Obr. 6.2. Byl opět zvolený OK s exponenciálním modelem variogramu, který prezentoval nejlepší výsledky při testech na vlastním datasetu.

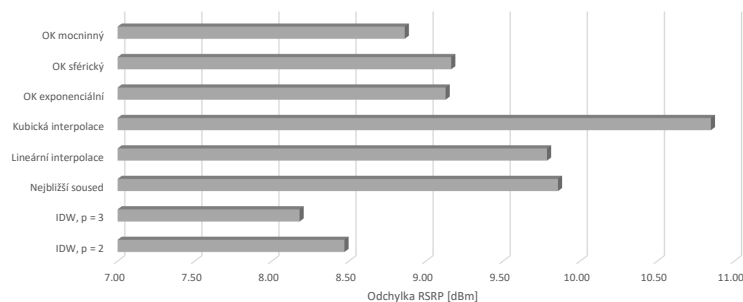
Pro omezení vzdálenosti nejbližších bodů pouze na 5 m došlo opět u většího počtu odhadovaných bodů k pádu aplikace, proto zde chybí měření pro 10 % a 20 %. Z výsledků je patrné podobné rozložení jako u brněnského datasetu. Odchylna zůstává konstantní bez ohledu na počet vstupních bodů.

6.2 Otestování na vybraných bodech

V předchozích kapitolách byly okomentované vybrané interpolační metody, které byly vždy otestované stejnou metodou. Tato metoda byla zaměřená na prověření celého datasetu, kde byly odhadované body vybrané náhodně. Tato testovací metodika byla dostatečně zprůměrována, ale stále je možné zanesení menší nepřesnosti, protože každé spuštění programu vygenerovalo výsledek s odchylkou.

Pro přesné porovnání metod mezi sebou je tato metoda upravená, aby byla vždy zajištěná stejná vstupní data. Každý výsledek je vypočítán deterministicky a není nutné zavádět průměrování opakováním, jak tomu bylo v předchozích kapitolách. Tím se snižuje i výpočetní náročnost.

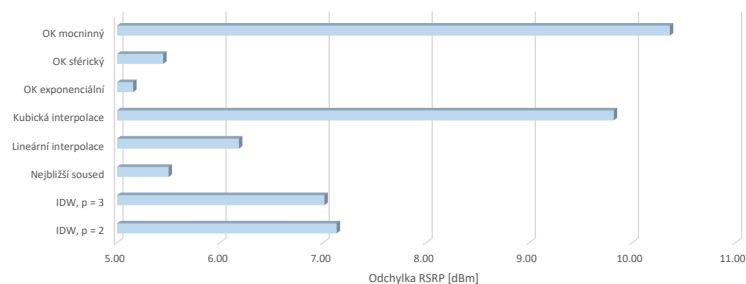
Výsledky jsou přehledně zobrazené ve formě sloupcových grafu na Obr. 6.3. Byly vybrané pouze interpolační algoritmy, které prezentovaly nejlepší výsledky v předchozích kapitolách.



Obr. 6.3: Srovnání všech metod pro dataset Brno Lužánky

Nejllepších výsledků opět bylo dosaženo metodou IDW, kde se rozdíl změřené a vypočítané hodnoty pohybuje v rozmezí od 8,18 to 8,5 dBm. Kriging podal uspokojivé výsledky, odchylka se pohybuje okolo 9 dBm. Nejhorší výsledky měly deterministické metody, které průměrně dosahovaly odchylky okolo 10 dBm. Tato skutečnost může být způsobená omezeným rozsahem těchto metod. Pokud byly dopočítávané postranní body, které nebylo možné interpolovat, byla jim přiřazena průměrná hodnota datasetu.

Stejný testovací postup byl proveden i pro vídeňský dataset. Jeho výsledky jsou prezentované na Obr. 6.4. V předchozí metodě byl vždy dopočítáván pouze jeden vzorek datasetu. Byl vybrán pouze jeden, protože u této možnosti interpolační modely dosahovaly nejlepšího odhadu. Tento postup byl okováván pro celý brněnský dataset. Ve Vídni byla zvolená podobná metoda z ohledem na větší počet vstupních dat. Byl vždy dopočítáván pouze 1 bod pro každou metodu. Tento postup byl opakován 10 krát a nakonec zprůměrován. Jedná se o malý vzorek, ale výstupem tohoto měření je vzájemné porovnání interpolačních metod mezi sebou.

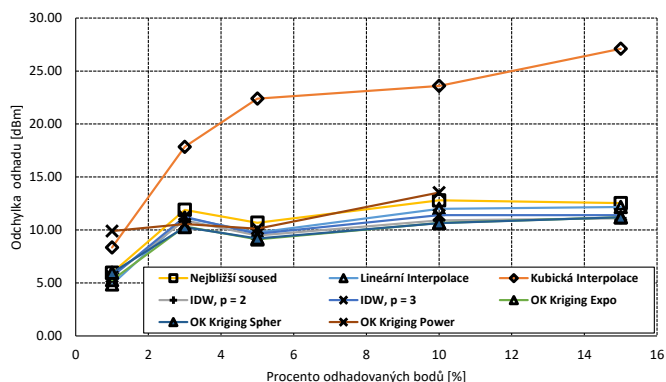


Obr. 6.4: Srovnání všech metod pro vídeňský dataset

Z výsledků je patrné, že opět jeden z nejhorších odhadů měla kubická interpolační metoda, 9,8 dBm. Metoda mocinného krigingu, která pro brněnský dataset měla nejlepší výsledky z krigovacích metod v tomto testu předvedla nejhorší odhad. Špatný odhad byl způsobený nesprávným automatickým dosazením parametrů mocinného krigingu. Nicméně model variogramu byl správně zvolený pro ostatní dvě krigovací metody, které prezentují nejnižší odchylky 5,4 a 5,2 dBm. Metoda nejbližšího souseda taktéž prezentovala dobrý odhad, s rozdílem 5,5 dBm. Tak dobrý odhad byl způsobený výběrem bodů, které byly blízko odhadovaného vzorku, tato metoda přejala hodnotu těchto vzorků bez jakékoliv aproximace s malou odchylkou. IDW metody měly průměrnou odchylku okolo 7 dBm.

6.3 Společný model

V předchozích kapitolách byly jednotlivé interpolační metody testované vždy samostatně. Každá metoda měla svůj program, kde bylo možné rychleji a přehledněji měnit parametry a zkoumat výsledky. Podle zadání práce byl vytvořený model, který obsahuje všechny zmíněné interpolační metody a umožňuje jejich současné porovnání. Nejdůležitějším parametrem je ověření správnosti odhadu, proto je zde vypočítávaná pouze odchylka. Výsledky jsou implementovány stejným způsobem jako v kapitole 6.1.



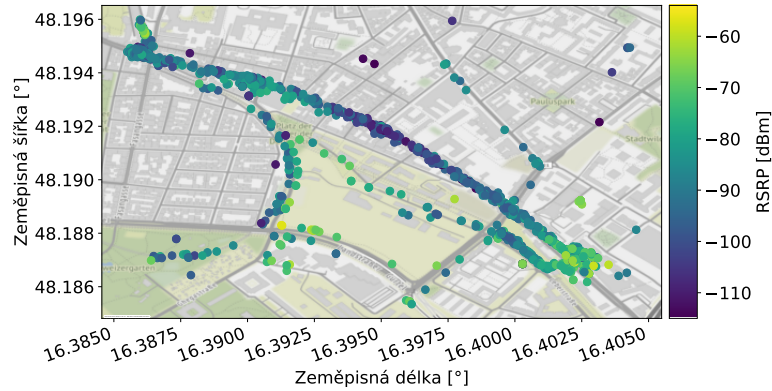
Obr. 6.5: Porovnání odchylek odhadu ve společném modelu ve Vídni

Z výsledků je zřejmá malá odchylka pro dopočítávání nižších procent z obsahu datasetu. Tato odchylka se pohybuje mírně nad hranicí 5 dBm pro všechny metody kromě mocninného krigingu a kubické interpolace. Kubická interpolace měla nejhorší výsledky v dosavadní práci. Po analýze výstupního souboru bylo zjištěno, že zvláště při odhadu slabší úrovně signálu *RSRP* (pod -80 dBm) docházelo často k odhadu až příliš nízkého signálu - až -120 dBm. Velká četnost těchto odhadů způsobila výrazně odlišné výsledky od ostatních metod. Ostatní metody se v odhadu lišily pouze nepatrně a lze pozorovat stejnou odchylku 10,5 dBm pro různý počet vstupních bodů.

6.3.1 Otestování datasetu na menší ploše

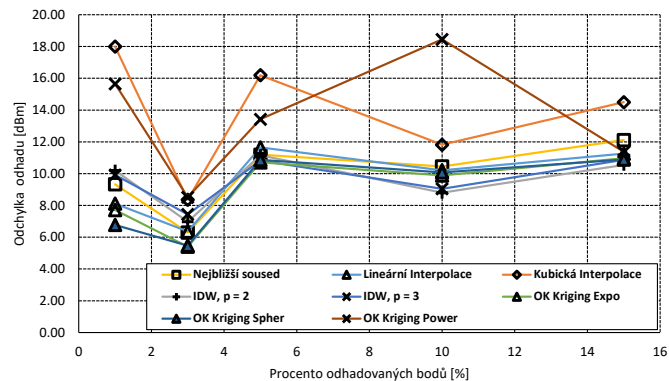
Současné výsledky byly vztahované na relativně velkou plochu celého města. Z testů je patrné, že pro malý počet dopočítávaných vzorků je odchylka okolo 5 až 7 dBm. Podle provedených časových měření v předešlé kapitole je možné tento rozsah považovat jako akceptovatelný. Bylo potvrzeno, že měření stejných míst po určité časové době také zaznamenává odchylku do 5 dBm v skoro 70 % měření.

Byla zvolena menší plocha (2 x 1,1 km), konkrétně část ulice v oblasti obsahující velký počet měření. Na této datové sadě je otestováno, jestli zvolení konkrétní metody má vliv na odhad síly signálu *RSRP* v oblasti kde je dostatek měření (více jak 500 měření na km²). Mapa této oblasti je zobrazená na Obr. 6.6. Počet vstupních měření je 1258.



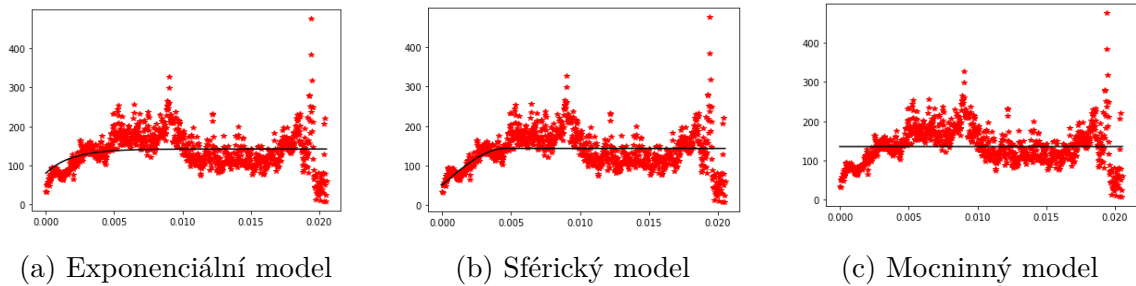
Obr. 6.6: Mapa části Vídně obsahující s početnou plošnou hustotou měření

Výsledky změny odchylky jsou prezentovány na Obr. 6.7, kde je opět patrné, že mezi nejhorší interpolační metody se řadí kubická interpolace a mocninný kriging. Překvapivá je nekonzistentnost výsledků se zvyšujícím se počtem bodů k odhadu. Z této datové sady je zřejmý vliv vybraných bodů na odhad modelu.



Obr. 6.7: Porovnání metod v části Vídně

Data, u kterých bylo odhadováno 5 % datasetu mají lepší výsledky, než kdy bylo k dispozici více vstupních vzorků. Většina měřených bodů se nachází v husté mřížce, dalo by se očekávat, že bod vybraný z této skupiny se bude lišit pouze minimálně v odhadu podle metody. V tomto případě je nutné klást větší důraz



Obr. 6.8: Porovnání různých variogramů pro část Vídně.

na symetrické rozmístění ve mřížce, než na jejich hustotu. Všechny metody mají podobné výsledky, nejnižší odchylku pro malý počet odhadovaných bodů má sférický kriging, s odchylkou 7 dBm.

Byly analyzovány variogramy, aby bylo prozkoumáno, proč mocninný mode krigingu prezentuje tak rozdílné odhady *RSRP*. Z grafů na Obr.6.8 je zřejmé, že nastavení vhodné funkce variogramu má důležitou roli na konečný výsledek. U prvních dvou modelů lze pozorovat kopírování křivky variogramu podle dostupných dat. U určení mocninného OK došlo ke špatnému výpočtu vhodné funkce. Tento rozdíl způsobil, že i blízké body neměly žádný kovarianční vliv na odhad. Zamezení nastavení špatného výpočtu je možné nastavením vlastních parametrů.

Závěr

V této diplomové práci byly v první kapitole popsány distribuční sady, které obsahují detailní informace o měřeních v síti LTE. Byly definovány nejdůležitější parametry, které se využívají při odhadování síly signálu v místech, kde měření chybí. Získání, zpracování a vyhodnocení datových sad bylo realizované v prostředí Google Colaboratory a neprogramové v jazyku Python. Byla prozkoumána plošná hustota měření pro oblasti jak o velikosti státu, tak i města. Na základě získaných informací byla navržena vlastní měřicí kampaň a byl definovaný způsob získávání dat pomocí mobilního telefonu. Byla naprogramovaná sada testů pro ověření věrohodnosti získaných odhadů. Následně proběhlo pilotní měření k získání vzorků. Byl proveden rozbor regresních a interpolačních metod a byly vybrány ty nejuvhodnější. Jednotlivé metody byly postupně popsány a byly otestovány na vlastním datasetu. Následně bylo nejlepší nastavení z každé metody implementováno a otestováno na reálných datech z nejpočetnějších datových sad.

Rozdíl mezi změřenou a odhadovanou hodnotou *RSRP* se pohyboval v rozmezí od 5 dBm do 14 dBm. Většina metod vykazovala odchylku odhadu okolo 10 dBm. Mezi metody s nejlepšími výsledky byla zařazená metoda inverzní vážení vzdálenosti a lineární interpolace. U komplikovanějších metod, například u krigingu, se jedná variantu s exponenciálním variogramem. Opakovaným časovým testem bylo zjištěno, že téměř 70 % získaných dat se liší v rozmezí síly signálu *RSRP* o 5 dBm ve stejných lokalitách.

Přesné odhadování síly signálu *RSRP* pouze za pomoci datových sad se ukázalo být velice obtížné. Informace v datových sadách jsou náhodně rozmístěné v prostoru i čase. Není známo, jestli měření probíhalo ve venkovním prostředí, či v budově. Protože se jedná o open-source databázi, samotná měření byla prováděné různými zařízeními s odlišnými parametry antén a za nespécifikovaných podmínek. Plošná hustota měření byla rozmístěna nerovnoměrně a obecně pokrytí bylo příliš řídké na korelované odhady.

Pro získání lepších výsledků odhadu by bylo nutné zajistit co nejrovnoměrnější rozložení známých bodů. Taktéž by se měl zvětšit počet známých měření na km², alespoň trojnásobně, v porovnání s daty získanými ve Vídni. Měření by měla probíhat v co nejkratším čase aby se eliminoval vliv změny stavu sítě za dobu měření. Měl by být omezený pohyb osob a vozidel kolem právě probíhajícího měření, které může razantně změnit výslednou úroveň síly signálu. Pro zlepšení modelu šíření by bylo vhodné pracovat s veřejně dostupnými informacemi o lokacích BTS stanic v nejbližší lokalitě.

Literatura

- [1] *Stupně otevřenosti otevřených dat a česká legislativa*. Otevřená data [online]. 2020 [cit. 2020-10-27] Dostupné z: <https://opendata.gov.cz/informace:stupn%C4%9B-otev%C5%99enosti-datov%C3%BDch-sad>.
- [2] WIMMER, Leonhard. *Open-RMBT* [online]. [cit. 2020-10-29]. Dostupné z: <https://github.com/rtr-nettest/open-rmbt>
- [3] WIMMER, Leonhard. *Platform for Measuring Mobile Broadband Performance: Analysis and Implementation*. Wien, 2019. Diploma Thesis. TU Wien.
- [4] WALACH, D. *Koexistence systémů DVB-T2 a LTE v pásmu 800 MHz*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2019. 66 s. Vedoucí bakalářské práce doc. Ing. Ladislav Polák, Ph.D..
- [5] *Souřadnicové systémy, Tvar zemského tělesa a referenční plochy* [online]. , 5 [cit. 2020-12-01]. Dostupné z: <https://kgm.zcu.cz/studium/gen1/html/ch02s03.html>
- [6] WEINTRIT, A. *So, What is Actually the Distance from the Equator to the Pole? – Overview of the Meridian Distance Approximations* [online]. June 2013, , 14 [cit. 2020-11-19]. Dostupné z: doi:10.12716/1001.07.02.14
- [7] *Calculate distance, bearing and more between Latitude/Longitude points* [online]. [cit. 2021-01-30]. Dostupné z: <http://www.movable-type.co.uk/scripts/latlong.html>
- [8] WINARNO, E., HADIKURNIAWATI, W. and ROSSO, R. N., "Location based service for presence system using haversine method," *2017 International Conference on Innovative and Creative Information Technology (ICITech)* Salatiga, 2017, pp. 1-4, doi: 10.1109/INNOCIT.2017.8319153.
- [9] REHÁK, Ondrej. *Vyhodnocení vlastností mobilních sítí na základě dlouhodobých měření*. Brno, 2020, 59 s. Diplomová práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav radioelektroniky. Vedoucí práce: doc. Ing. Martin Slanina, Ph.D
- [10] <https://gsmweb.cz/mapa/>
- [11] Mgr. Ing. arch. Petr Kurfürst, Ph.D. *Interpolace a regrese* [online]. , 11 [cit. 2020-12-05]. Dostupné z: https://is.muni.cz/do/sci/KTFaFyz/el/pocetni_praktikum/apC_s02.html

- [12] SUTHAHARAN, S., "A correlation-based subspace analysis for data confidentiality and classification as utility in CPS, "2016 IEEE Conference on Communications and Network Security (CNS) Philadelphia, PA, 2016, pp. 426-431, doi: 10.1109/CNS.2016.7860531.
- [13] TYSLIAK, W., "Regression Analysis of Intrinsic Linear Models with Automated Transformations of Monotone Predictors,"2005 IEEE *Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*, Sofia, 2005, pp. 620-623, doi: 10.1109/IDAACS.2005.283059.
- [14] SPIESS, AN., NEUMEYER, N. *An evaluation of R2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach*. BMC Pharmacol 10, 6 (2010). <https://doi.org/10.1186/1471-2210-10-6>
- [15] LUO, X., XU, Y. and SHI, Y., "Comparison of interpolation methods for spatial precipitation under diverse orographic effects,"2011 *19th International Conference on Geoinformatics*, Shanghai, 2011, pp. 1-5, doi: 10.1109/GeoInformatics.2011.5980666.
- [16] LANGHAMMER, Jakub. *3D analýza dat, Proces interpolace* [online]. In: . s. 18 [cit. 2021-04-07]. Dostupné z: https://web.natur.cuni.cz/langhamr/lectures/vtfg2/prednasky/surfer_1/Surfer_prednaska_1_2012.pdf
- [17] YUE, Y., SHEN, J. and LIU, R., "An Improved Adaptive Weighted Gaussian Nearest Neighbor Classification Method,"2019 *Chinese Control And Decision Conference (CCDC)*, Nanchang, China, 2019, pp. 2712-2715, doi: 10.1109/CCDC.2019.8832849.
- [18] HORMANN, Kai. *Barycentric Interpolation* [online]. 2014 [cit. 2021-5-4]. Dostupné z: <https://www.inf.usi.ch/hormann/papers/Hormann.2014.BI.pdf>
- [19] DENGWEN Z., "An edge-directed bicubic interpolation algorithm, "2010 *3rd International Congress on Image and Signal Processing Yantai*, 2010, pp. 1186-1189, doi: 10.1109/CISP.2010.5647190.
- [20] MANN, Stephan. *Cubic precision Clough-Tocher interpolation: Research Report CS-98-15* [online]. In: . s. 11 [cit. 2021-5-4]. Dostupné z: <https://cs.uwaterloo.ca/research/tr/1998/15/CS-98-15.pdf>
- [21] How inverse distance weighted interpolation works. *Argis Pro*[online]. [cit. 2020-12-08]. Dostupné z: <https://pro.arcgis.com/en/pro-app/help/analysis/geostatistical-analyst/how-inverse-distance-weighted-interpolation-works.htm>

- [22] RAPPAPORT, T., *Wireless Communications: Principles and Practice* 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2001
- [23] RAMESH, Rakshit, Mukunth ARUNACHALAM, Hari KRISHNAATLURI, Chetan KUMAR S, S.V.R. ANAND, Paventhan ARUMUGAM a Bharadwaj. *LoRaWAN for smart cities: experimental study in a campus deployment* . 2020. ISBN 9780128188804. Dostupné také z: <https://www.sciencedirect.com/science/article/pii/B9780128188804000168>
- [24] *LTE Radio Link Budgeting and RF Planning* [online]. [cit. 2021-03-15]. Dostupné z: <https://sites.google.com/site/lteencyclopedia/lte-radio-link-budgeting-and-rf-planning>
- [25] FAHRMEIR, Ludwig, Thomas KNEIB, Stefan LANG a Brian MARX. Regression [online]. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013 [cit. 2020-05-19]. DOI: 10.1007/978-3-642-34333-9. ISBN 978-3-642-34332-2.
- [26] *Kriging* [online]. [cit. 2020-12-7]. Dostupné z: https://vsp.pnnl.gov/help/Vsample/Kriging_Variogram_Model.htm
- [27] CHUNG, Sang Yong, S. VENKATRAMANAN, Hussam Eldin ELZAIN, S. SELVAM, M.V. PRASANNA,. *Chapter 4 - Supplement of Missing Data in Groundwater-Level Variations of Peak Type Using Geostatistical Methods*, [online]. , Pages 33-41 [cit. 2020-12-09]. ISSN ISBN 9780128154137. Dostupné z: [doi:https://doi.org/10.1016/B978-0-12-815413-7.00004-3](https://doi.org/10.1016/B978-0-12-815413-7.00004-3).

Seznam symbolů, veličin a zkratek

BTS	Base Transceiver Station
ČTÚ	Český telekomunikační úřad
IDW	Inverse distance weighing
LTE	Long Term Evolution
OFDM	Orthogonal Frequency Division Multiplexing
OK	Ordinary Kriging
PLE	Path Loss Exponent
RSRP	Reference Signal Receive Power
RSRQ	Reference Signal Received Quality
RSSI	Received Signal Strength Indicator
RTR	Rakouský regulační úřad pro vysílání a telekomunikace
TLS	Transport Layer Security
UK	Universal Kriging
WGS 84	World Geodetic System 1984