

UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA

DIPLOMOVÁ PRÁCE

Modelování množství srážek pomocí metody
kriging



Katedra matematické analýzy a aplikací matematiky
Vedoucí diplomové práce: **doc. RNDr. Eva Fišerová, Ph.D.**
Vypracoval(a): **Bc. Lucie Chwistková**
Studijní program: N1103 Aplikovaná matematika
Studijní obor Aplikace matematiky v ekonomii
Forma studia: prezenční
Rok odevzdání: 2016

BIBLIOGRAFICKÁ IDENTIFIKACE

Autor: Bc. Lucie Chwistková

Název práce: Modelování množství srážek pomocí metody kriging

Typ práce: Diplomová práce

Pracoviště: Katedra matematické analýzy a aplikací matematiky

Vedoucí práce: doc. RNDr. Eva Fišerová, Ph.D.

Rok obhajoby práce: 2016

Abstrakt: Tato práce se zabývá geostatistickým zpracováním prostorových dat. Cílem práce je popsat a následně aplikovat metody Českého hydrometeorologického ústavu při odhadování množství srážek. Zpracování dat bylo provedeno ve statistickém softwaru R.

Klíčová slova: Kriging, variogram, metoda inverzní vážené vzdálenosti, srážky, R software, Český hydrometeorologický ústav

Počet stran: 61

Počet příloh: 1

Jazyk: český

BIBLIOGRAPHICAL IDENTIFICATION

Author: Bc. Lucie Chwistková

Title: Modelling of precipitation amount using kriging

Type of thesis: Master's thesis

Department: Department of Mathematical Analysis and Application of Mathematics

Supervisor: doc. RNDr. Eva Fišerová, Ph.D.

The year of presentation: 2016

Abstract: This thesis deals with geostatistical processing of spatial data. The aim of this thesis is to describe and apply the methods used by the Czech hydrometeorological office for estimating precipitation amount. Statistical software R was used for data processing.

Key words: Kriging, variogram, inverse distance weighting, precipitation, R software, Czech hydrometeorological office

Number of pages: 61

Number of appendices: 1

Language: Czech

Prohlášení

Prohlašuji, že jsem diplomovou práci zpracovala samostatně pod vedením paní doc. RNDr. Evy Fišerové, Ph.D. a všechny použité zdroje jsem uvedla v seznamu literatury.

V Olomouci dne 18. dubna 2016

.....
podpis

Obsah

Úvod	10
1 Sběr a zpracování dat	11
1.1 Český hydrometeorologický ústav	11
1.2 Použitý souřadnicový systém	12
2 Geostatistika	14
2.1 Metoda inverzní vzdálenosti	15
2.2 Prostorová proměnná	17
2.2.1 Regionalizovaná proměnná	18
2.2.2 Náhodná funkce	18
2.2.3 Stacionarita	20
2.2.4 Prostorová kovarianční funkce	22
2.3 Variogram	24
2.3.1 Experimentální variogram	25
2.4 Teoretický variogram	25
2.4.1 Dosah a oblast vlivu	27
2.4.2 Chování kolem počátku	28
2.4.3 Modely variogramu	29
2.4.4 Anizotropie	31
3 Kriging	34
3.1 Odvození krigingových rovnic	34
3.2 Typy krigingu	35
3.2.1 Základní kriging	35
3.2.2 Jednoduchý kriging	38
3.2.3 Univerzální kriging	39
3.2.4 Lognormální kriging	40

4 Praktická část	41
4.1 Průzkumová analýza dat	41
4.2 Metoda IDW	43
4.3 Metoda Kriging	45
4.3.1 Stacionarita	46
4.3.2 Modelování variogramu	48
4.3.3 Kriging	50
4.3.4 Křížová validace	54
Závěr	58
Literatura	60

Seznam obrázků

1.1	Měřicí stanice - zeměpisné souřadnice	13
1.2	Měřicí stanice - systém S-42	13
2.1	Rozhodovací strom pro volbu interpolační metody	14
2.2	IDW: ilustrační příklad	16
2.3	Regionalizovaná proměnná jako realizace náhodné funkce	19
2.4	Teoretický variogram	27
2.5	Variogram: nugget model	29
2.6	Variogram: sférický model	29
2.7	Variogram: exponenciální model	30
2.8	Variogram: gaussovský model	30
2.9	Variogram: sinový model	31
2.10	Variogramy: geometrická anizotropie	31
2.11	Elipsy: hlavní a vedlejší osy v případě geometrické anizotropie	32
4.1	Histogram: původní data (vlevo), zlogaritmovaná data (vpravo)	42
4.2	3D bodový graf - zlogaritmovaná data	43
4.3	IDW - srážková mapa, $p = 2$	44
4.4	IDW - srážková mapa, $p = 10$	44
4.5	Geodata souhrnně	45
4.6	Konstantní trend - rezidua, zeměpisná šířka	46
4.7	Konstantní trend - rezidua, zeměpisná délka	46
4.8	Lineární trend - rezidua, zeměpisná šířka	47
4.9	Lineární trend - rezidua, zeměpisná délka	47
4.10	Polynom 4. řádu - rezidua, zeměpisná šířka	48
4.11	Polynom 4. řádu - rezidua, zeměpisná délka	48
4.12	Směrový variogram	49
4.13	Variogram - fitovaný anizotropní model	50
4.14	Srážková mapa - univerzální kriging (polynom 4. řádu)	51
4.15	Směrodatné odchylky krigingu - univerzální kriging (polynom 4. řádu)	51
4.16	Srážková mapa - základní kriging	52
4.17	Srážková mapa - univerzální kriging (lineární trend)	52

4.18	Směrodatné odchytky krigingu - základní kriging	52
4.19	Směrodatné odchytky krigingu - univerzální kriging (lineární trend)	52
4.20	Pozorované, predikované hodnoty - logaritmovaná data	56
4.21	Pozorované, predikované hodnoty - odlogaritmovaná data	56
4.22	Pozorované hodnoty, rezidua - odlogaritmovaná data	57
4.23	Odhadnuté hodnoty, rezidua - odlogaritmovaná data	57

Poděkování

Na tomto místě bych ráda poděkovala vedoucí mé diplomové práce doc. RNDr. Evě Fišerové, Ph.D. za veškeré rady a drahocenný čas, který mi v průběhu práce věnovala. Dále chci poděkovat RNDr. Radimovi Tolaszovi, Ph.D. z Českého hydrometeorologického ústavu za ochotné poskytnutí cenných informací. Poděkování patří v neposlední řadě mým blízkým, kteří mě po celou dobu mého studia plně podporují.

Úvod

Počasí je nedílnou součástí života každého z nás. Ovlivňuje spoustu našich aktivit ať už volnočasových, tak profesních. Každý z nás jistě někdy sleduje předpověď počasí, aby si co nejlépe naplánoval např. volný víkend. Pro zemědělce či řidiče pak může být vývoj počasí přímo klíčový.

Proto je jistě užitečné, když už počasí nemůžeme ovlivnit, mu alespoň porozumět, a tak se na něj připravit. A právě tímto se bude zabývat tato práce. Pro tuto analýzu jsem si vybrala hydrometeorologická data, a to úhrny srážek. Na území České republiky zajišťuje pravidelná měření srážek Český hydrometeorologický ústav (ČHMÚ). Měření provádí síť srážkoměrných stanic a protože není možné změřit množství srážek v každém bodě, jsou pro zbylé lokality tyto hodnoty odhadovány na základě známých hodnot v jejich okolí. Cílem práce bude vytvořit mapu úhrnů srážek v České republice za měsíc listopad 2015.

Protože se jedná o data, která se vztahují ke konkrétním místům, tzv. geodata, budou základní nástroje pro jejich zpracování z oblasti geostatistiky. Geostatistika nachází v současné době široké využití, ať už v oblasti ropné geologie, hydrologie, geochemie, hutnictví či krajinné ekologii, epidemiologii, zemědělství, atp. Její počátky však sahají již do 50. let 20. století, kdy byla využívána pro odhad rozložení ložisek rud pro těžební průmysl.

Nejprve se seznámíme s nejpoužívanějšími geostatistickými metodami a se základními pojmy s nimi souvisejícími jako je variogram, náhodná funkce, regionalizovaná proměnná, atp. Následně tyto metody aplikujeme na reálná data poskytnutá Českým hydrometeorologickým ústavem.

1 Sběr a zpracování dat

K praktické části mé diplomové práce jsem čerpala data výhradně ze stránek Českého hydrometeorologického ústavu v [9].

1.1. Český hydrometeorologický ústav

Historie Českého hydrometeorologického ústavu (ČHMÚ) sahá až do roku 1920, kdy vznikl rozhodnutím ministerské rady Československé republiky Meteorologický ústav, který působil pouze jako meteorologický ústav. Později došlo ke začlenění hydrologie do působnosti tohoto ústavu a s rostoucím významem ochrany životního prostředí také ochrana čistoty ovzduší, čímž od roku 1967 funguje v podobě, v jaké jej známe dnes. [15]

ČHMÚ v současné době poskytuje široké spektrum informací týkajících se meteorologie, klimatologie, hydrologie a dalších. Pro širokou veřejnost jsou data upravená a předzpracovaná ke snadnější a rychlejší orientaci. Pro účely této práce byla však vhodnější data neupravená. Nejprve bylo nutné nashromáždit denní úhrny srážek za určitý časový úsek. Já jsem si zvolila listopad 2015. ČHMÚ explicitně neposkytuje souřadnice srážkoměrných stanic, ty mi byly poskytnuty na požádání. Ne všechny stanice vykazují tato data v průběhu celého časového období, proto byly stanice s chybějícími hodnotami vyřazeny. Ve výsledku jsem pracovala s daty naměřenými v síti 319 stanic, které udávaly denní úhrny srážek za celé období a pro které byly dostupné přesné souřadnice.

ČHMÚ poskytuje 2 typy plošných odhadů srážek. Prvním z nich je kombinace radarového odhadu, který je upřesňován měřeními z pozemních srážkoměrů. Od roku 2011 přibyl druhý typ plošného odhadu srážek založený pouze na měřeních

z pozemních srážkoměrů. Každá z metod má svá úskalí. Nevýhodou odhadů založených hlavně na radarovém měření je, že se ve výsledné mapě mohou objevovat falešné (nemeteorologické) odrazy. Naopak nevýhodou plošných odhadů založených na měřeních pozemních srážkoměrů je nízká pravděpodobnost, že zachytí izolované hydrometeorologické jevy. [10] Výhodou druhého ze zmiňovaných typů odhadů je, že reflektují jen to, co bylo skutečně naměřeno.

Z praktických důvodů se budu v této práci zabývat plošnými odhady srážek založenými pouze na měřeních z pozemních srážkoměrů. V tomto případě ČHMÚ používá dva druhy interpolačních metod. První z nich je metoda kriging, která pro své častější využití bude tvořit stěžejní část práce. Druhou z metod je metoda vážené inverzní vzdálenosti (IDW), kterou ČHMÚ používá v případě, že je k dispozici málo nenulových srážek.

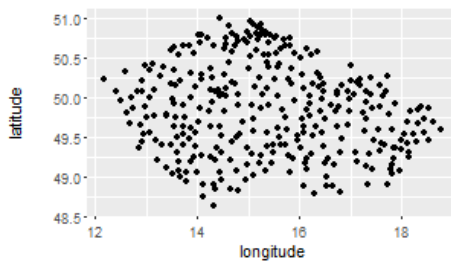
1.2. Použitý souřadnicový systém

Cílem souřadnicových systémů obecně je, že chceme všem bodům na Zemi přiřadit jednoznačnou polohu, která umožňuje tyto body zobrazit na mapě. Protože není snadné matematicky popsat tvar Země, aproximujeme její referenční plochou. Obvykle jde o určitý druh elipsoidu. Dále je třeba zvolit typ zobrazení referenční plochy na rozvinutelnou plochu tak, abychom získali rovinnou mapu. [14]

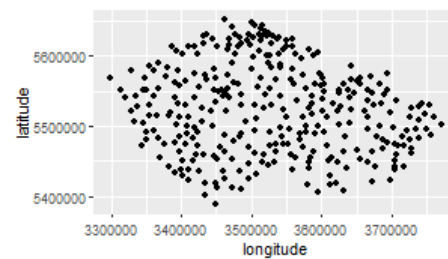
Já jsem si v mé práci zvolila souřadnicový systém S-42, který používá jako referenční plochu tzv. Krasovského elipsoid. Samotné zobrazení bodů z této referenční plochy na rozvinutelnou plochu, která je v tomto případě eliptický válec, probíhá tak, že rozdělíme povrch elipsoidu na pásy o šesti stupních zeměpisné délky, tj. první pás se nachází mezi nultým a šestým poledníkem, druhý pás mezi šestým a dvanáctým poledníkem, atp. Každý pás je přesně určen svým číslem a navíc středním poledníkem, který leží uprostřed pásu, např. pás číslo šest má střední poledník 33, protože $6 \times 6 = 36$, $36 - 6 = 30$, šířka pásu je tedy $\langle 30, 36 \rangle$ a středem tohoto intervalu je číslo 33. Následně se odpovídající pás obalí eliptickým válcem, který se dotýká referenčního elipsoidu v místě středního poledníku. Zobrazení pak provedeme pomocí pravoúhlého promítnutí bodů z elipsoidu na válcovou plochu.

Jejím následným narovnáním již získáme rovinnou mapu uvažovaného pásu.

Protože se na narovnané válcové ploše odečítá vzdálenost v metrech východně od středního poledníku, dostali bychom pro polohy západně od středního poledníku záporné souřadnice. Tomu se chceme vyhnout, proto ještě přičteme konstantnu 500 000. Aby byla informace o poloze jednoznačná, připisuje se ještě před takto získané souřadnice číslo pásu. [16]



Obrázek 1.1: Měřicí stanice - zeměpisné souřadnice



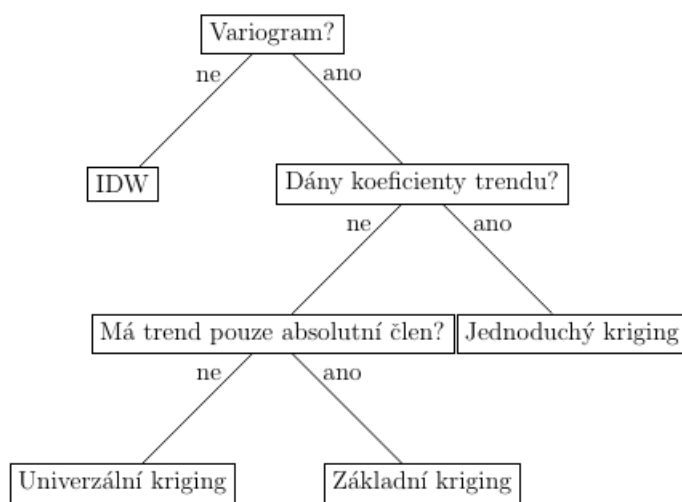
Obrázek 1.2: Měřicí stanice - systém S-42

V obrázku 1.1 můžeme vidět síť použitých srážkoměrných stanic v zeměpisných souřadnicích a na obrázku 1.2 také po převedení do souřadnicového systému S-42, s nímž budu dále pracovat. Můžeme si povšimnout, že převodem z jednoho souřadnicového systému na druhý nedochází ke zdatelné deformaci.

2 Geostatistika

Jak už název napovídá, geostatistika je odvětví statistiky zabývající se zemí, tedy statistika s geografickým kontextem. Toto odvětví statistiky zahrnuje techniky s důrazem na data vztahující se k určitému území. Geostatistika je silně spjata se základními interpolačními metodami. Metody geostatistiky stojí na statistickém modelu, jehož základem je teorie náhodné funkce, která slouží k modelování nejistoty související s prostorovými odhady.

Mezi jednodušší interpolační metody patří metoda vážené inverzní vzdálenosti (IDW). Ačkoliv je metoda IDW v práci začleněna pod geostatistické metody, je třeba podotknout, že se jedná pouze o numerickou metodu. Schéma na obrázku 2.1 znázorňuje rozhodovací strom pro volbu vhodné interpolační metody v rámci těch metod, kterými se tato práce zabývá.



Obrázek 2.1: Rozhodovací strom pro volbu interpolační metody

2.1. Metoda inverzní vzdálenosti

Metoda inverzní vzdálenosti je deterministická nelineární interpolační metoda, která používá vážený průměr hodnot sledovaného atributu z okolí bodu odhadu, za účelem odhadnout hodnotu daného atributu v lokalitách, kde nebylo provedeno měření. Jako odhadovaný atribut si můžeme představit například obsah organické hmoty v půdě či množství srážek spadných v určité lokalitě. Druhý z příkladů odpovídá mému případu. Váha příslušného měření závisí na vzdálenosti od bodu, ve kterém chceme provést odhad. Název této metody tedy plyne z toho, že podobnost hodnot určitého znaku by měla klesat s rostoucí vzdáleností mezi body měření.

Při definování určitého okolí kolem odhadovaného bodu můžeme postupovat dvěma způsoby:

- zvolit fixní tvar okolí,
- určit pevně daný počet nejbližších bodů, ve kterých bylo provedeno pozorování. [11]

Odhad hodnoty sledovaného atributu v určitém bodě \mathbf{x}_0 je počítán dle následujícího vztahu:

$$\hat{Z}(\mathbf{x}_0) = \frac{\sum_{i=1}^n w(\mathbf{x}_i) Z(\mathbf{x}_i)}{\sum_{i=1}^n w(\mathbf{x}_i)}, \quad (2.1)$$

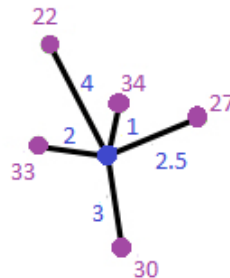
kde váhy daného pozorování jsou počítány jako reciproká hodnota p -té mocniny Eukleidovské vzdálenosti k bodu interpolace

$$w(\mathbf{x}_i) = \|\mathbf{x}_i - \mathbf{x}_0\|^{-p}. \quad (2.2)$$

$\hat{Z}(\mathbf{x}_0)$ značí odhadovanou hodnotu daného atributu v bodě \mathbf{x}_0 , $Z(\mathbf{x}_i)$ naměřené hodnoty daného atributu v bodech \mathbf{x}_i , $i=1, \dots, n$ a p je kladné reálné číslo, tzv. silový parametr. Silový parametr p je definován uživatelem, nejčastější volbou je $p = 2$. Silový parametr určuje, do jaké míry jsou pro odhad preferována blízká pozorování oproti těm vzdálenějším. [2]

Metoda IDW je populární především pro svoji jednoduchost a rychlost výpočtů. Další výhodou je, že jsme schopni přesně kontrolovat vliv vzdáleností. Skýtá však také nevýhody, které již v dnešní době byly překonány složitějšími interpolačními metodami. Jednou z nevýhod této metody je, že váhy jsou z intervalu $\langle 0, 1 \rangle$, z čehož plyne, že interpolované hodnoty budou vždy v rozmezí naměřených hodnot. IDW při odhadování navíc zcela ignoruje prostorové uspořádání dat, což vede k nežádoucím efektům v případě, že struktura celé oblasti měření není homogenní a liší se např. v závislosti na směru, nikoli jen na vzdálenosti. Metoda je založena na místní variabilitě dat, je tedy vysoce citlivá na volbu vhodného okolí odhadovaného bodu. [11]

Příklad 1 Uvažujme situaci zobrazenou na obrázku 2.2, kdy chceme dopočítat hodnotu atributu na základě několika málo okolních pozorování metodou IDW. Úkolem je odhadnout hodnotu atributu v bodě označeném modře. Výpočet je proveden pro různé volby silového parametru p tak, aby ilustroval vliv hodnoty tohoto parametru na výslednou hodnotu odhadu.



Obrázek 2.2: IDW: ilustrační příklad

Naměřené hodnoty atributu v bodech označených fialově nabývají hodnot 22, 34, 27, 30 a 33. Na obrázku je navíc modře znázorněna vzdálenost každého z měření k bodu odhadu. Uvedeným měřením přísluší vzdálenosti 4, 1, 2.5, 3 a 2 postupně.

- Pro $p = 1$:

$$\hat{Z}(\mathbf{x}_0) = \frac{\frac{1}{2^1}33 + \frac{1}{3^1}30 + \frac{1}{2.5^1}27 + \frac{1}{1^1}34 + \frac{1}{4^1}22}{\frac{1}{2^1} + \frac{1}{3^1} + \frac{1}{2.5^1} + \frac{1}{1^1} + \frac{1}{4^1}} = 30.92617.$$

- Typická volba $p = 2$:

$$\hat{Z}(\mathbf{x}_0) = \frac{\frac{1}{2^2} 33 + \frac{1}{3^2} 30 + \frac{1}{2.5^2} 27 + \frac{1}{1^2} 34 + \frac{1}{4^2} 22}{\frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{2.5^2} + \frac{1}{1^2} + \frac{1}{4^2}} = 32.38063.$$

- Nyní pro $p = 5$:

$$\hat{Z}(\mathbf{x}_0) = \frac{\frac{1}{2^5} 33 + \frac{1}{3^5} 30 + \frac{1}{2.5^5} 27 + \frac{1}{1^5} 34 + \frac{1}{4^5} 22}{\frac{1}{2^5} + \frac{1}{3^5} + \frac{1}{2.5^5} + \frac{1}{1^5} + \frac{1}{4^5}} = 33.87473.$$

Vidíme, že s rostoucím parametrem p roste i váha nejbližšího pozorování. V posledním případě má zřejmě největší vliv na odhad pozorování s hodnotou 34, kterému přísluší nejmenší vzdálenost k bodu odhadu, a to 1. Pro velké hodnoty tohoto parametru metoda IDW konverguje k interpolaci nejbližším sousedem (známým pozorováním).

2.2. Prostorová proměnná

V mnoha vědních disciplínách je třeba data chápat včetně prostorového či časového kontextu. Protože geostatistická data poskytují informace o proměnné (či více proměnných), jejichž chování chceme pro danou oblast prostoru charakterizovat, zavádí se v geostatistické analýze pojem prostorová (regionalizovaná) proměnná, jejíž rozložení hodnot chceme modelovat. V pravděpodobnostním modelu je regionalizovaná proměnná $z(\mathbf{x})$ považována za realizaci náhodné funkce $Z(\mathbf{x})$. Výhodou tohoto přístupu je, že v mnohých případech postačuje charakterizovat základní vlastnosti náhodné funkce $Z(\mathbf{x})$ a ne vlastnosti realizací $z(\mathbf{x})$. [6]

V nejobecnější formě můžeme náš datový soubor zapsat následovně:

$$\begin{bmatrix} t_1 & x_1^1 & x_1^2 & x_1^3 & z_1^1 & \dots & z_1^i & \dots & z_1^N \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ t_\alpha & x_\alpha^1 & x_\alpha^2 & x_\alpha^3 & z_\alpha^1 & \dots & z_\alpha^i & \dots & z_\alpha^N \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ t_n & x_n^1 & x_n^2 & x_n^3 & z_n^1 & \dots & z_n^i & \dots & z_n^N \end{bmatrix} \quad (2.3)$$

Řádky matice představují jednotlivé vzorky, jejichž celkový počet je n . Proměnné jsou značeny indexem i , celkový počet proměnných je N . Jednotlivé vzorky mohou být pořizeny v různých časech t_α , $\alpha = 1, \dots, n$, a na různých místech \mathbf{x}_α , kde \mathbf{x}_α je typicky třírozměrný vektor obsahující souřadnice daného bodu $(x_\alpha^1, x_\alpha^2, x_\alpha^3)$. [5], [6]

2.2.1. Regionalizovaná proměnná

Předpokládejme nyní, že na objektech v prostoru je měřena pouze jedna proměnná z (v obecném zápise tedy $N = 1$) a čas, ve kterém byla měření provedena není zaznamenáván. Mějme n pozorování

$$z(\mathbf{x}_\alpha), \alpha = 1, \dots, n, \quad (2.4)$$

pořízená v bodech \mathbf{x}_α v daném regionu \mathcal{D} . Teoreticky je možné provést nekonečně mnoho měření v daném regionu, což je v praxi neproveditelné a mnohdy z důvodu nákladnosti měřících metod se počet pozorování snažíme redukovat. Za předpokladu, že by bylo možné provést nekonečně mnoho měření dané proměnné v určitém regionu, mohli bychom vynechat index α a definovat regionalizovanou proměnnou (REV) jako:

$$z(\mathbf{x}), \forall \mathbf{x} \in \mathcal{D}. \quad (2.5)$$

Datová sada $\{z(\mathbf{x}_\alpha), \alpha = 1, \dots, n\}$ je nyní chápána jako sada hodnot REV. [6]

2.2.2. Náhodná funkce

Každá naměřená hodnota v datové sadě přísluší určitému bodu ve zvoleném regionu \mathcal{D} , proto je nazývána regionalizovaná hodnota. Podívejme se nyní na problematiku geostatistických dat z úhlu pohledu, kdy považujeme regionalizované hodnoty za výsledek náhodného mechanismu. Formálně tento mechanismus nazýváme náhodná proměnná a výběrová hodnota $z(\mathbf{x}_\alpha)$ reprezentuje realizaci náhodné proměnné $Z(\mathbf{x}_\alpha)$. V každém bodě \mathbf{x}_α mechanismus zodpovědný za generování hodnot $z(\mathbf{x}_\alpha)$ může být odlišný, proto $Z(\mathbf{x}_\alpha)$ může mít různé vlastnosti v každém z bodů daného regionu \mathcal{D} .

Tímto se dostáváme k pojmu náhodné funkce. Uvažujme regionalizované hodnoty ve všech bodech daného regionu, přidružená funkce $z(\mathbf{x})$ pro $\mathbf{x} \in \mathcal{D}$ je regionalizovaná proměnná. Množinu hodnot

$$\{z(\mathbf{x}), \mathbf{x} \in \mathcal{D}\} \quad (2.6)$$

můžeme považovat za jeden výběr z nekonečné množiny náhodných proměnných (jedna náhodná proměnná v každém z bodů regionu).

Definice Třída náhodných proměnných (veličin)

$$\{Z(\mathbf{x}), \mathbf{x} \in \mathcal{D}\} \quad (2.7)$$

se nazývá náhodná funkce (RAF).

$$\begin{array}{ccc} RAF & Z(\mathbf{x}) \longrightarrow & Z(\mathbf{x}_0) \\ \downarrow & \downarrow & \downarrow \\ REV & z(\mathbf{x}) \longrightarrow & z(\mathbf{x}_0) \end{array}$$

Obrázek 2.3: Regionalizovaná proměnná jako realizace náhodné funkce

Na obrázku 2.3 je schéma, které ilustruje, jak je náhodná funkce založena na dvou různých přístupech. Prvním z aspektů je, že data pocházejí z určitého prostoru (1-2-3D) a jsou závislá na poloze v daném regionu - jsou regionalizovaná. Druhým z nich je fakt, že regionalizované výběrové hodnoty $z(\mathbf{x}_\alpha)$ nemohou být obecně modelovány deterministickou funkcí $z(\mathbf{x})$, ale musí být zvolen pravděpodobnostní přístup, kdy mechanismus generující data je považován za náhodný. RAF je tedy koncept spojující regionalizovanost a náhodnost. V našem schématu je tedy REV $z(x)$ jednou realizací RAF $Z(\mathbf{x})$. Regionalizovaná hodnota $z(\mathbf{x}_0)$ v konkrétní oblasti \mathbf{x}_0 je realizací náhodné proměnné $Z(\mathbf{x}_0)$, která samotná je součástí třídy náhodných proměnných. [5], [6]

2.2.3. Stacionarita

Náhodný mechanismus $Z(\mathbf{x}_0)$ působící v bodě \mathbf{x}_0 daného regionu \mathcal{D} generuje realizace, které se řídí rozdělením pravděpodobnosti s distribuční funkcí $F_{\mathbf{x}_0}$:

$$P(Z(\mathbf{x}_0) \leq z) = F_{\mathbf{x}_0}(z), \quad (2.8)$$

kde P je pravděpodobnost, že hodnota Z v bodě \mathbf{x}_0 je menší nebo rovna než určitá hodnota z .

Definici distribuční funkce můžeme rozšířit na dvourozměrný případ, kdy máme dvě náhodné proměnné $Z(\mathbf{x}_1)$, $Z(\mathbf{x}_2)$ ve dvou různých bodech

$$P(Z(\mathbf{x}_1) \leq z_1, Z(\mathbf{x}_2) \leq z_2) = F_{\mathbf{x}_1, \mathbf{x}_2}(z_1, z_2), \quad (2.9)$$

nebo obecně pro n -rozměrný případ, kdy máme n náhodných proměnných v n různých bodech

$$P(Z(\mathbf{x}_1) \leq z_1, \dots, Z(\mathbf{x}_n) \leq z_n) = F_{\mathbf{x}_1, \dots, \mathbf{x}_n}(z_1, \dots, z_n), \quad (2.10)$$

kde P je sdružená pravděpodobnost, že $Z(\mathbf{x}_i)$ je menší nebo rovno než z_i , pro $\forall i = 1, \dots, n$ současně.

Tímto způsobem lze vytvořit obecný model, kterým lze popsat jakýkoli přírodní či technologický proces. V praxi však máme jen určité množství dat, která jsou jednou či více realizacemi náhodné funkce, z nichž není možné odvodit všechny jednorozměrné či mnohorozměrné distribuční funkce pro jakoukoli množinu bodů. Je nutné přijmout určité omezení, kterým bude předpoklad stacionarity. [6]

Striktní stacionarita

Definice Řekneme, že náhodná funkce $Z(\mathbf{x})$ je striktně stacionární, jestliže pro jakoukoli n -tici bodů $\mathbf{x}_1, \dots, \mathbf{x}_n$, $n \in N$ a pro jakýkoli vektor \mathbf{h} platí:

$$F_{\mathbf{x}_1, \dots, \mathbf{x}_n}(z_1, \dots, z_n) = F_{\mathbf{x}_1 + \mathbf{h}, \dots, \mathbf{x}_n + \mathbf{h}}(z_1, \dots, z_n). \quad (2.11)$$

Tato definice nám říká, že posun n bodů z jedné části definičního oboru do jiné nemění mnohorozměrnou distribuci. Tuto vlastnost nazýváme invariancí vůči posunutí a týká se náhodné funkce RAF, nikoli regionalizované proměnné.

Striktní stacionarita vyžaduje, aby všechny momenty byly invariantní vůči posunutí. To však není možné ověřit z omezeného množství experimentálních dat. Proto se spokojíme se stacionaritou prvních dvou momentů. [6]

Slabá stacionarita

Jednodušší strategie je uvažovat pouze dvojice bodů $\mathbf{x}_1, \mathbf{x}_2$ z definičního oboru a pokusit se charakterizovat pouze dva první momenty náhodné funkce, ne plnou distribuci. Tato strategie je optimální v případě Gaussovske distribuce, kde první dva momenty plně charakterizují rozdělení. [6] V praxi tento přístup funguje obecně v případě dat, jejichž histogram nemá příliš silné chvosty.

Takovou stacionaritu nazýváme stacionarita druhého řádu nebo také slabá stacionarita. V tomto případě je očekávaná hodnota $E(Z(\mathbf{x}))$ konstantní pro všechny body z dané oblasti \mathcal{D} :

$$E[Z(\mathbf{x})] = m(\mathbf{x}) = m, \forall \mathbf{x} \in \mathcal{D} \quad (2.12)$$

a kovarianční funkce mezi jakoukoli dvojicí bodů $\mathbf{x}, \mathbf{x} + \mathbf{h}$ závisí na vzdálenosti těchto bodů, ne na bodu \mathbf{x} :

$$E[Z(\mathbf{x}) \cdot Z(\mathbf{x} + \mathbf{h})] - m^2 = C(\mathbf{h}), \forall \mathbf{x}, \mathbf{x} + \mathbf{h} \in \mathcal{D}. \quad (2.13)$$

Není třeba dělat předpoklad o rozptylu, protože ten získáme z kovarianční funkce dosazením za $\mathbf{h} = 0$, $C(0)$.

Druhou z možností je předpokládat slabou stacionaritu přírůstků, tj. průměr a rozptyl přírůstků $Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})$ existují a jsou nezávislé na bodu \mathbf{x} :

$$E(Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})) = 0, \quad (2.14)$$

$$var(Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})) = 2\gamma(\mathbf{h}). \quad (2.15)$$

Toto platí pro jakékoli dva body z \mathcal{D} , které jsou odděleny stejnou vzdáleností a směrem, bez ohledu na polohu těchto dvou bodů v definičním oboru. Tuto stacionaritu nazýváme vnořenou stacionaritou a jak si ukážeme později, vede k zavedení pojmu semivariogram. [1], [3]

Stacionaritou můžeme tedy rozumět, že případná závislost prostorové proměnné sledované ve dvou místech neplyne z konkrétního místa měření, ale ze vzdálenosti těchto míst.

2.2.4. Prostorová kovarianční funkce

Kovarianční funkce $C(\mathbf{h})$ je založena na předpokladu stacionarity prvních dvou momentů náhodné funkce.

Základní vlastnosti prostorové kovarianční funkce [1]:

1.

$$C(0) = \sigma^2, \quad (2.16)$$

kde σ^2 je obecné označení rozptylu náhodné veličiny, který můžeme vyjádřit jako $\sigma^2 = \text{var}[Z(\mathbf{x})] = E[Z(\mathbf{x}) - E(Z(\mathbf{x}))]^2$.

Důkaz. Pro stacionární veličiny existuje průměr m . Kovarianční funkce je definovaná takto:

$$C(\mathbf{h}) = E[Z(\mathbf{x} + \mathbf{h}) - m][Z(\mathbf{x}) - m],$$

dosadíme za $\mathbf{h} = 0$, z čehož plyne

$$C(0) = E[Z(\mathbf{x} + 0) - m][Z(\mathbf{x}) - m] = \sigma^2$$

□

2. Kovarianční funkce je sudá

$$C(\mathbf{h}) = C(-\mathbf{h}) \quad (2.17)$$

Důkaz. Podle definice

$$C(-\mathbf{h}) = E[Z(\mathbf{x} - \mathbf{h}) - m][Z(\mathbf{x}) - m],$$

dosadíme za $t = \mathbf{x} - \mathbf{h}$, z čehož plyne

$$C(-\mathbf{h}) = E[Z(t) - m][Z(t + \mathbf{h}) - m] = C(\mathbf{h})$$

□

3. Schwarzova nerovnost

$$|C(\mathbf{h})| \leq C(0) = \text{var}(Z(\mathbf{x})) \quad (2.18)$$

říká, že kovarianční funkce je omezená a její absolutní hodnota nemůže přesáhnout rozptyl.

Důkaz. Chceme dokázat, že

$$C(\mathbf{h}) \leq C(0) \text{ a } -C(0) \leq C(\mathbf{h}).$$

Vycházíme ze vztahu

$$\begin{aligned} 0 &\leq E[(Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x}))^2] \\ 0 &\leq E[Z(\mathbf{x} + \mathbf{h}) - m]^2 + (Z(\mathbf{x}) - m)^2 - 2(Z(\mathbf{x} + \mathbf{h}) - m)(Z(\mathbf{x}) - m)] \\ 0 &\leq 2C(0) - 2C(\mathbf{h}) \\ C(\mathbf{h}) &\leq C(0) \end{aligned}$$

Stejným způsobem bychom získali ze vztahu $0 \leq E[(Z(\mathbf{x} + \mathbf{h}) + Z(\mathbf{x}))^2]$, že platí $C(\mathbf{h}) \geq -C(0)$. Proto $|C(\mathbf{h})| \leq C(0)$. \square

4. Kovarianční funkce dělená rozptylem je korelační funkce

$$\rho(\mathbf{h}) = \frac{C(\mathbf{h})}{C(0)}, \quad (2.19)$$

která je omezená

$$-1 \leq \rho(\mathbf{h}) \leq 1. \quad (2.20)$$

5. Kovarianční funkce je pozitivně semidefinitní, tzn. že při výpočtu rozptylu lineární kombinace $n + 1$ náhodných proměnných $Z(\mathbf{x}_\alpha)$ pomocí kovarianční funkce $C(\mathbf{h})$, je tento rozptyl nezáporný. Z toho vyplývá pozitivní semidefinitní matice \mathbf{C} kovariancí:

$$\underbrace{\text{var} \left(\sum_{\alpha=0}^n w_\alpha Z(\mathbf{x}_\alpha) \right)}_{\geq 0} = \sum_{\alpha=0}^n \sum_{\beta=0}^n w_\alpha w_\beta C(\mathbf{x}_\alpha - \mathbf{x}_\beta) = \mathbf{w}^T \mathbf{C} \mathbf{w} \quad (2.21)$$

pro libovolnou množinu bodů \mathbf{x}_α a váhy w_α uspořádané ve vektoru \mathbf{w} . [6]

2.3. Variogram

Pro zavedení pojmu variogram si nejprve řekneme, jakým způsobem měříme variabilitu regionalizované proměnné $z(\mathbf{x})$. Nepodobnost dvojice výběrových hodnot hodnotíme pomocí čtvercového rozdílu mezi jejich hodnotami děleného dvěma. Takto spočítané nepodobnosti mezi dvojicemi bodů si zobrazíme proti jejich vzdálenostem v geografickém prostoru, výsledkem je tzv. variogram cloud. Takto vytvořené mračno bodů rozdělíme do tříd podle jejich separací v prostoru a průměrné nepodobnosti v každé třídě pak tvoří hodnoty tzv. experimentálního variogramu. [5], [6]

Nyní si tuto hlavní myšlenku popíšeme precizněji. Jak už bylo řečeno, variabilitu regionalizované proměnné $z(\mathbf{x})$ měříme pomocí nepodobnosti mezi dvojicemi datových hodnot, z_α, z_β naměřených v bodech $\mathbf{x}_\alpha, \mathbf{x}_\beta$ v prostoru \mathcal{D} . Měřítko nepodobnosti mezi dvěma hodnotami značíme γ^* a spočteme jej dle:

$$\gamma_{\alpha,\beta}^* = \frac{(z_\alpha - z_\beta)^2}{2}, \quad (2.22)$$

tj. polovina druhé mocniny rozdílu mezi hodnotami.

Dva body $\mathbf{x}_\alpha, \mathbf{x}_\beta$ v geografickém prostoru mohou být spojeny vektorem $\mathbf{h} = \mathbf{x}_\beta - \mathbf{x}_\alpha$. Takto získáme nepodobnost γ^* , která závisí na vzdálenosti a orientaci dvojic bodů popsaných vektorem \mathbf{h} :

$$\gamma^*(\mathbf{h}) = \frac{1}{2}(z(\mathbf{x}_\alpha + \mathbf{h}) - z(\mathbf{x}_\alpha))^2 \quad (2.23)$$

Vzhledem k tomu, že nepodobnost je počítána jako kvadrát rozdílu mezi dvěma hodnotami, pořadí, ve kterém uvažujeme body $\mathbf{x}_\alpha, \mathbf{x}_\beta$, nehraje roli. Řekneme, že nepodobnost je symetrická vzhledem k \mathbf{h} . Variogram cloud bude proto zobrazovat nepodobnosti mezi všemi dvojicemi hodnot proti absolutním hodnotám vektorů \mathbf{h} . Nepodobnost obvykle roste s rostoucí vzdáleností mezi body a pozorování sobě bližší mají tendenci se podobat. Variogram cloud je užitečným nástrojem pro pochopení distribuce výběrových hodnot v geografickém prostoru. Slouží k detekci anomálií či nehomogenity, které se mohou vyskytnout i na krátkých vzdálenostech mezi body. [6]

2.3.1. Experimentální variogram

Principem experimentálního variogramu je spočítat průměrné nepodobnosti $\gamma^*(\mathbf{h})$ pro dané třídy vektorů \mathfrak{h} . Pro všechny dvojice hodnot, které mohou být spojené vektory \mathbf{h} patřícími do třídy vektorů \mathfrak{h} , spočítáme průměrnou nepodobnost příslušející dané třídě. Třída vektorů \mathfrak{h} sestává z množiny vektorů délky spadající do určitého intervalu a jejichž orientace je stejná až do určitého tolerančního úhlu. Obecně volíme nepřekrývající se třídy vektorů \mathfrak{h} . Průměrná nepodobnost příslušející dané třídě vektorů \mathfrak{h}_k je hodnota experimentálního variogramu

$$\gamma^*(\mathfrak{h}_k) = \frac{1}{2n_c} \sum_{\alpha=1}^{n_c} (z(\mathbf{x}_\alpha + \mathbf{h}) - z(\mathbf{x}_\alpha))^2, \mathbf{h} \in \mathfrak{h}_k, \quad (2.24)$$

kde n_c je počet dvojic bodů, které mohou být spojeny vektorem $\mathbf{h} \in \mathfrak{h}_k$. [6]

2.4. Teoretický variogram

Experimentální variogram je prokládán funkcí, která jej nejlépe aproximuje. Tuto funkci nazýváme teoretický variogram, pomocí kterého jsme schopni popsat chování náhodné funkce $Z(\mathbf{x})$ v prostoru. Změny náhodné funkce v prostoru zkoumáme pomocí rozdílů mezi hodnotami v párech bodů \mathbf{x} a $\mathbf{x} + \mathbf{h}$:

$$Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x}), \quad (2.25)$$

které nazýváme přírůstky.

Teoretický semivariogram $\gamma(\mathbf{h})$ je definován pomocí vnořené stacionarity týkající se náhodné funkce, kdy jsme zformulovali dva předpoklady o přírůstcích:

- průměr přírůstků $E(Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})) = m(\mathbf{h})$, tzv. drift, je invariantní vůči jakékoli změně daného vektoru $\mathbf{h} \in \mathcal{D}$. Navíc se předpokládá, že je drift roven nule.
- rozptyl přírůstků $var(Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x}))$ má konečnou hodnotu $2\gamma(\mathbf{h})$, která závisí na délce a orientaci daného vektoru \mathbf{h} , nikoli na pozici $\mathbf{h} \in \mathcal{D}$.

Tyto dvě vlastnosti náhodné funkce dávají definici teoretického semivariogramu $\gamma(\mathbf{h})$. Protože již víme, že průměr přírůstků $Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})$ je roven

nule, potom je semivariogram spočítán jako střední čtvercová diference [6]:

$$\gamma(\mathbf{h}) = \frac{1}{2}E[(Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x}))^2]. \quad (2.26)$$

Teoretický variogram je ve tvaru $2\gamma(\mathbf{h}) = E[(Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x}))^2]$, ačkoliv tyto pojmy obvykle v literatuře nejsou rozlišovány a teoretickým variogramem se nazývá také teoretický semivariogram.

Vlastnosti teoretického variogramu [6]:

1. Hodnota variogramu je pro $\mathbf{h} = 0$ podle definice nula

$$\gamma(0) = 0. \quad (2.27)$$

2. Hodnoty variogramu jsou pozitivní

$$\gamma(\mathbf{h}) \geq 0. \quad (2.28)$$

3. Variogram je sudá funkce

$$\gamma(-\mathbf{h}) = \gamma(\mathbf{h}). \quad (2.29)$$

4. Variogram roste pomaleji než $|\mathbf{h}^2|$

$$\lim_{|\mathbf{h}| \rightarrow \infty} \frac{\gamma(h)}{|\mathbf{h}^2|} = 0. \quad (2.30)$$

5. Variogram lze vyjádřit pomocí kovarianční funkce následovně [1]

$$\gamma(\mathbf{h}) = C(0) - C(\mathbf{h}), \quad (2.31)$$

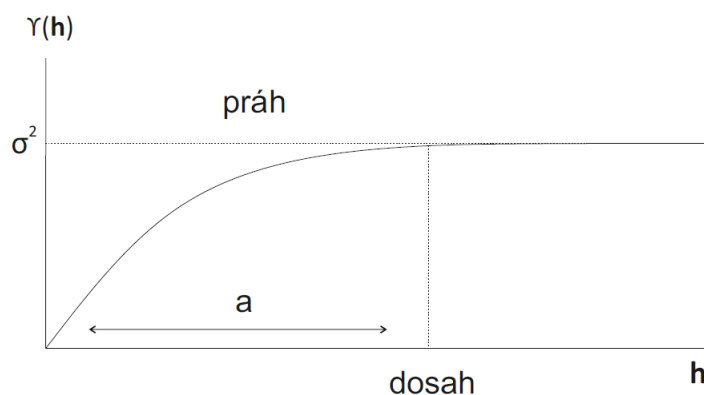
opačně toto obecně neplatí, protože variogram nemusí být nutně omezená funkce.

Důkaz. Důkaz vychází z tvaru pro výpočet variogramu:

$$\begin{aligned} 2\gamma(\mathbf{h}) &= E[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})]^2 \\ 2\gamma(\mathbf{h}) &= E[(Z(\mathbf{x} + \mathbf{h}) - m)^2 + (Z(\mathbf{x}) - m)^2 - 2(Z(\mathbf{x} + \mathbf{h}) - m)(Z(\mathbf{x}) - m)] \\ 2\gamma(\mathbf{h}) &= 2C(0) - 2C(\mathbf{h}) \\ \gamma(\mathbf{h}) &= C(0) - C(\mathbf{h}) \end{aligned}$$

□

Tento vztah ukazuje, že odpovídající kovariance získáme otočením variogramu o 180° . Zatímco variogram začíná od nuly (pro $\mathbf{h} = 0$) a s rostoucím \mathbf{h} roste až po určitý limit (práh), kovarianční funkce začíná od rozptylu a s rostoucím \mathbf{h} klesá. Toto je však možné za předpokladu shora ohraničených variogramů. Lze dokázat, že pouze stacionární regionalizované proměnné mají shora ohraničené variogramy. [1], [8]



Obrázek 2.4: Teoretický variogram

Na obrázku 2.4 vidíme typickou podobu teoretického variogramu. Variogram v určité vzdálenosti dosáhne svého maxima a dále již zůstává konstantní. Tuto maximální hodnotu variogramu nazýváme práh. Vzdálenosti, ve které dosahuje svého prahu říkáme dosah, který je v obrázku označen jako a .

2.4.1. Dosah a oblast vlivu

Míra růstu variogramu v závislosti na růstu vzdálenosti \mathbf{h} značí, jak rychle klesá vliv vzorku s rostoucí vzdáleností. Poté, co variogram dosáhne svého limitu (prahu), kterému odpovídá určitá vzdálenost \mathbf{h} (dosah), jsou již vzorky nekorelované. Tato kritická vzdálenost, od které již vzorky nejsou vzájemně ovlivněny, nám pomůže přesněji formulovat tzv. zónu vlivu neboli dosah. Pro stacionární proměnné je $\gamma(\mathbf{h})$ rovno rozptylu, což platí pro vzdálenosti větší než je hodnota dosahu variogramu:

$$\gamma(\mathbf{h}) = \frac{1}{2} \text{var}[(Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x}))] \quad (2.32)$$

$$\gamma(\mathbf{h}) = \frac{1}{2} [\text{var}(Z(\mathbf{x} + \mathbf{h})) - \text{var}Z((\mathbf{x}))] = \sigma^2.$$

Je nutno zmínit, že ne všechny variogramy dosahují prahu. Existují případy, kdy hodnota variogramu stále roste s rostoucí vzdáleností \mathbf{h} . Také mohou nastat případy, kdy dosah variogramu není stejný ve všech směrech. Toto ověříme pomocí tzv. směrového variogramu, kdy je variogram spočítán zpravidla pro čtyři základní směry, a to pro 0° , 45° , 90° a 135° . Tato odlišnost variogramů pro různé směry úzce souvisí s anizotropií, kterou si vysvětlíme později. [1]

2.4.2. Chování kolem počátku

Do teď jsme se zabývali chováním variogramu pro velké vzdálenosti \mathbf{h} . Chování kolem počátku však souvisí se spojitostí a prostorovou pravidelností proměnné, nemělo by být tedy opominuto. Existují čtyři typy chování kolem počátku:

- **Kvadratické** - indikuje kontinuitu a diferencovatelnost regionalizované proměnné. Kvadratický tvar může souviset s přítomností driftu.
- **Lineární** - regionalizovaná proměnná je kontinuální, ne však diferencovatelná.
- **Diskontinuální kolem počátku** - $\gamma(\mathbf{h})$ se neblíží nule s \mathbf{h} jdoucím k nule. To vypovídá o vysoké nepravidelnosti proměnné na krátkých vzdálenostech.
- **Ploché** - regionalizované proměnné $Z(\mathbf{x} + \mathbf{h})$ a $Z(\mathbf{x})$ jsou nekorelované pro všechny hodnoty \mathbf{h} , nezáleží tedy na tom, jak blízko sobě jsou. Jde tedy o čistou náhodnost či bílý šum.

Diskontinuita kolem počátku se nazývá nugget efekt. Nugget efekt je spojován s variabilitou na krátké vzdálenosti (mikrovariabilitou) nebo také chybou měření. Představit si můžeme příklad, kdy provádíme opakovaná měření na stejných místech, ale výsledky se odlišují. [1]

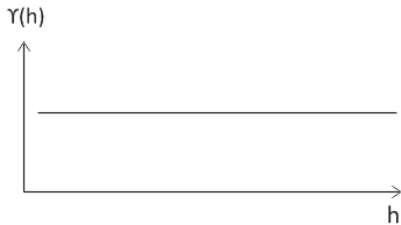
2.4.3. Modely variogramu

Představíme si několik nejčastěji používaných modelů variogramů. Modely, které mají práh korespondují stacionárním regionalizovaným proměnným. Modely bez práhu pak pouze regionalizovaným proměnným, které splňují vnořenou stacionaritu. Variogramy jsou vyjádřeny pomocí kovarianční funkce C . Konstantou c je pak značena hodnota variogramu, která již zůstává od určité vzdálenosti konstantní. [1]

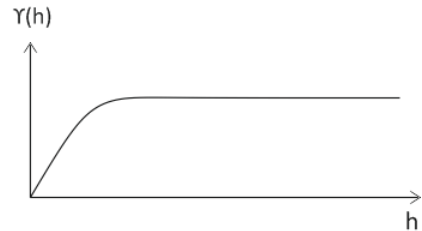
Nugget efekt

$$\gamma(\mathbf{h}) = \begin{cases} 0 & \mathbf{h} = 0 \\ c & |\mathbf{h}| > 0 \end{cases} \quad (2.33)$$

Tento model odpovídá čistě náhodnému jevu, kdy mezi hodnotami není žádná korelace bez ohledu na to, jak blízko sobě jsou. Tento model variogramu můžeme vidět na obrázku 2.5.



Obrázek 2.5: Variogram: nugget model



Obrázek 2.6: Variogram: sférický model

Sférický

$$\gamma(\mathbf{h}) = \begin{cases} C \left(\frac{3}{2} \frac{|\mathbf{h}|}{a} - \frac{1}{2} \left(\frac{|\mathbf{h}|}{a} \right)^3 \right) & |\mathbf{h}| < a \\ c & |\mathbf{h}| \geq a \end{cases} \quad (2.34)$$

Často užívaný model, který je vyjádřený polynomiálně a jeho tvar zachycuje to, co je v praxi často vysledováno - téměř lineární růst až do určité vzdálenosti,

který následuje stabilizace viz. obrázek 2.6.

Exponenciální

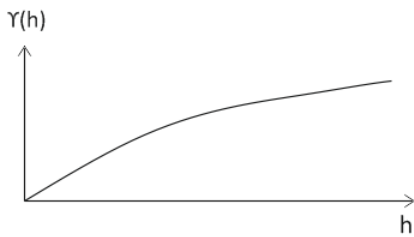
$$\gamma(\mathbf{h}) = C \left(1 - \exp\left(-\frac{|\mathbf{h}|}{a}\right) \right). \quad (2.35)$$

Stejně jako sférický model je i exponenciální model lineární na krátké vzdálenosti. V praxi se za dosah bere hodnota $3a$, protože právě v této vzdálenosti dosahuje variogram 95 % hodnoty prahu. Tento model variogramu je znázorněn na obrázku 2.7.

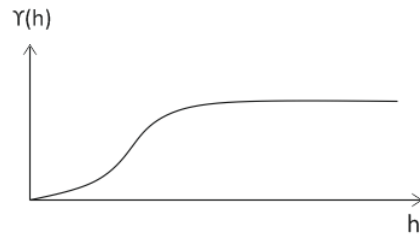
Gaussovský

$$\gamma(\mathbf{h}) = C \left(1 - \exp\left(-\frac{|\mathbf{h}|^2}{a^2}\right) \right). \quad (2.36)$$

Gaussovský model je vysoce kontinuální. V praxi používaný dosah je $1.73a$. Tento model variogramu je zachycen obrázkem 2.8.



Obrázek 2.7: Variogram: exponenciální model

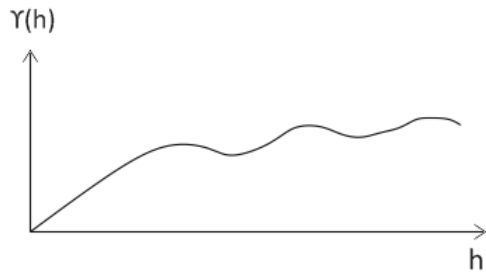


Obrázek 2.8: Variogram: gaussovský model

Základní sinový model

$$\gamma(\mathbf{h}) = C \left(1 - \frac{\sin(r)}{|\mathbf{h}|} \right), \quad (2.37)$$

kde $r = \frac{h}{a}$. Tento model odpovídá velmi kontinuálním strukturám a můžeme jej vidět na obrázku 2.9.



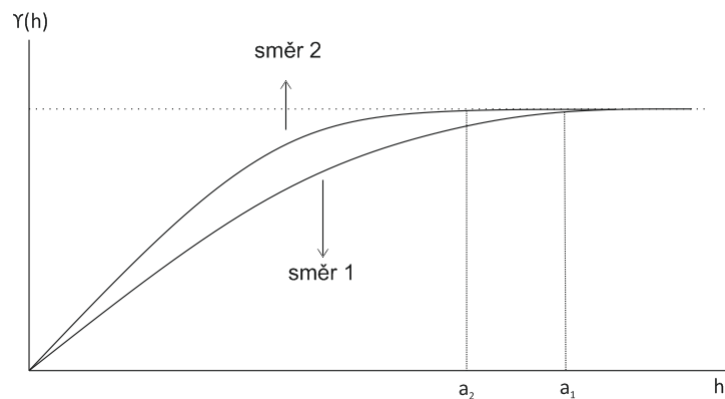
Obrázek 2.9: Variogram: sinový model

2.4.4. Anizotropie

Spočítáme-li variogram pro různé směry, může dojít k tomu, že se chová v jednotlivých směrech odlišně. Tento jev svědčí o tom, že hodnoty variogramu jsou ovlivněny nejen vzdáleností mezi dvěma body, ale také směrem, který vůči sobě mají. V takovém případě se jedná o anizotropní variogram. Pokud jsou variogramy ve všech směrech stejné či podobné, pak variogram závisí jen na vzdálenosti mezi body a řekneme, že je izotropní, tj. invariantní vůči rotaci. [1]

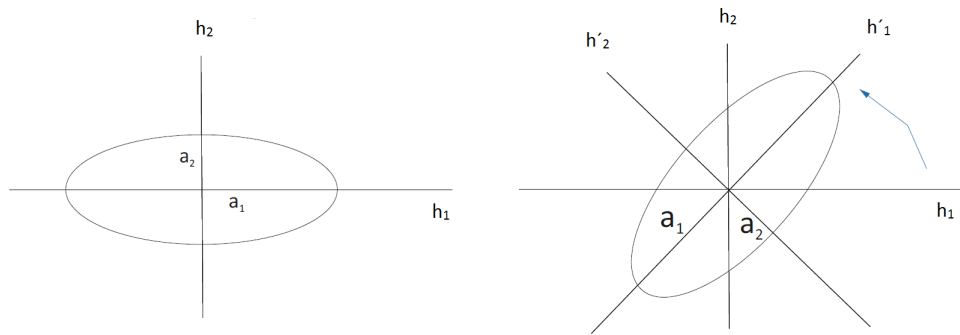
Existují dva druhy anizotropie:

- **Geometrická** - v tomto případě dosahují variogramy pro různé směry stejného nebo podobného prahu, avšak dosah se liší, což je ilustrováno na obrázku 2.10, kde a_1 značí dosah v hlavním směru a a_2 dosah ve vedlejším směru. [1]



Obrázek 2.10: Variogramy: geometrická anizotropie

Můžeme si nakreslit diagram, který znázorňuje dosah jako funkci směru viz. obrázek 2.11. V případě, že jde o elipsu, můžeme jednoduchou změnou souřadnic transformovat elipsu na kruh a eliminovat tak anizotropii. Změna souřadnic probíhá tak, že původní souřadnicový systém $\mathbf{h} = (h_1, h_2)$ rotujeme do nového souřadnicového systému $\mathbf{h}' = (h'_1, h'_2)$. Změna souřadnic je jednoduchá v případě, že hlavní osa elipsy je shodná se souřadnicovou osou. V tomto případě spočítáme variogram γ_1 ve směru 1. [1], [6]



Obrázek 2.11: Elipsy: hlavní a vedlejší osy v případě geometrické anizotropie

Celkový variogram po korekci je pak ve tvaru:

$$\gamma(\mathbf{h}) = \gamma_1(\sqrt{h_1^2 + k^2 h_2^2}), \quad (2.38)$$

kde h_1, h_2 jsou složky \mathbf{h} a $k = \frac{a_1}{a_2}$.

Při výpočtu variogramu je důležité provést výpočet nejméně ve čtyřech směrech. Pokud by byl variogram spočítán pouze ve dvou kolmých směrech, nemuseli bychom zachytit anizotropii vůbec. To by odpovídalo situaci, kdy hlavní osa by byla ve směru 45° vůči tomu směru, ve kterém byl variogram spočten. [1]

- **Zonální** - tento druh anizotropie představuje komplexnější formu anizotropie, která zahrnuje případ, kdy práh variogramu není stejný ve všech směrech. S tímto jevem se setkáváme zejména v situacích, kdy se každé pozorování vztahuje k místu popsanému třemi souřadnicemi - zeměpisnou délkou a šířkou a nadmořskou výškou, tedy $\mathbf{h} = (h_1, h_2, h_3)$. Vertikální vzdálenost hraje obvykle významnou roli, protože právě v tomto směru se vyskytuje větší variabilita v datech. [1]

Standardní postup je rozdělit variogram na dvě komponenty - izotropní variogram $\gamma_0(\sqrt{h_1^2 + h_2^2 + h_3^2})$ a druhý variogram počítaný pouze ve vertikálním směru $\gamma_1(h_3)$. Výsledný variogram je pak ve tvaru [1]:

$$\gamma(\mathbf{h}) = \gamma_0(\mathbf{h}) + \gamma_1(\mathbf{h}). \quad (2.39)$$

3 Kriging

Kriging je metoda odhadování, jejímž výsledkem je nejlepší nestranný lineární odhad (BLUE - best linear unbiased estimator) hodnoty v daném bodě, popř. blokového průměru.

V případě, že provádíme měření na vzorcích odebraných v určitých bodech, získáme přesnou informaci o tom, co se děje v těchto bodech. V praxi většinou potřebujeme zjistit také, co se děje mimo tyto body, popř. zjistit průměry sledované proměnné v blocích.

Kvalita odhadů závisí na několika faktorech. Prvním z nich je počet vzorků a kvalita dat v každém z bodů měření. Pozice vzorků v dané oblasti (definičním oboru) hraje také významnou roli, kdy pravidelně rozmístěné vzorky poskytují lepší obraz o tom, co se v oblasti děje než vzorky seskupené kolem několika málo míst v dané oblasti. V neposlední řadě výsledný odhad ovlivňuje vzdálenost mezi vzorky a bodem odhadu, popř. blokem. Obecně platí, že spoléháme více na vzorky pořízené v okolí odhadu než na ty vzdálené, což je v souladu s myšlenkou, že jevy, které jsou sledovány v blízkých bodech mají tendenci se sobě více podobat. V potaz je nutno vzít také prostorovou kontinuitu, kdy je snazší odhadovat hodnotu pravidelných proměnných než těch nepravidelně rozložených. [1]

3.1. Odvození krigingových rovnic

Základním cílem metody kriging je na základě n datových hodnot $z(\mathbf{x}_1), \dots, z(\mathbf{x}_n)$ z definičního oboru \mathcal{D} odhadnout hodnotu $Z(\mathbf{x}_0)$ v konkrétním bodě \mathbf{x}_0 .

Pro odhad $\hat{Z}(\mathbf{x}_0)$ uvažujeme vážený průměr dat:

$$\hat{z}(\mathbf{x}_0) = \sum_{i=1}^n \lambda_i z(\mathbf{x}_i), \quad (3.1)$$

kde λ_i značí váhy. Problém tedy přechází v to, jak odhadnout váhy λ_i nejlepším způsobem. Uvažujme regionalizovanou proměnnou:

$$\hat{Z}(\mathbf{x}_0) = \sum_{i=1}^n \lambda_i Z(\mathbf{x}_i). \quad (3.2)$$

Váhy jsou vybrány tak, aby výsledný odhad byl BLUE, tj.:

1. Nestranný: $E[\hat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0)] = 0$,
2. Nejlepší ve smyslu minimálního rozptylu: $var[\hat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0)]$.

Tento rozptyl nazýváme rozptylem krigingu. [1], [13]

3.2. Typy krigingu

Rozlišujeme několik typů krigingu v závislosti na tom, zda průměr prostorové proměnné je známý, neznámý, popř. zda se mění v závislosti na tom, v jaké části definičního oboru (bloku) je spočten.

Ve všech případech je potřeba vyřešit soustavu lineárních rovnic nazývanou krigingový systém, jehož řešením dostaneme váhy krigingu a rozptyl krigingu. [1]

3.2.1. Základní kriging

V této kapitole jsem čerpala ze zdrojů [1], [13]. V případě základního krigingu předpokládáme, že regionalizovaná proměnná $Z(\mathbf{x})$ je stacionární a její průměr, m , je neznámý.

Nejprve potřebujeme ověřit:

1. **Nestrannost:** Předpokládáme, že proměnná $Z(\mathbf{x})$ je stacionární s průměrem m :

$$E[Z(\mathbf{x})] = m = E[Z(\mathbf{x}_0)]. \quad (3.3)$$

Odhad je lineární kombinací známých pozorování, což můžeme zapsat takto:

$$\hat{Z}(\mathbf{x}_0) = \sum_{i=1}^n \lambda_i Z(\mathbf{x}_i). \quad (3.4)$$

Aby byl odhad nestranný, musí platit, že $E[\hat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0)] = 0$, což lze vyjádřit jako:

$$E\left[\sum_{i=1}^n \lambda_i Z(\mathbf{x}_i) - Z(\mathbf{x}_0)\right] = \sum_{i=1}^n \lambda_i m - m = m\left[\sum_{i=1}^n \lambda_i - 1\right]. \quad (3.5)$$

To je splněno buď v případě, že $m = 0$ nebo součet vah krigingu $\sum \lambda_i$ je roven 1. V našem případě je ale průměr m neznámý, tudíž součet vah krigingu musí být roven 1.

2. Minimální rozptyl: Rozptyl krigingu $var[\hat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0)]$ můžeme vyjádřit pomocí kovariance takto:

$$\begin{aligned} E[(\hat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0))^2] &= var\left[\sum_{i=1}^n \lambda_i Z(\mathbf{x}_i)\right] + var[Z(\mathbf{x}_0)] - 2C\left[\sum_{i=1}^n \lambda_i Z(\mathbf{x}_i), Z(\mathbf{x}_0)\right] \\ &= \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C[Z(\mathbf{x}_i), Z(\mathbf{x}_j)] + var[Z(\mathbf{x}_0)] - 2 \sum_{i=1}^n \lambda_i C[Z(\mathbf{x}_i), Z(\mathbf{x}_0)]. \end{aligned} \quad (3.6)$$

Za účelem minimalizace rozptylu pod podmínkou, že součet krigingových vah musí být roven 1, zavedeme do výrazu, který má být minimalizován, Lagrangeův multiplikátor μ :

$$L(\lambda_i, \mu) = var[\hat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0)] + 2\mu\left(\sum_{i=1}^n \lambda_i - 1\right). \quad (3.7)$$

S ohledem na omezení $\sum_{i=1}^n \lambda_i = 1$ vidíme, že přidání Lagrangeova multiplikátoru nezmění hodnotu výrazu. Do tohoto výrazu dosadíme rozptyl krigingu:

$$\begin{aligned} L(\lambda_i, \mu) &= \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C[Z(\mathbf{x}_i), Z(\mathbf{x}_j)] + var[Z(\mathbf{x}_0)] - 2 \sum_{i=1}^n \lambda_i C[Z(\mathbf{x}_i), Z(\mathbf{x}_0)] + \\ &\quad + 2\mu\left(\sum_{i=1}^n \lambda_i - 1\right). \end{aligned} \quad (3.8)$$

System krigingových rovnic získáme derivováním tohoto výrazu podle jednotlivých neznámých λ_i a Lagrangeova multiplikátoru μ a položením těchto parciálních derivací rovno nule:

$$\frac{\partial L(\lambda_i, \mu)}{\partial \lambda_i} = 2 \sum_{j=1}^n \lambda_j C[Z(\mathbf{x}_i), Z(\mathbf{x}_j)] - 2C[Z(\mathbf{x}_i), Z(\mathbf{x}_0)] + 2\mu = 0$$

$$\frac{\partial L(\lambda_i, \mu)}{\partial \mu} = 2 \sum_{i=1}^n \lambda_i - 2 = 0$$

Výsledný systém krigingových rovnic lze přepsat do tvaru:

$$\sum_{j=1}^n \lambda_j C[Z(\mathbf{x}_i), Z(\mathbf{x}_j)] + \mu = C[Z(\mathbf{x}_i), Z(\mathbf{x}_0)] \quad (3.9)$$

$$\sum_{i=1}^n \lambda_i = 1 \quad (3.10)$$

System krigingových rovnic lze přepsat do maticové podoby $\mathbf{AX} = \mathbf{B}$:

$$\underbrace{\begin{pmatrix} C_{11} & C_{12} & \cdots & C_{1n} & 1 \\ C_{21} & C_{22} & \cdots & C_{2n} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \vdots \\ \lambda_n \\ \mu \end{pmatrix}}_{\mathbf{X}} = \underbrace{\begin{pmatrix} C(x_1, x_0) \\ C(x_2, x_0) \\ \vdots \\ \vdots \\ C(x_n, x_0) \\ 1 \end{pmatrix}}_{\mathbf{B}} \quad (3.11)$$

Rozptyl krigingu (minimální rozptyl) je dán pomocí střední čtvercové chyby MSE:

$$MSE = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C[Z(\mathbf{x}_i), Z(\mathbf{x}_j)] + var[Z(\mathbf{x}_0)] - 2C\left[\sum_{i=1}^n \lambda_i Z(\mathbf{x}_i), Z(\mathbf{x}_0)\right],$$

kde

$$\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C[Z(\mathbf{x}_i), Z(\mathbf{x}_j)] = \sum_{i=1}^n \lambda_i \sum_{j=1}^n \lambda_j C[Z(\mathbf{x}_i), Z(\mathbf{x}_j)]$$

$$= \sum_{i=1}^n \lambda_i (C[Z(\mathbf{x}_i), Z(\mathbf{x}_j)] - \mu).$$

Rozptyl krigingu pak můžeme zapsat jako:

$$\sigma_{OK}^2 = \sigma^2 - \sum_{i=1}^n \lambda_i (C[Z(\mathbf{x}_i), Z(\mathbf{x}_j)] - \mu). \quad (3.12)$$

nebo pomocí maticového zápisu:

$$\sigma_{OK}^2 = \sigma^2 - \mathbf{X}^T \mathbf{B}. \quad (3.13)$$

3.2.2. Jednoduchý kriging

V této kapitole si ukážeme odvození systému krigingových rovnic pro případ, že průměr m regionalizované proměnné $Z(\mathbf{x})$ je známý. Pro tuto kapitolu jsem čerpala ze zdroje [1].

Nejprve uvažujme regionalizovanou proměnnou $Y(\mathbf{x})$ s nulovým průměrem. Pak původní prostorovou proměnnou dostaneme jako $Z(\mathbf{x}) = Y(\mathbf{x}) + m$.

Odhad $Y(\mathbf{x})$ je ve tvaru:

$$\hat{Y}(\mathbf{x}_0) = \sum_{i=1}^n \lambda_i Y(\mathbf{x}_i). \quad (3.14)$$

Opět požadujeme, aby tento odhad byl nestranný a měl minimální rozptyl.

Pro nestrannost odhadu musí platit:

$$E[\hat{Y}(\mathbf{x}_0) - Y(\mathbf{x}_0)] = E\left[\sum_{i=1}^n \lambda_i Y(\mathbf{x}_i) - Y(\mathbf{x}_0)\right] = 0. \quad (3.15)$$

Již víme, že průměr $Y(\mathbf{x}) = 0$, tudíž je odhad automaticky nestranný, a proto v případě jednoduchého krigingu není žádný požadavek na součet vah.

Rozptyl krigingu je:

$$\text{var}[\hat{Y}(\mathbf{x}_0) - Y(\mathbf{x}_0)] = E\left[\sum_{i=1}^n \lambda_i Y(\mathbf{x}_i) - Y(\mathbf{x}_0)\right]^2 \quad (3.16)$$

$$= \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C[Y(\mathbf{x}_i), Y(\mathbf{x}_j)] + \text{var}[Y(\mathbf{x}_0)] - 2 \sum_{i=1}^n \lambda_i C[Y(\mathbf{x}_i), Y(\mathbf{x}_0)].$$

Vzhledem k tomu, že zde není podmínka pro váhy, není potřeba zavádět Lagrangeův multiplikátor.

Systém krigingových rovnic pro jednoduchý kriging je ve tvaru:

$$\sum_{j=1}^n \lambda_j C[Y(\mathbf{x}_i), Y(\mathbf{x}_j)] = C[Y(\mathbf{x}_i), Y(\mathbf{x}_0)], i = 1, \dots, n. \quad (3.17)$$

Rozptyl jednoduchého krigingu je dán vztahem:

$$\sigma_{SK}^2 = \sigma^2 - \sum_{i=1}^n \lambda_i C[Y(\mathbf{x}_i), Y(\mathbf{x}_0)]. \quad (3.18)$$

Povšimněme si, že vztahy pro systém krigingu a rozptyl krigingu jsou shodné jako v případě základního krigingu z hlediska kovariance po vynechání členu s Lagrangeovým multiplikátorem.

3.2.3. Univerzální kriging

Informace o univerzálním krigingu jsem čerpala ze zdrojů [8], [12] a [13]. V případě nesplnění podmínek stacionarity regionalizované proměnné $Z(\mathbf{x})$ použijeme tzv. univerzální kriging. Regionalizovanou proměnnou rozdělíme na dvě složky - trendovou (drift), která určuje průměrnou hodnotu a reziduální složku.

K modelování trendu se v praxi nejčastěji používají polynomy 1. či 2. řádu:

$$m(\mathbf{x}) = \mu + b_1 x_i + b_2 y_i$$

pro polynom 1. řádu,

$$m(\mathbf{x}) = \mu + b_1 x_i + b_2 y_i + b_3 x_i^2 + b_4 x_i y_i + b_5 y_i^2$$

pro polynom 2. řádu. x_i, y_i značí souřadnice i -tého bodu \mathbf{x}_i , μ, b_1, \dots, b_5 jsou koeficienty trendu.

Soustavu rovnic pro výpočet vah krigingu rozšíříme o členy popisující trend. Krigování pak není provedeno na původní data, ale na data očištěná od zjištěného

trendu $(z(x_1, y_1) - m(x_1, y_1), \dots, z(x_n, y_n) - m(x_n, y_n))$. K výsledným hodnotám krigingu je na závěr přičtena hodnota trendu v konkrétním bodě.

V případě trendu popsaného polynomem 1. stupně vypadá soustava rovnic následovně:

$$\begin{pmatrix} C_{11} & C_{12} & \cdots & C_{1n} & 1 & x_1 & y_1 \\ C_{21} & C_{22} & \cdots & C_{2n} & 1 & x_2 & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} & 1 & x_n & y_n \\ 1 & 1 & \cdots & 1 & 0 & 0 & 0 \\ x_1 & x_2 & \cdots & x_n & 0 & 0 & 0 \\ y_1 & y_2 & \cdots & y_n & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \vdots \\ \lambda_n \\ \mu \\ b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} C(\mathbf{x}_1, \mathbf{x}_0) \\ C(\mathbf{x}_2, \mathbf{x}_0) \\ \vdots \\ \vdots \\ C(\mathbf{x}_n, \mathbf{x}_0) \\ 1 \\ x_0 \\ y_0 \end{pmatrix}, \quad (3.19)$$

kde x_i, y_i jsou souřadnice vstupních bodů \mathbf{x}_i , $i = 1, \dots, n$; x_0, y_0 souřadnice bodu odhadu \mathbf{x}_0 ; b_1, b_2 koeficienty polynomu.

3.2.4. Lognormální kriging

V praxi je často vyskytujícím se jevem pozitivní šikmost dat. V takovém případě je vhodné původní data nejprve zlogaritmovat. Já jsem pro převod zvolila přirozený logaritmus:

$$Y(\mathbf{x}) = \ln(Z(\mathbf{x})), \quad (3.20)$$

kde předpokládáme, že $Y(\mathbf{x})$ je stacionární.

Pokud jsou transformovaná data $(y(\mathbf{x}_1), \dots, y(\mathbf{x}_n))$ přibližně normálně rozdělená, řekneme, že původní data $(z(\mathbf{x}_1), \dots, z(\mathbf{x}_n))$ jsou lognormálně rozdělená. Výpočet a modelování semivariogramu, stejně jako krigování, provádíme na těchto transformovaných datech. [8] Takto získaný výsledek pak musíme zpětně transformovat, protože odhadnuté hodnoty jsou v logaritmické míře. Výsledek ještě upravíme o systematickou chybu. Výsledný vztah pro zpětnou transformaci je ve tvaru:

$$\hat{Z}(\mathbf{x}_0) = \exp(\hat{Y}(\mathbf{x}_0) + \frac{\hat{\sigma}_Y}{2}), \quad (3.21)$$

kde $\hat{Y}(\mathbf{x}_0)$ je odhadnutá hodnota v bodě \mathbf{x}_0 , $\hat{\sigma}_Y$ je rozptyl odhadu $\hat{Y}(\mathbf{x}_0)$ v bodě \mathbf{x}_0 . [12]

4 Praktická část

V praktické části této práce budu aplikovat výše popsané metody na reálná data. V mém případě se jedná o hydrometeorologická data - srážky, kdy mám k dispozici denní úhrny srážek za období listopad roku 2015 z 319 měřících stanic. Protože by bylo velmi rozsáhlé provádět výpočty pro každý jednotlivý den za zvolené období, data jsem agregovala a jednotlivé metody použiji pro měsíční úhrn srážek.

Pro výpočty jsem použila statistický software R. Jedná se o volně dostupný software, jehož prostředí umožňuje statistické výpočty a poskytuje rozmanité nástroje pro vizualizaci dat. Mezi použité knihovny patří zejména knihovna `gstat`, která obsahuje stěžejní funkce této práce jako je `variogram()`, `fit.variogram()`, `krige()`. Další knihovnou, kterou jsem v práci využívala pro vytvoření prostorových objektů je knihovna `sp`. Okrajově byla také využita knihovna `geoR`, a to pro převedení dat na tzv. geodata pomocí funkce `as.geodata()` a jejich následné vykreslení, jako to můžeme vidět na obrázku 4.5. Pro vytvoření některých grafických výstupů byly použity knihovny `ggplot2` a `scatterplot3d`.

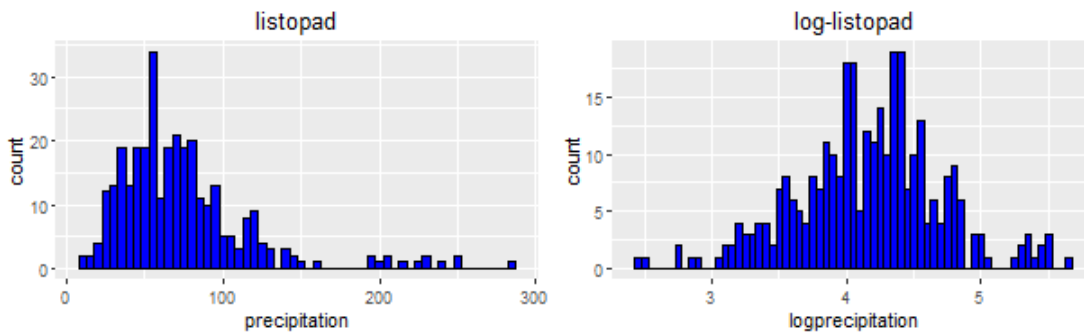
4.1. Průzkumová analýza dat

Než přejdeme k samotným interpolačním metodám, je třeba si data připravit pro vlastní práci. Při práci s geologickými daty je vhodné převést souřadnice ze zeměpisné soustavy souřadnic do souřadnicového systému, kde jsou již data zobrazena v rovině. V této práci budu pracovat s daty, jejichž poloha je zadána v souřadnicovém systému S-42.

Jedním z aspektů, který může snižovat kvalitu výsledků či je zkreslovat, je

distribuce dat. Toto se týká zejména modelování variogramu při použití metody kriging, kdy šikmé rozdělení dat může činit obtížným odhad parametrů variogramu. Logaritmická transformace lognormálně rozdělených dat může zlepšit model variogramu.

Pro získání představy o rozdělení dat si vykreslíme jejich histogram.



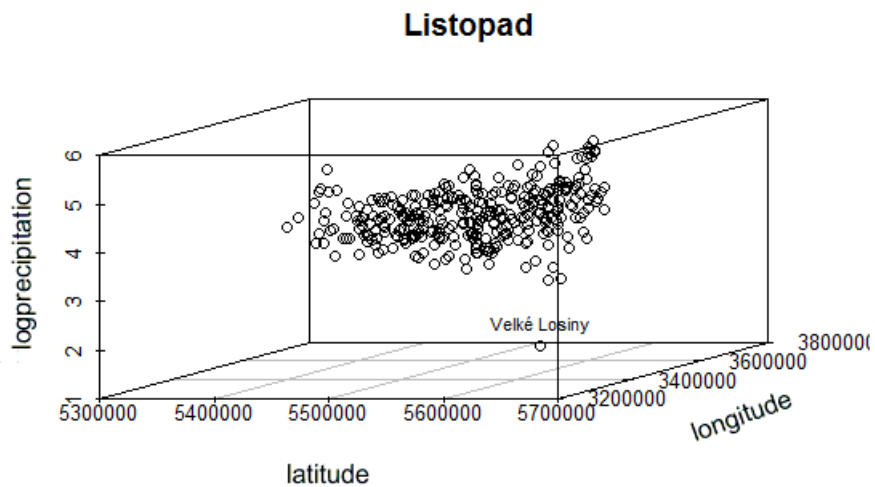
Obrázek 4.1: Histogram: původní data (vlevo), zlogaritmovaná data (vpravo)

V obrázku 4.1 vidíme, že data vykazují pravostranné šikmé rozdělení. Po zlogaritmování se data blíží normálnímu rozdělení.

V případě metody kriging budu tedy pracovat se zlogaritmovanými daty. Výsledek krigingu následně transformuji zpět, abychom mohli přímo interpretovat výsledné srážkové mapy. Navíc tak budou porovnatelné s výsledkem metody IDW, kterou provedu na původní data.

Pro získání prvotní představy o datech jsou vhodné zejména nástroje vizualizace. Podívejme se tedy ještě na třírozměrný bodový graf, který zachycuje hodnoty úhrnů srážek příslušející jednotlivým stanicím.

Na obrázku 4.2 vidíme, že v měsíci listopad se vyskytuje jedna hodnota úhrnu srážek příslušející měřicí stanici ve Velkých Losinách, která se výrazně odlišuje od ostatních. Takovéto hodnoty nazýváme odlehlými. Pro další práci ji odstráním z datového souboru.



Obrázek 4.2: 3D bodový graf - zlogaritmovaná data

4.2. Metoda IDW

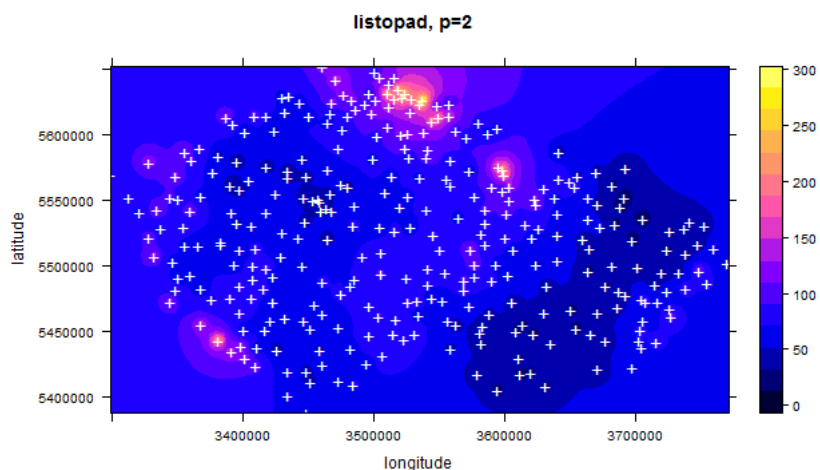
Nejprve se podíváme na výsledky interpolace úhrnu srážek metodou IDW. Vyzkoušela jsem dvě volby silového parametru p . Ve srážkových mapách jsou navíc znázorněny měřící stanice bílými křížky.

- $p = 2$:

Tato volba patří mezi nejčastěji používané. Výsledek můžeme vidět na obrázku 4.3. Můžeme si povšimnout, že vysokou srážkovostí se vyznačuje sever republiky, kde bylo v listopadu například naměřeno 102.1 mm srážek v Liberci či 207.1 mm o 5 km východněji v Josefově Dole. Na základě srážkové mapy lze konstatovat, že jihovýchod republiky se vyznačuje nižšími hodnotami úhrnů srážek, které se pohybují kolem 30 mm. Uveďme například situaci v Kroměříži, kde v listopadu spadlo 30.2 mm srážek nebo v Holešově, kde bylo naměřeno 39.4 mm.

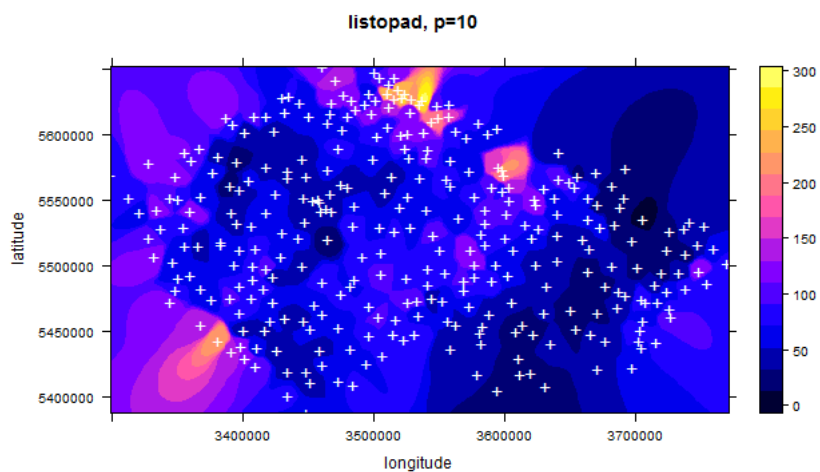
- $p = 10$:

Tato volba slouží spíše pro ilustraci toho, jak s rostoucí hodnotou parametru p roste vliv nejbližšího pozorování na odhadovanou hodnotu. Na obrázku 4.4



Obrázek 4.3: IDW - srážková mapa, $p = 2$

vidíme, že při interpolaci s touto hodnotou silového parametru se více odráží vliv lokálních jevů, v důsledku čehož je výsledná mapa srážek více členitá oproti té s volbou silového parametru $p = 2$.



Obrázek 4.4: IDW - srážková mapa, $p = 10$

I zde vidíme stejný trend jako v předchozí mapě, tedy jihovýchod republiky obecně patří mezi oblast s nižšími hodnotami srážek. Severní hranice republiky opět patří mezi nejdeštivější oblasti ČR.

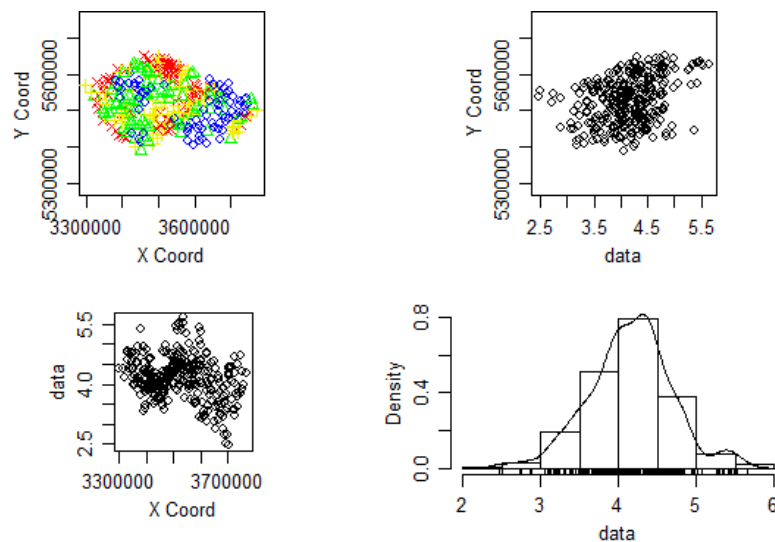
Nutno však zmínit, že v hraničních oblastech je kvalita odhadu zkreslená

v důsledku toho, že jsou měření, na nichž je interpolace založena, dostupná pouze na území České republiky. Hraniční body ČR tedy postrádají úplné okolí.

4.3. Metoda Kriging

Nyní přejdeme k odhadování množství srážek metodou kriging. Tato metoda je provedena pro logaritmovaná data, výsledné srážkové mapy budou následně převedené zpět do původního měřítka.

V předchozí kapitole s názvem Průzkumová analýza dat jsme již detekovali odlehle hodnoty. Jednalo se pouze o jedno měření pořizené ve Velkých Losinách. Přesto může být užitečné, podívat se na data ještě souhrnně.



Obrázek 4.5: Geodata souhrnně

Na obrázku 4.5 můžeme vidět 4 grafy shrnující naše geologická data. První graf vlevo nahoře znázorňuje množství srážek současně s jejich polohou. Hodnoty srážek jsou ještě rozlišeny barevně a symbolicky do čtyř tříd. Grafy na vedlejší diagonále zachycují hodnoty srážek vzhledem k jednotlivým souřadnicím. Na posledním grafu vpravo dole můžeme vidět histogram dat proložený odhadem jejich hustoty.

4.3.1. Stacionarita

Prvním z předpokladů, který je potřeba ověřit než přejdeme k modelování variogramu a následnému krigování, je předpoklad stacionarity. V případě, že se bude průměrná hodnota srážek lišit v závislosti na konkrétní oblasti daného regionu, proložíme souřadnice lineárním trendem či polynomem vyššího řádu. Správnost použitého trendu v praxi zjistíme ověřením podmínky stacionarity prvního řádu, aplikovanou na rezidua $R(\mathbf{x}) = Z(\mathbf{x}) - m(\mathbf{x})$, kde $Z(\mathbf{x})$ značí naměřenou hodnotu v bodě \mathbf{x} , $m(\mathbf{x})$ hodnotu trendu v tomto bodě. [7]

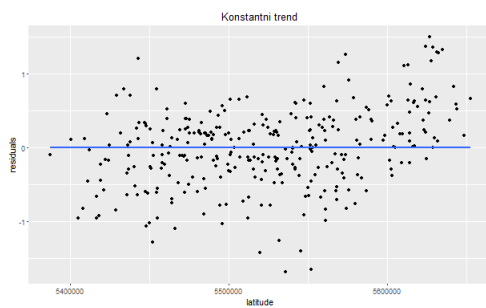
• Konstantní trend

V případě konstantního trendu je průměrná hodnota srážek stálá v celé oblasti. Tato situace patří mezi nejjednodušší a vede k základnímu krigingu v případě, že je průměr neznámý; k jednoduchému, pokud je průměr známý.

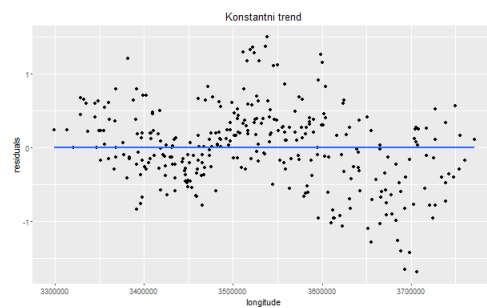
Předpokládaný model je ve tvaru:

$$E[Z(\mathbf{x}_i)] = m(x_i, y_i) = \mu,$$

kde μ je absolutní člen konstantního trendu a x_i, y_i jsou souřadnice. [7]



Obrázek 4.6: Konstantní trend - rezidua, zeměpisná šířka



Obrázek 4.7: Konstantní trend - rezidua, zeměpisná délka

Na obrázcích 4.6 a 4.7 vidíme vykreslená rezidua proti jednotlivým souřadnicím. Za předpokladu stacionarity a použití správného trendu by měla rezidua náhodně oscilovat kolem nuly. Vidíme, že toto zřejmě splněno není, obzvlášť pak u reziduí vykreslených proti zeměpisné délce.

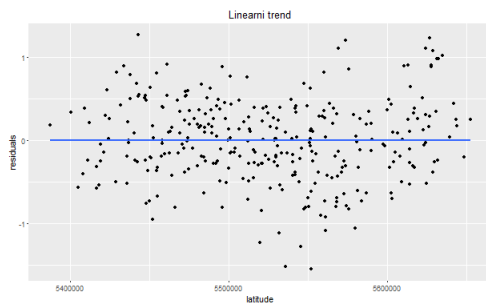
Budeme tedy prokládat souřadnice trendem, proto již víme, že vhodnou krigingovou metodou bude univerzální kriging.

• Lineární trend

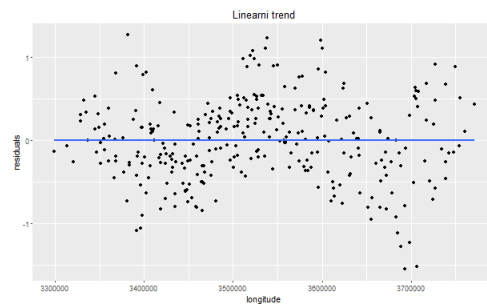
Konstantní trend nevyhovoval našim datům, proto vyzkoušíme souřadnicemi proložit polynom nejprve prvního řádu a uvažovat tak lineární trend, který je ve tvaru:

$$E[Z(\mathbf{x}_i)] = m(x_i, y_i) = \mu + b_1x_i + b_2y_i,$$

kde μ, b_1, b_2 jsou koeficienty trendu. [7]



Obrázek 4.8: Lineární trend - rezidua, zeměpisná šířka



Obrázek 4.9: Lineární trend - rezidua, zeměpisná délka

Na obrázcích 4.8 a 4.9 opět vidíme vykreslená rezidua. Použitím lineárního trendu jsme také nedocílili stacionárního chování reziduí.

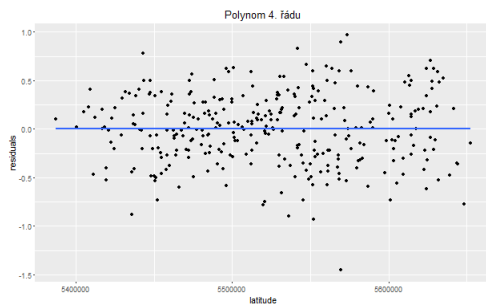
Následný postup je obdobný s postupným obměňováním polynomu na vyšší řády, a to na kvadratický, kubický trend a polynom 4. řádu. Protože zejména vývoj reziduí vzhledem k zeměpisné délce má střídavě klesající a rostoucí charakter, zkusila jsem souřadnice proložit goniometrickou funkcí kombinací sinu a cosinu. Přestože v případech polynomu 2., 3. řádu a goniometrické funkce došlo k mírnému zlepšení oproti konstantnímu trendu, rezidua stále vykazovala trend.

• Polynom 4. řádu

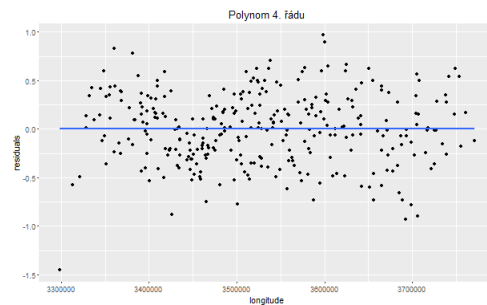
Při proložení souřadnic polynomem 4. řádu, který je ve tvaru:

$$E[Z(\mathbf{x}_i)] = m(x_i, y_i) = \mu + b_1x_i + b_2y_i + b_3x_i^2 + b_4y_i^2 + b_5x_iy_i + b_6x_i^3 + b_7y_i^3 + b_8x_i^2y_i + b_9x_iy_i^2 + b_{10}x_i^4 + b_{11}y_i^4 + b_{12}x_i^2y_i^2 + b_{13}x_i^3y_i + b_{14}x_iy_i^3,$$

kde μ, b_1, \dots, b_{14} jsou koeficienty trendu, jsou rezidua stacionární, jak je vidět na obrázku 4.10 a 4.11. Proto budu tento trend považovat za vhodný a použiji jej k následné práci.



Obrázek 4.10: Polynom 4. řádu - rezidua, zeměpisná šířka



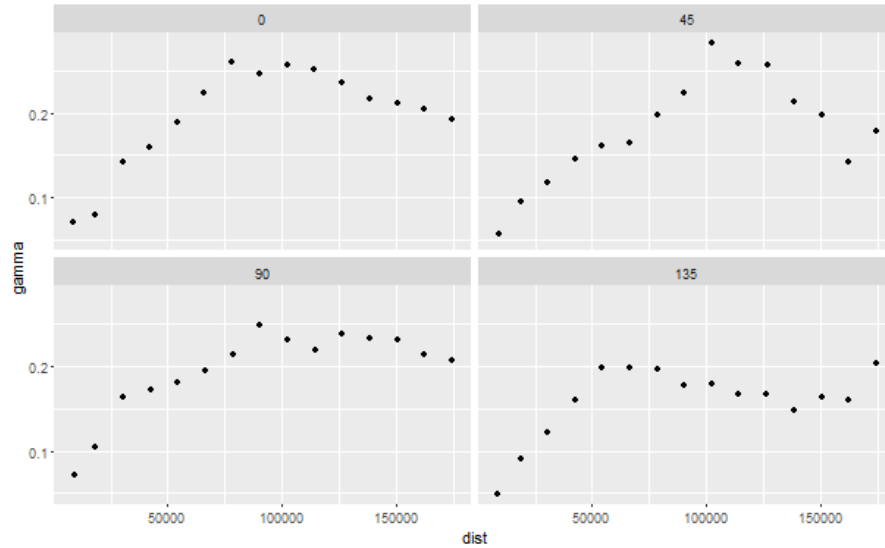
Obrázek 4.11: Polynom 4. řádu - rezidua, zeměpisná délka

4.3.2. Modelování variogramu

Teorie regionalizované proměnné je založena na předpokladu, že variabilita proměnné je stejná ve všech směrech. V tomto případě je prostorová variabilita izotropní a je možné využít pro výpočet variogramu všechny možné dvojice datových bodů. [7]

Případnou anizotropii vyšetřujeme pomocí vykreslení experimentálního variogramu alespoň pro 4 základní směry: 0° , 45° , 90° a 135° . Na obrázku 4.12 můžeme vidět směrový variogram pro naše data s uvažovaným trendem ve tvaru polynomu 4. řádu.

Již zde si můžeme všimnout určitých odlišností ve variogramech. Abychom zjistili konkrétní parametry variogramu v každém ze směrů a zjistili tak s jistotou,



Obrázek 4.12: Směrový variogram

zda jde o anizotropní model, budeme experimentální variogramy v jednotlivých směrech prokládat křivkou - teoretickým variogramem, který nejlépe fituje hodnoty experimentálního variogramu. Nejčastější metodou fitování je minimalizace váženého součtu čtverců chyb $\sum w_j(\gamma(\mathbf{h}) - \hat{\gamma}(\mathbf{h}))^2$. [2]

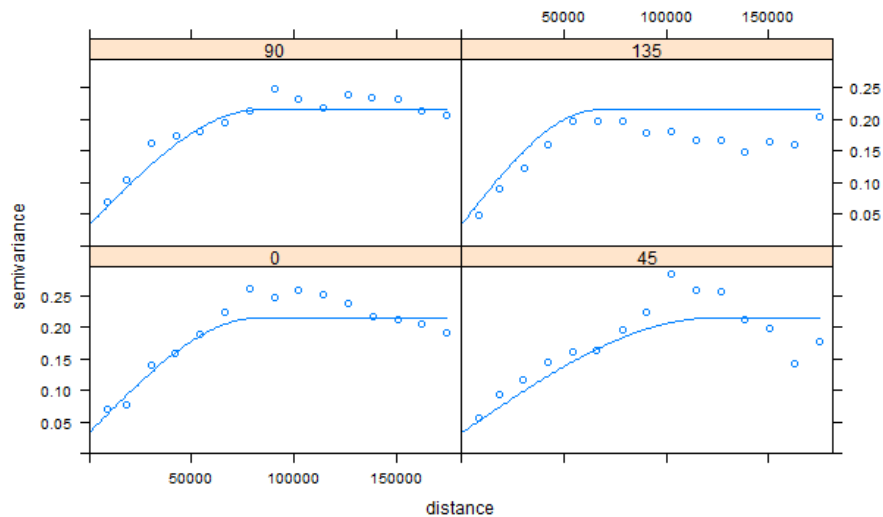
Úhel	Zbytkový rozptyl	Částečný práh	Celkový práh	Dosah
0°	0.03	0.21	0.24	94 085
45°	0.04	0.19	0.23	113 642
90°	0.047	0.17	0.217	81 642
135°	0.013	0.17	0.18	61 793.7

Tabulka 4.1: Parametry směrových variogramů

V tabulce 4.1 můžeme vidět odhadnuté parametry směrových variogramů. Lze konstatovat, že variogramy v jednotlivých směrech dosahují podobných prahů, dosah se ovšem liší. V tomto případě se jedná o geometrickou anizotropii, kdy influenční okolí není kruh, jako by tomu bylo v případě izotropního modelu, ale elipsa. [4] Z tabulky 4.1 je patrné, že hlavním směrem je směr 45° s dosahem 113 642, což představuje délku hlavní poloosy elipsy. Vedlejší směr je tedy 135° s dosahem 61 793.7, který tvoří vedlejší poloosu elipsy.

Jedním z v praxi využívaných způsobů modelování anizotropního variogramu

probíhá tak, že vytvoříme variogram s parametry variogramu v hlavním směru. Navíc je ještě nutné definovat parametry anizotropie, kterými jsou hlavní směr - v našem případě tedy 45° - a poměr anizotropie, který spočítáme jako poměr mezi vedlejším a hlavním dosahem. Proto bude poměr vždy menší než 1. Pro naše data je roven 0.544.



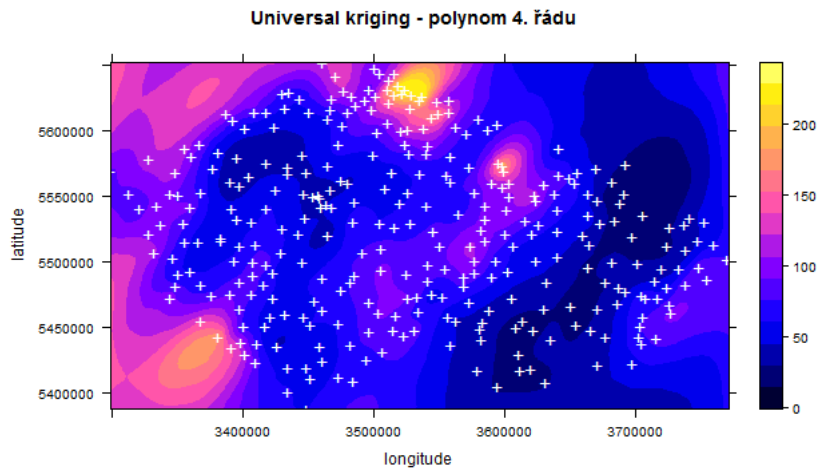
Obrázek 4.13: Variogram - fitovaný anizotropní model

Výsledný fitovaný model je zobrazený na obrázku 4.13 modrou čarou. Navíc jsou zde zobrazeny hodnoty experimentálních variogramů pro 4 základní směry.

4.3.3. Kriging

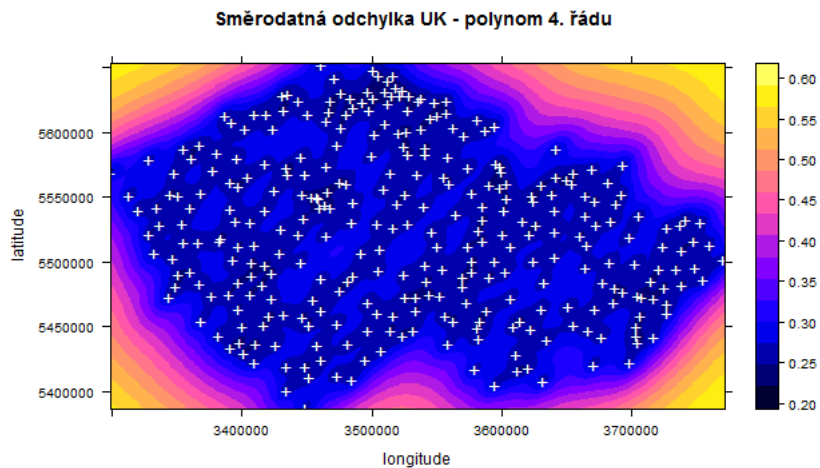
Nyní již můžeme přejít k odhadům úhrnů srážek pomocí metody kriging s použitím výše namodelovaného anizotropního variogramu.

Výsledná srážková mapa je zobrazena na obrázku 4.14. Bílými křížky jsou znázorněny použité měřící stanice v ČR. Odhadnuté hodnoty úhrnů srážek jsou zobrazeny v původním měřítku, tedy po zpětné transformaci z měřítka logaritmickeho. V měsíci listopadu patřil mezi nejméně deštivé oblasti ČR východ republiky společně s částí severozápadu ČR, kde množství srážek spadných za měsíc nepřesáhlo 50 mm. Sever a také část středovýchodu republiky lze naopak označit za nejdeštivější oblast, kde srážky přesahovaly 100 mm. V obrázku 4.15 je ještě



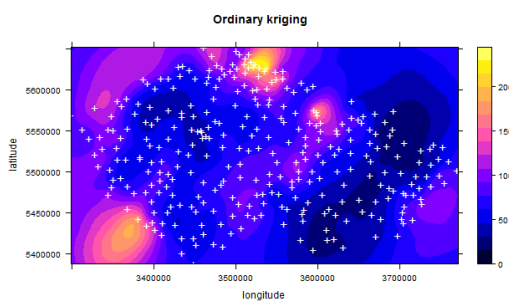
Obrázek 4.14: Srážková mapa - univerzální kriging (polynom 4. řádu)

znázorněna mapa směrodatných odchylek krigingu. Tyto jsou zobrazeny pro logaritmovaná data, protože v případě směrodatných odchylek se narozdíl od úhrnů srážek nejedná o hodnoty, které bychom potřebovali znát v jejich skutečném měřítku a převod rozptylu (a tedy i směrodatné odchylky) logaritmicke transformované veličiny na původní měřítko, narozdíl od zpětné transformace odhadnutých úhrnů srážek, vedl k velké deformaci výsledné mapy tak, že spíše než vývoj směrodatných odchylek v jednotlivých oblastech zachycovala vývoj srážek.

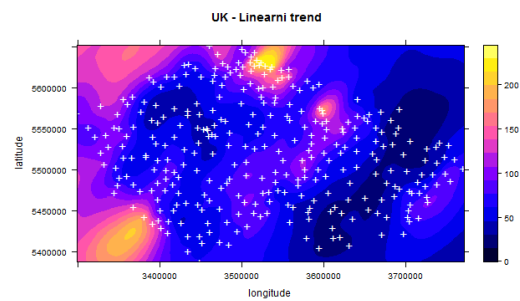


Obrázek 4.15: Směrodatné odchylky krigingu - univerzální kriging (polynom 4. řádu)

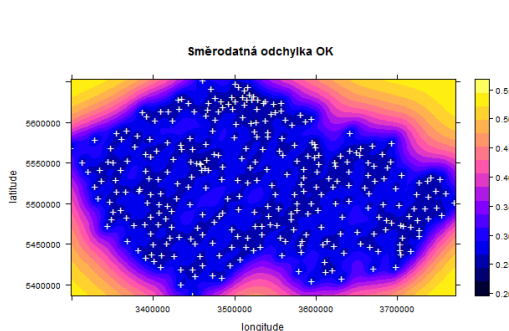
Při provádění univerzálního krigingu z důvodu snadnější interpretace preferujeme trend v datech modelovaný polynomem co nejnižšího řádu. Přestože je model s trendem polynomu 4. řádu z formálního hlediska korektní, může být obtížné z něj dělat závěry. Zkusila jsem proto vytvořit mapy srážkových úhrnů pomocí základního krigingu a univerzálního krigingu s lineárním trendem. Následně otestuji, zda porušení předpokladu stacionarity výrazně ovlivnilo schopnost modelu odhadovat hodnoty. Stejně jako u univerzálního krigingu s polynomem 4. řádu se jedná o anizotropní modely s hlavním směrem 45° a vedlejším 135° . V případě základního krigingu je poměr anizotropie 0.669, u univerzálního krigingu s lineárním trendem to je 0.568.



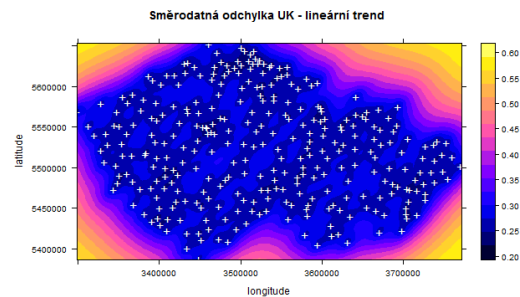
Obrázek 4.16: Srážková mapa - základní kriging



Obrázek 4.17: Srážková mapa - univerzální kriging (lineární trend)



Obrázek 4.18: Směrodatné odchylky krigingu - základní kriging



Obrázek 4.19: Směrodatné odchylky krigingu - univerzální kriging (lineární trend)

Na obrázcích 4.16 a 4.17 jsou zobrazeny výsledné srážkové mapy. V obrázcích 4.18, 4.19 pak mapy jejich směrodatných odchylek krigingu. Můžeme konstatovat

vat, že základní rysy srážkových map, stejně jako map směrodatných odchylek, zůstávají ve všech případech stejné.

Pro posouzení vhodnosti použitého modelu si uvedeme dvě často užívaná měřítka, kterými jsou střední čtvercová chyba ($RMSE$) a koeficient determinace (R^2).

- $RMSE$ - neboli střední čtvercová chyba, která charakterizuje, jak přesně zvolený model fituje naše data. Můžeme ji interpretovat jako průměrnou odchylku odhadnutých a napozorovaných hodnot v lokalitách \mathbf{x}_i . Preferujeme nízké hodnoty.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n [Z(\mathbf{x}_i) - \hat{Z}(\mathbf{x}_i)]^2}{n}},$$

kde $Z(\mathbf{x}_i)$ je pozorovaná hodnota a $\hat{Z}(\mathbf{x}_i)$ odhadnutá hodnota v bodě \mathbf{x}_i . [7]

- R^2 - koeficient determinace, který udává poměr variability vysvětlené modelem, proto nabývá hodnot od 0 do 1. Žádoucí jsou hodnoty blízké jedné.

$$R^2 = \frac{\sum_{i=1}^n [\hat{Z}(\mathbf{x}_i) - \bar{Z}(\mathbf{x})]^2}{\sum_{i=1}^n [Z(\mathbf{x}_i) - \bar{Z}(\mathbf{x})]^2},$$

kde $\bar{Z}(\mathbf{x})$ je průměrná hodnota $Z(\mathbf{x}_i)$. [7]

Tyto charakteristiky jsem napočítala pro všechny tři zkoumané modely, a to pro základní kriging - ZK, univerzální kriging s lineárním trendem - UK (lineární trend) a univerzální kriging s trendem ve tvaru polynomu 4. řádu - UK (polynom 4. řádu). Výsledné hodnoty těchto charakteristik zachycuje tabulka 4.2. Hodnoty jsou uvedeny pro odlogaritovaná data. Pořadí vhodnosti modelu, stejně jako koeficient determinace, by zůstali stejné také v případě výpočtu s logaritmovanými daty.

Podle koeficientu determinace ($R^2 = 1$) vidíme, že všechny uvažované modely vysvětlily téměř veškerou variabilitu v datech. Téměř veškerou, protože jeho hodnota nebyla přímo rovna jedné. Abychom nedostali hodnotu rovnu jedné vlivem zaokrouhlování, museli bychom koeficient determinace uvádět s až 26-ti místným

Model	RMSE	R^2
ZK	8.847 e-15	1
UK (lineární trend)	3.076 e-14	1
UK (polynom 4. řádu)	9.073 e-10	1

Tabulka 4.2: Charakteristiky kvality modelu - odlogaritmovaná data

desetinným rozvojem. Podle charakteristiky $RMSE$ lze modely seřadit od nejlepšího následovně - základní kriging, univerzální kriging s lineárním trendem a nakonec univerzální kriging s polynomem 4. řádu, avšak hodnoty této charakteristiky jsou u všech modelů velice malé, proto také rozdíly mezi nimi jsou zanedbatelné.

4.3.4. Křížová validace

Pro posouzení predikční kvality modelu jsem použila křížovou validaci, při které je odstraněna vždy jedna měřící stanice z výpočtu krigingu a na základě hodnot zbývajících stanic je predikce provedena pro odstraněnou stanici. [4] Takto je postup zopakován tolikrát, dokud není postupně odstraněna každá měřící stanice a nebyly tak vypočteny odhady pro všechna místa měření.

Na základě známých a predikovaných hodnot jsem pak spočítala charakteristiky kvality modelu jako jsou již zmiňované R^2 a $RMSE$. V případě křížové validace je však interpretace $RMSE$ odlišná. Zde nám $RMSE$ vyjadřuje, jak přesně zvolený model predikuje naše data. Jde tedy o průměrnou odchylku predikovaných a napozorovaných hodnot v lokalitách \mathbf{x}_i . Jako další ukazatele si uvedeme korelační koeficient mezi pozorovanými a odhadnutými hodnotami a korelační koeficient mezi odhadnutými hodnotami a rezidui.

- $cor_{Z(\mathbf{x}_i), \hat{Z}(\mathbf{x}_i)}$ - korelační koeficient mezi pozorovanými a odhadnutými hodnotami. Nabývá hodnot od -1 do 1. V tomto případě preferujeme hodnoty co nejbližší 1.

$$cor_{Z(\mathbf{x}_i), \hat{Z}(\mathbf{x}_i)} = \frac{cov[Z(\mathbf{x}_i), \hat{Z}(\mathbf{x}_i)]}{\sqrt{var[Z(\mathbf{x}_i)]var[\hat{Z}(\mathbf{x}_i)]}},$$

kde $var[Z(\mathbf{x}_i)]$ značí rozptyl pozorovaných hodnot $Z(\mathbf{x}_i)$, $var[\hat{Z}(\mathbf{x}_i)]$ rozptyl odhadnutých hodnot $\hat{Z}(\mathbf{x}_i)$.

• $cor_{residual(\mathbf{x}_i), \hat{Z}(\mathbf{x}_i)}$ - korelační koeficient mezi odhadnutými hodnotami a rezidui. Preferujeme hodnoty kolem 0.

$$cor_{residual(\mathbf{x}_i), \hat{Z}(\mathbf{x}_i)} = \frac{cov[residual(\mathbf{x}_i), \hat{Z}(\mathbf{x}_i)]}{\sqrt{var[residual(\mathbf{x}_i)]var[\hat{Z}(\mathbf{x}_i)]}},$$

kde $residual(\mathbf{x}_i) = Z(\mathbf{x}_i) - \hat{Z}(\mathbf{x}_i)$.

Křížová validace je prováděna na logaritmovaných datech, pro která byl namodelován také variogram a proveden následný kriging. Výsledné hodnoty charakteristik stejně jako diagnostické grafy jsou uvedeny pro odlogaritmovaná data, což může případně pomoci identifikovat oblasti s problematickými hodnotami srážkových úrhnů.

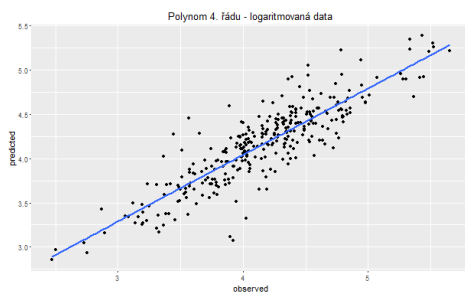
Model	RMSE	R^2	$cor_{Z(s_i), \hat{Z}(s_i)}$	$cor_{residual(s_i), \hat{Z}(s_i)}$
ZK	21.424	0.750	0.868	0.114
UK (lineární trend)	21.683	0.744	0.864	0.101
UK (polynom 4. řádu)	22.260	0.730	0.855	0.078

Tabulka 4.3: Křížová validace - odlogaritmovaná data

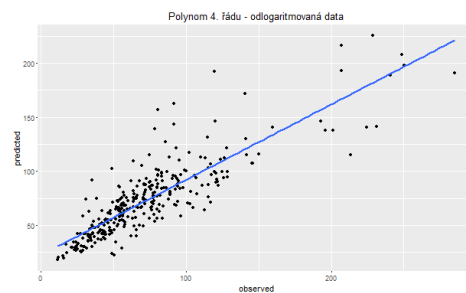
Z tabulky 4.3 vidíme, že charakteristiky kvality modelu $RMSE$ a R^2 , stejně jako korelační koeficienty, se mezi jednotlivými modely liší minimálně. Proto lze konstatovat, že nesplnění podmínky stacionarity nesnižuje predikční schopnosti modelů.

Podívejme se ještě na některé diagnostické grafy. Zde se budu již zabývat pouze matematicky správným modelem, kterým je univerzální kriging s trendem ve tvaru polynomu 4. řádu. Nejprve se podívejme na graf znázorňující pozorované

a odhadnuté hodnoty. Tento graf můžeme vidět na obrázku 4.20 pro logaritmovaná data, na obrázku 4.21 pro odlogaritmovaná data. Na základě obou grafů lze říci, že model nemá tendenci systematicky nadhodnocovat či podhodnocovat predikované hodnoty. Problematická měření nabývají spíše průměrných hodnot a tak lze usuzovat, že velké odchylky predikovaných hodnot od pozorovaných plynou spíše ze specifity příslušných stanic.

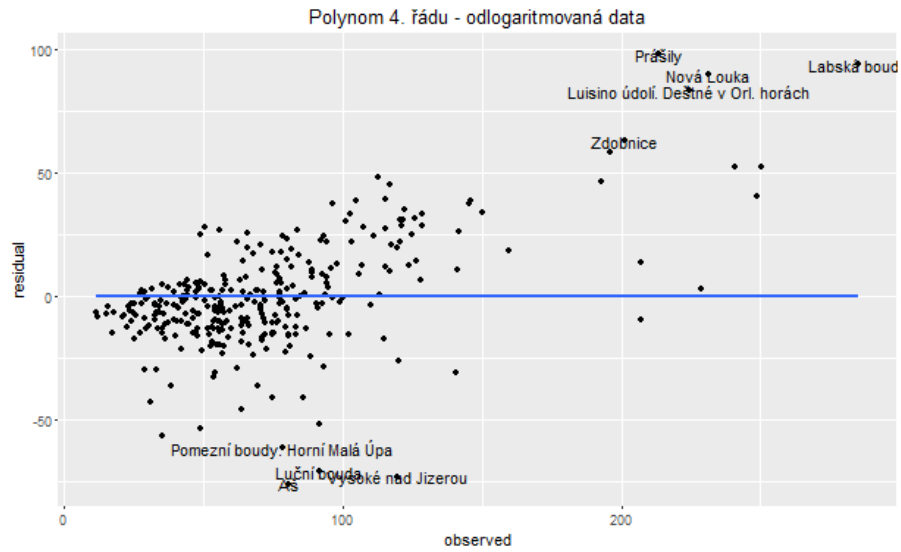


Obrázek 4.20: Pozorované, predikované hodnoty - logaritmovaná data

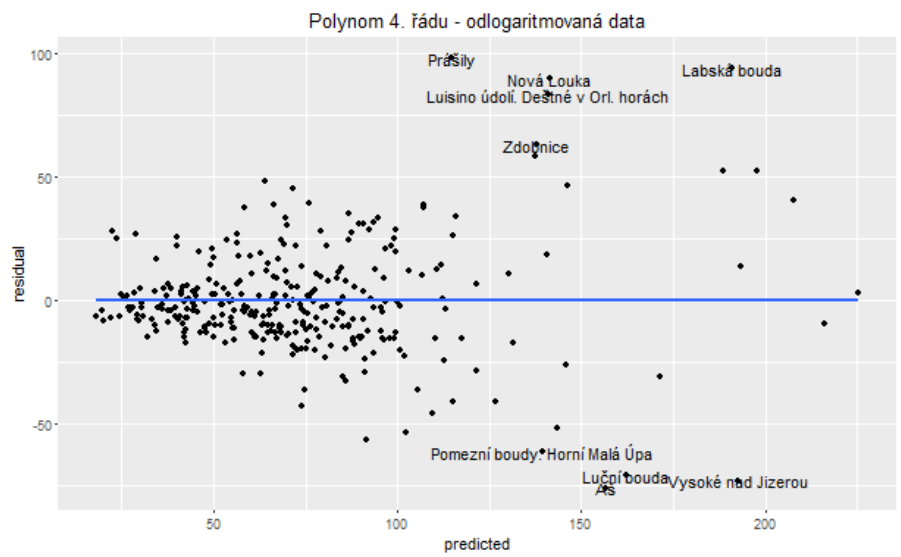


Obrázek 4.21: Pozorované, predikované hodnoty - odlogaritmovaná data

Graf 4.22 zachycuje vykreslená pozorování proti reziduům, a to pro odlogaritmovaná data. Navíc jsou v něm uvedeny názvy měřících stanic, kterým odpovídají hodnoty reziduí v absolutní hodnotě větší než 60. Můžeme si povšimnout, že některým lokalitám, ve kterých dosahovaly úhrny srážek kolem 200 mm a více, přísluší vysoké kladné hodnoty reziduí. Mezi tyto stanice patří např. měřící stanice ve Zdobenicích, Luisunu údolí, v Prášilech, Nové Louce či Labské boudě. Navíc jsou pak problematická měření pořízená v Aši, Luční boudě, Vysokém nad Jizerou, Pomezní boudě a Horní Malé Úpě, kde se úhrny srážek pohybovaly kolem 100 mm, a tak patřily do intervalu hodnot, kde se pohybovalo velké množství měření. Obdobný graf můžeme vidět na obrázku 4.23, kde jsou zachyceny odhadnuté hodnoty a rezidua s vyznačenými problematickými stanicemi. Z grafů 4.22 a 4.23 je patrné, že zkonstruovaný model podhodnocuje některá měření, která dosahovala nadprůměrných hodnot a nadhodnocuje několik málo měření, která dosahují průměrných hodnot. Vzhledem k tomu, že se jedná jen o několik měření z celkového počtu 318, není tento jev systematický.



Obrázek 4.22: Pozorované hodnoty, rezidua - odlogaritmovaná data



Obrázek 4.23: Odhadnuté hodnoty, rezidua - odlogaritmovaná data

Závěr

V této práci jsem se zabývala daty, která jsou vztažena k určitému území. Proto byly nástroje k jejich zpracování z oblasti geostatistiky.

V první kapitole je představen Český hydrometeorologický ústav, z jehož internetových stránek jsem čerpala data. Také je zde popsán souřadnicový systém S-42, který je pro zpracování prostorových dat vhodnější než klasický zeměpisný souřadnicový systém.

V druhé kapitole jsou nadefinovány stěžejní pojmy jako jsou náhodná funkce, regionalizovaná proměnná či variogram. Pro srovnání je v této kapitole uvedena také interpolační metoda IDW, která je však čistě numerickou metodou.

Třetí kapitola pak pojednává o teorii krigingu, kde byl také odvozen systém krigingových rovnic. V této kapitole byly rovněž uvedeny různé typy krigingu, a to základní, jednoduchý, univerzální a lognormální kriging.

Hlavní částí práce je pak čtvrtá kapitola, která obsahuje praktickou část. Zde jsem výše získané poznatky aplikovala na reálná data - měsíční úhrny srážek. Nejprve byly vytvořeny srážkové mapy metodou IDW. Poté následovala tvorba map metodou kriging, kdy jsem nejprve prostřednictvím reziduí vyšetřovala stacionaritu dat. Rezidua vykazovala stacionární chování až při proložení souřadnic trendem ve tvaru polynomu 4. řádu. Pro srovnání jsem dále uvažovala i modely s konstantním a lineárním trendem. Po ověření stacionarity následovalo modelování variogramu, kdy se ukázalo, že variogram nemá stejné chování ve všech směrech, což vedlo k anizotropnímu modelu variogramu s hlavním směrem 45° a vedlejším 135° . Výstupem této metody byly opět srážkové mapy, které můžeme vizuálně porovnat se srážkovými mapami vytvořenými metodou IDW. Pro posou-

zení kvality modelu byly zvoleny charakteristiky RMSE a R^2 . Podle obou charakteristik jsou všechny modely vytvořené pro kriging uspokojivé. Kvalita predikcí modelů pak byla posuzována na základě křížové validace. Také predikční schopnosti modelů se ukázaly jako obstojné ve všech případech a tak lze konstatovat, že nesplnění stacionarity nemá v tomto případě výrazný vliv na schopnost modelů predikovat hodnoty.

Při psaní diplomové práce jsem získala mnoho zkušeností. Měla jsem poprvé možnost zpracovat text takového rozsahu v typografickém systému \TeX . Vůbec poprvé jsem pak pracovala v programu CorelDRAW. Také jsem si prohloubila dovednosti v oblasti použití statistického softwaru R. Za nejpřínosnější však považuji práci s reálnými daty, kdy člověk dopředu neví, co všechno si na něj „připraví“. Díky tomu, že data, která jsem zpracovávala, porušovala podmínku stacionarity a při modelování variogramu se vyskytovaly potíže s anizotropií, jsem měla možnost nahlédnout do problematiky krigingu snad i trochu hlouběji.

Literatura

- [1] Armstrong, M.: *Basic linear geostatistics*. Springer: Berlin, Heidelberg, 1998.
- [2] Bivand, R.S., Pebesma, E.J., Gómez-Rubio, V.: *Applied Spatial Data Analysis with R*. Springer, New York, 2008.
- [3] Cressie, N.A.C.: *Statistics for Spatial Data*. Wiley, New York, 1993.
- [4] Kastelec, D., Košmelj, K.: *Spatial Interpolation of Mean Yearly Precipitation using Universal Kriging*. Metodološki zvezki, Ljubljana, 2002.
- [5] Kovalová, K.: *Geostatistika pro kompoziční data*. Univerzita Palackého, Olomouc, 2014.
- [6] Wackernagel, H.: *Multivariate Geostatistics: An Introduction with Applications*. Springer: Berlin, Heidelberg, 2010.
- [7] Zůvala, R., Fišerová, E., Marek, L.: *Mathematical aspects of the kriging applied on landslide in Halenkovice (Czech Republic)*. Open Geosci, 2016.
- [8] Zůvala, R.: *Analýza sesuvu svahu pomocí metody kriging*. Univerzita Palackého, Olomouc, 2011.
- [9] ČHMÚ – Aktuální informace hydrologické předpovědní služby [online]. [cit. 2015-11-1]. Dostupné z: http://hydro.chmi.cz/hpps/hpps_act_rain.php.
- [10] ČHMÚ – Aktuální odhad srážek z radarů a srážkoměrů [online]. [cit. 2016-2-10]. Dostupné z: http://hydro.chmi.cz/hpps/hpps_srzinfo.php.
- [11] Geographic Information Technology Training Alliance - Distance Based Interpolation [online]. [cit. 2016-2-10]. Dostupné z: http://www.gitta.info/ContiSpatVar/en/html/Interpolation_learningObject2.xhtml.
- [12] Geomatika - Kriging [online]. [cit. 2016-2-13]. Dostupné z: <http://geomatika.kma.zcu.cz/studium/ugi/referaty/05/ArchivReferatu/KRIGING.pdf>.

- [13] Solution to Ordinary and Universal Kriging Equations [online]. [cit. 2016-4-7]. Dostupné z: <http://www-hsc.usc.edu/~mereditf/KrigingEquations.pdf>.
- [14] Souřadnicové systémy [online]. [cit. 2016-2-10]. Dostupné z: <http://tvorbamap.shocart.cz/kartografie/systemy.htm>.
- [15] Wikipedie – Český hydrometeorologický ústav [online]. [cit. 2016-2-4]. Dostupné z: https://cs.wikipedia.org/wiki/Český_hydrometeorologický_ústav
- [16] Wordpress – Systém S-42 a GPS [online]. [cit. 2016-2-10]. Dostupné z: <https://hrus.wordpress.com/2008/12/01/system-s-42-a-gps/>.