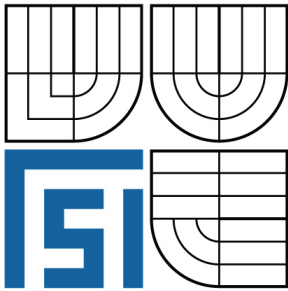


VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA STROJNÍHO INŽENÝRSTVÍ  
ÚSTAV MATEMATIKY

FACULTY OF MECHANICAL ENGINEERING  
INSTITUTE OF MATHEMATICS

# REGRESNÍ METODY PRO STATISTICKOU ANALÝZU PROSTOROVÝCH DAT

REGRESSION METHODS FOR STATISTICAL ANALYSIS OF SPATIAL DATA

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. LUCIE KLIMPROVÁ

VEDOUCÍ PRÁCE

SUPERVISOR

doc. RNDr. JAROSLAV MICHÁLEK,  
CSc.

BRNO 2009



Vysoké učení technické v Brně, Fakulta strojního inženýrství

Ústav matematiky

Akademický rok: 2008/2009

## **ZADÁNÍ DIPLOMOVÉ PRÁCE**

student(ka): Bc. Lucie Klimprová

který/která studuje v **magisterském navazujícím studijním programu**

obor: **Matematické inženýrství (3901T021)**

Ředitel ústavu Vám v souladu se zákonem č.111/1998 o vysokých školách a se Studijním a zkušebním řádem VUT v Brně určuje následující téma diplomové práce:

### **Regresní metody pro statistickou analýzu prostorových dat**

v anglickém jazyce:

### **Regression methods for statistical analysis of spatial data**

Stručná charakteristika problematiky úkolu:

Častou úlohou při analýze prostorových dat je nalezení optimální predikce (vzhledem ke střední kvadratické chybě) kovariančně stacionárního procesu definovaného v dané oblasti. Problémem při tom je, že není známa kovarianční struktura tohoto procesu, je pouze popsána variogramem a je potřeba odhadnout ji z dat. V literatuře je popsána řada přístupů k odhadu variogramu a není vždy zcela jasný vliv tohoto odhadu na sledovanou predikci.

Cíle diplomové práce:

Cílem práce je popsat vybrané regresní metody pro predikci kovariančně stacionárního procesu, podle literatury vybrat případně navrhnout metody pro odhad variogramu a studovat vliv tohoto odhadu na predikci. Dále srovnat různé modely pro odhad variogramu a na simulovaných nebo reálných datech posoudit jejich vliv na predikci procesu. Programové prostředí doporučené pro programovou implementaci navržených metod je MATLAB.

Seznam odborné literatury:

Cressie, N.A.C.: Statistics for spatial data. John-Wiley Inc., New York, 1993.

Michálek, J. Dvě poznámky o metodě kriging. Sborník Robust 2000, JČMF Praha, 2000, s. 189 - 197.

Shapiro, A. and Botha, J.D.: Variogram fitting with a general class of conditionally positive definite functions. Computational Statistics and Data Analysis. 11, 1991 p.87-96.

Vedoucí diplomové práce: doc. RNDr. Jaroslav Michálek, CSc.

Termín odevzdání diplomové práce je stanoven časovým plánem akademického roku 2008/2009.

V Brně, dne 20.11.2008

L.S.

---

prof. RNDr. Josef Šlapal, CSc.  
Ředitel ústavu

---

doc. RNDr. Miroslav Doupovec, CSc.  
Děkan fakulty

## **Abstrakt**

Regresní metodou používanou pro vyhodnocování prostorových spojitých procesů je metoda kriging. Pokud je neznámá kovarianční struktura procesu, je potřeba odhadnout ji z dat. První, teoretická část, je věnována právě popisu metody kriging a odhadu variogramu, který popisuje kovarianční strukturu uvažovaného procesu. Druhá, praktická část, obsahuje programovou implementaci v MATLABu metody kriging na simulovaných a reálných datech.

## **Abstract**

Kriging techniques are regression methods used for evaluation of continuous spatial processes. If the covariance structure of process is unknown, then it's necessary to estimate it from the data. The first part of this Master's thesis is devoted to description the kriging method and to estimate of a variogram function, which describes the covariance structure of considered process. The second part includes the implementation of kriging method in MATLAB for simulated and real data.

## **Klíčová slova**

Regresní model, prostorová data, kriging odhady, univerzální kriging model, odhad variogramu.

## **Keywords**

Regression model, spatial data, kriging estimations, universal kriging model, estimation of variogram function.

## **Bibliografická citace**

KLIMPROVÁ, L. Regresní metody pro statistickou analýzu prostorových dat. Brno: Vysoké učení technické v Brně, Fakulta strojního inženýrství, 2009. 83 s. Vedoucí diplomové práce doc. RNDr. Jaroslav Michálek, CSc.



Prohlašuji, že jsem svou diplomovou práci *Regresní metody pro statistickou analýzu prostorových dat* vypracovala samostatně pod vedením doc. RNDr. Jaroslava Michálka, CSc., s použitím materiálů uvedených v seznamu literatury.

.....  
Bc. Lucie Klimprová





Děkuji svému vedoucímu doc. RNDr. Jaroslavu Michálkovi, CSc. za cenné rady, vstřícnost, ochotu, poskytnuté materiály a výborné vedení při vytváření mé diplomové práce. Dále bych chtěla poděkovat své rodině za podporu během celého studia.



# Obsah

<b>1</b>	<b>Úvod</b>	<b>10</b>
<b>2</b>	<b>Základní pojmy</b>	<b>11</b>
2.1	Pravděpodobnostní prostor . . . . .	11
2.2	Náhodná veličina a její charakteristiky . . . . .	11
2.2.1	Diskrétní náhodná veličina . . . . .	12
2.2.2	Spojité náhodná veličina . . . . .	12
2.2.3	Číselné charakteristiky náhodné veličiny . . . . .	13
2.2.4	Podmíněná střední hodnota . . . . .	14
<b>3</b>	<b>Regresní model</b>	<b>16</b>
3.1	Regresní funkce . . . . .	16
3.2	Lineární regresní model . . . . .	16
<b>4</b>	<b>Prostorová data</b>	<b>20</b>
4.1	Náhodný proces a jeho charakteristiky . . . . .	20
4.2	Stacionarita, kovariogram a variogram . . . . .	20
4.2.1	Stacionarita . . . . .	20
4.2.2	Variogram . . . . .	21
4.2.3	Kovariogram a korelogram . . . . .	22
4.3	Lineární modely se statisticky závislou chybou . . . . .	22
4.3.1	Model s plnou hodnotí . . . . .	23
4.4	Příklady prostorových dat . . . . .	25
4.4.1	Simulovaná data . . . . .	25
4.4.2	Reálná data . . . . .	27
<b>5</b>	<b>Kriging metoda</b>	<b>29</b>
5.1	Univerzální kriging model . . . . .	29
5.2	Kriging odhady . . . . .	30
5.2.1	Rovnice univerzálního krigingu . . . . .	31
5.2.2	Kriging rozptyl a intervaly predikce . . . . .	32
5.2.3	Odhad střední hodnoty . . . . .	33
5.3	Další kriging modely . . . . .	33
5.3.1	Ordinární kriging . . . . .	33
5.3.2	Simple kriging . . . . .	33
<b>6</b>	<b>Odhad variogramu</b>	<b>34</b>
6.1	Klasický odhad variogramu . . . . .	34
6.2	Robustní odhady variogramu . . . . .	34
6.3	Základní typy modelových variogramů . . . . .	34
6.4	Jiný přístup k modelování variogramu . . . . .	39
<b>7</b>	<b>Popis programu</b>	<b>40</b>
7.1	Zadání vstupních dat . . . . .	40
7.2	Funkce Mezikruží . . . . .	41
7.3	Klasický odhad variogramu . . . . .	42
7.4	Základní typy modelových variogramů . . . . .	42
7.5	Kriging odhad . . . . .	43

7.6	Celkový odhad . . . . .	43
<b>8</b>	<b>Kriging metoda pro regresní rovinu</b>	<b>44</b>
8.1	Odhady regresní roviny . . . . .	45
8.1.1	Lineární variogram . . . . .	45
8.1.2	Sférický variogram . . . . .	46
8.1.3	Racionální kvadratický variogram . . . . .	48
8.1.4	Exponenciální variogram . . . . .	49
8.1.5	Mocinný variogram . . . . .	50
8.1.6	Vlnový variogram . . . . .	51
8.2	Srovnání výsledků . . . . .	52
8.3	Zjemnění $x$ -ové a $y$ -ové škály měřítka souřadnic odhadu . . . . .	53
<b>9</b>	<b>Kriging metoda pro obecnou lineární regresní plochu</b>	<b>57</b>
9.1	Kriging odhady . . . . .	58
9.1.1	Odhady $Z^*$ regresní funkce . . . . .	58
9.1.2	Grafické znázornění výsledků . . . . .	60
9.1.3	Reziduální analýza . . . . .	63
9.2	Zjemnění měřítek vektorů $x_0$ a $y_0$ . . . . .	65
<b>10</b>	<b>Reálný příklad pro Coal-Ash Data</b>	<b>69</b>
10.1	Kriging odhady . . . . .	70
10.1.1	Odhady regresní funkce . . . . .	70
10.1.2	Grafické znázornění výsledků . . . . .	72
10.1.3	Reziduální analýza . . . . .	75
10.2	Srovnání výsledků . . . . .	77
10.3	Zjemnění škály měřítka souřadnic bodů predikce . . . . .	78
<b>11</b>	<b>Závěr</b>	<b>82</b>
<b>12</b>	<b>Seznam použitých zdrojů</b>	<b>83</b>

# 1 Úvod

Diplomová práce zpracovává téma Regresní metody pro statistickou analýzu prostorových dat.

Regresní analýza je často používanou technikou hledání a zkoumání závislostí proměnných, jejichž hodnoty získáme při realizaci experimentu.

Statistika se v minulosti hodně rozvíjela a určitý metodický rozvoj nejruznějších odvětví statistiky přetrvává dodnes. Jednou z oblastí, která se v posledních letech významně rozvinula, je oblast prostorové statistiky. Prostorová statistika byla zpočátku spojována zejména s přírodními vědami (např. s geologií), a proto francouzský matematik G. Matheron použil jako první pro prostorovou statistiku termín *geostatistika*. Název geostatistika se dodnes celosvětově užívá jako označení disciplíny zahrnující specifické metody zpracování dat měřených v prostoru nebo v ploše.

V geostatistice je jednou z nejčastěji používaných metod odhadu metoda kriging. Je oblíbená mezi inženýry (zejména v geologii, meteorologii, apod.) a je pojmenována po D. E. Krigeovi, jihoafrickém inženýrovi, který v padesátých letech rozvinul empirické statistické metody v důlním inženýrství.

Podstatou této metody je nalézt optimální predikci kovariančně stacionárního náhodného procesu definovaného v dané oblasti vzhledem ke střední kvadratické chybě. Problémem je, že obvykle není známa kovarianční struktura uvažovaného procesu a je potřeba odhadnout ji z dat.

První, teoretická část diplomové práce, obsahuje připomenutí základních pojmů a označení z matematické statistiky, popis regresního modelu a vlastností prostorových dat, zavedení kriging metody a možnost odhadu kovarianční struktury, zejména odhadu variogramu.

Druhá, praktická část diplomové práce, obsahuje popis programové implementace v prostředí MATLAB a demonstraci kriging metody na simulovaných a reálných datech.

## 2 Základní pojmy

Nejprve uvedeme základní pojmy potřebné pro porozumění zpracovávanému tématu. Jelikož jde o zavedení označení a připomenutí známých pojmů, nebudeme je striktně definovat.

### 2.1 Pravděpodobnostní prostor

Modely pravděpodobnostní a statistické patří mezi matematické modely, které se používají při studiu reálných dějů. Reálné děje se snažíme nejen studovat a popisovat, ale hlavně nalézt jejich charakteristické vlastnosti a využít je k predikci chování uvažovaných dějů v prostoru nebo v čase.

Pokud realizujeme pokus, jednotlivé výsledky označíme  $\omega$  a nazýváme je *elementárními jevy*. Množinu všech možných výsledků pokusu (tj. množinu všech elementárních jevů) nazýváme *prostorem elementárních jevů* a značíme ji  $\Omega$ .

Symbolem  $\mathcal{A}$  značíme  $\sigma$ -algebru podmnožin prostoru  $\Omega$ . Prvky  $A \in \mathcal{A}$  se nazývají *náhodné jevy*. Vlastnosti  $\sigma$ -algebry jsou:

- Jestliže  $A \in \mathcal{A}$ , potom  $\bar{A} \in \mathcal{A}$ .
- Jestliže  $A_i \in \mathcal{A}$ ,  $i = 1, 2, \dots$ , potom  $\bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$ .

Dále definujeme *pravděpodobnost*  $P(A)$  náhodného jevu  $A \in \mathcal{A}$ . Pravděpodobnost  $P(A)$  je reálná funkce definovaná na  $\sigma$ -algebře  $\mathcal{A}$  s vlastnostmi:

- $P(A) \geq 0$  pro všechny náhodné jevy  $A \in \mathcal{A}$ .
- $P(\Omega) = 1$ .
- Jestliže  $A_i \in \mathcal{A}$ ,  $i = 1, 2, \dots$  jsou po dvou disjunktní náhodné jevy, potom

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

Trojice  $(\Omega, \mathcal{A}, P)$  se nazývá *pravděpodobnostní prostor*.

### 2.2 Náhodná veličina a její charakteristiky

*Náhodnou veličinou* (vzhledem k jevovému poli  $\mathcal{A}$ ) rozumíme zobrazení  $X : \Omega \rightarrow \mathbf{R}$ , pro které je množina  $\{\omega \in \Omega : X(\omega) \leq x\}$  jevem v  $\mathcal{A}$  pro každé  $x \in \mathbf{R}$ .

Pro náhodnou veličinu  $X$  definujeme *distribuční funkci*  $F(x) = P(X \leq x)$ , která je reálnou funkcí definovanou na  $(-\infty, +\infty)$  s vlastnostmi:

- $0 \leq F(x) \leq 1$  pro  $\forall x \in (-\infty, +\infty)$ .
- $F(x)$  je neklesající, zprava spojitá a má nejvýše spočetně mnoho bodů nespojitosti na  $(-\infty, +\infty)$ .
- $\lim_{x \rightarrow -\infty} F(x) = F(-\infty) = 0$  a  $\lim_{x \rightarrow +\infty} F(x) = F(+\infty) = 1$ .

- $P(a < X \leq b) = F(b) - F(a)$  pro libovolná reálná čísla  $a < b$ , speciálně:  
 $P(a < X) = 1 - F(a)$ ,  $P(X \leq b) = F(b)$
- $P(X = c) = F(c) - \lim_{x \rightarrow c^-} F(x)$  pro libovolné reálné číslo  $c$ .

### 2.2.1 Diskrétní náhodná veličina

Máme náhodnou veličinu  $X$  definovanou na pravděpodobnostním prostoru  $(\Omega, \mathcal{A}, P)$  a množinu  $M$ , která obsahuje nejvýše spočetně mnoho reálných čísel takových, že  $\sum_{x \in M} P(X = x) = 1$ . O náhodné veličině  $X$  potom řekneme, že má *diskrétní rozdělení pravděpodobnosti*. Množina  $M$  se nazývá obor hodnot  $X$  a funkce  $p(x)$  definovaná vztahy:

$$\begin{cases} p(x) = P(X = x), & x \in M, \\ p(x) = 0, & \text{jinak,} \end{cases} \quad (2.1)$$

se nazývá *pravděpodobnostní funkce* náhodné veličiny  $X$ . Označením  $X \sim (M, p)$  rozumíme, že náhodná veličina  $X$  má diskrétní rozdělení pravděpodobnosti s oborem hodnot  $M$  a pravděpodobnostní funkcí  $p$ .

Nechť  $X \sim (M, p)$ , potom platí:

- $p(x) \geq 0$  a  $\sum_{x \in M} p(x) = 1$ ,
- $p(x) = F(x) - \lim_{y \rightarrow x^-} F(y)$ ,
- $P(X \in B) = \sum_{x \in B \cap M} p(x)$ ,  $B$  je libovolná podmnožina  $\mathbf{R}$ ,
- $F(x) = \sum_{t \in M \cap (-\infty, x]} p(t)$ .

### 2.2.2 Spojitá náhodná veličina

Říkáme, že náhodná veličina  $X$  je *spojitá* (má spojité rozdělení pravděpodobnosti), pokud existuje nezáporná funkce  $f(x)$ ,  $x \in \mathbf{R}$ , taková, že  $\int_{-\infty}^{\infty} f(x) dx = 1$ . Distribuční funkci  $F(x)$  náhodné veličiny  $X$  lze pomocí  $f(x)$  napsat ve tvaru

$$F(x) = \int_{-\infty}^x f(t) dt.$$

Funkce  $f$  se nazývá *hustota* (rozdělení pravděpodobnosti) náhodné veličiny  $X$ .

Pokud máme náhodnou veličinu  $X$  spojitěho typu s distribuční funkcí  $F$  a hustotou  $f$ , potom platí:

- $f(x) \geq 0$ ,  $x \in \mathbf{R}$ ,
- $\int_{-\infty}^{\infty} f(x) dx = 1$ ,
- $\int_{-\infty}^x f(t) dt = F(x)$ ,
- $f(x) = F'(x)$  pokud derivace existuje,
- $P(X \in B) = \int_B f(x) dx$  pro libovolnou borelovskou množinu  $B$ ,
- pro libovolné  $h > 0$ :  $P(x < X < x + h) = f(x)h + o(h)$ , když existuje derivace funkce  $F$  v bodě  $x$  a  $\frac{o(h)}{h} \rightarrow 0$ , když  $h \rightarrow 0$ .

### 2.2.3 Číselné charakteristiky náhodné veličiny

Pro náhodné veličiny definujeme *střední hodnotu* jako:

$$E(X) = \begin{cases} \sum_x xp(x), & \text{pro diskrétní náhodnou veličinu} \\ & \text{(pokud řada konverguje absolutně),} \\ \int_{-\infty}^{\infty} xf(x), & \text{pro spojitou náhodnou veličinu} \\ & \text{(pokud integrál konverguje absolutně).} \end{cases}$$

Střední hodnota má vlastnosti (pokud všechny uvažované střední hodnoty existují):

- $E(aX + b) = aE(X) + b$  pro všechna  $a, b \in \mathbb{R}$ .
- $E(\sum_{i=1}^n X_i) = \sum_{i=1}^n E(X_i)$ .
- $E(\prod_{i=1}^n X_i) = \prod_{i=1}^n E(X_i)$ , jsou-li náhodné veličiny  $X_1, \dots, X_n$  nezávislé.

Dále definujeme *rozptyl* náhodné veličiny  $X$  vztahem:

$$D(X) = E([X - E(X)]^2),$$

pokud uvedená střední hodnota existuje, a *směrodatnou odchylku* náhodné veličiny  $X$ :

$$\sigma(X) = \sqrt{D(X)}.$$

Pro rozptyly náhodných veličin  $X, X_1, X_2, \dots, X_n$  platí:

- $D(X) \geq 0$ .
- $D(aX + b) = a^2D(X)$  pro všechna  $a, b \in \mathbb{R}$ .
- $D(X) = E(X^2) - [E(X)]^2$ , pokud potřebné střední hodnoty existují.
- $D(\sum_{i=1}^n X_i) = \sum_{i=1}^n D(X_i)$ , jsou-li náhodné veličiny  $X_1, X_2, \dots, X_n$  nezávislé.

Pokud máme náhodné veličiny  $X$  a  $Y$  definované na pravděpodobnostním prostoru  $(\Omega, \mathcal{A}, P)$ , jejich vzájemný vztah můžeme vyjádřit *kovariancí*, kterou definujeme vztahem:

$$\text{cov}(X, Y) = E([X - E(X)][Y - E(Y)]) = E(XY) - E(X)E(Y),$$

za předpokladu, že uvedené střední hodnoty existují. Kovariance má vlastnosti:

- $\text{cov}(X, Y) = \text{cov}(Y, X)$ ,
- $\text{cov}(X, X) = D(X)$ ,
- $D(X + Y) = D(X) + D(Y) + 2\text{cov}(X, Y)$ ,
- $X, Y$  nezávislé  $\Rightarrow \text{cov}(X, Y) = 0$  a  $E(X, Y) = E(X)E(Y)$ , pokud  $E(X)$  a  $E(Y)$  existují,
- $\text{cov}(aX + b, cY + d) = ac \text{cov}(X, Y)$  pro libovolná reálná čísla  $a, b, c, d$ ,
- $|\text{cov}(X, Y)| \leq \sqrt{D(X)D(Y)} = \sigma_x \sigma_y$ .



Je-li  $\mathbf{X} = (X_1, \dots, X_n)'$  náhodný vektor a existují  $C(X_i, X_j); i, j = 1, \dots, n$ . Potom matice  $var(\mathbf{X})$ ,

$$var(\mathbf{X}) = \begin{pmatrix} D(X_1) & C(X_1, X_2) & \dots & C(X_1, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ C(X_n, X_1) & C(X_n, X_2) & \dots & D(X_n) \end{pmatrix},$$

se nazývá varianční, je symetrická a pozitivně definitní.

Vzájemný vztah náhodných veličin  $X$  a  $Y$  můžeme vyjádřit také pomocí *korelačního koeficientu*:

$$\rho(X, Y) = cov\left(\frac{X - E(X)}{\sigma(X)}, \frac{Y - E(Y)}{\sigma(Y)}\right) = \frac{cov(X, Y)}{\sigma(X)\sigma(Y)} = \frac{cov(X, Y)}{\sqrt{D(X)D(Y)}},$$

pokud  $E(X)$  a  $E(Y)$  existují. Korelační koeficient má vlastnosti:

- $\rho(X, Y) = \rho(Y, X)$ ,
- $\rho(X, X) = 1$ ,
- $-1 \leq \rho(X, Y) \leq 1$ ,
- $\rho(aX + b, cY + d) = \frac{ac}{|ac|}\rho(X, Y)$  pro libovolná reálná čísla  $a, b, c, d; ac \neq 0$ ,
- $Y = aX + b \Leftrightarrow |\rho(X, Y)| = 1$ , kde  $a, b \in \mathbb{R}; a \neq 0$ ,
- $X, Y$ - nezávislé  $\Rightarrow \rho(X, Y) = 0$ .

Pokud  $\rho(X, Y) = 0$ , říkáme, že náhodné veličiny jsou nekorelované. Nezávislé náhodné veličiny jsou nekorelované, ale nekorelované náhodné veličiny nemusí být nezávislé.

Koeficienty korelace můžeme podobně jako kovariance zapsat do matice a dostaneme tak tzv. *korelační matici* náhodného vektoru  $\mathbf{X} = (X_1, \dots, X_n)'$ :

$$\rho(\mathbf{X}) = \begin{pmatrix} 1 & \rho(X_1, X_2) & \dots & \rho(X_1, X_n) \\ \rho(X_2, X_1) & 1 & \dots & \rho(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(X_n, X_1) & \rho(X_n, X_2) & \dots & 1 \end{pmatrix},$$

která je symetrická a pozitivně semidefinitní.

## 2.2.4 Podmíněná střední hodnota

- Diskrétní případ:

Nechť  $(X, Y) \sim (M, p)$ ,  $p_x, p_y$  jsou marginární pravděpodobnostní funkce,  $p(x, y) = P(X = x, Y = y), (x, y) \in M, p_x(x) \neq 0$ ,  $M_x$  je definiční obor  $x$  a  $M_y$  je definiční obor  $y$ . Potom funkce

$$p(y|x) = \frac{p(x, y)}{p_x(x)}$$

se nazývá *podmíněná pravděpodobnostní funkce* náhodné veličiny  $Y$  za podmínky  $X = x$ . Dále za daných podmínek definujeme vztahem

$$E(Y|X = x) = \sum_{y \in M_y} yp(y|x)$$

*podmíněnou střední hodnotu* náhodné veličiny  $Y$  za podmínky  $X = x$ .

- Spojitý případ:

Nechť  $(X, Y)$  má sdruženou hustotu  $f(x, y)$  a marginální hustota  $f_x(x) \neq 0$ . Potom funkci

$$f(y|x) = \frac{f(x, y)}{f_x(x)}$$

nazýváme *podmíněnou hustotou* náhodné veličiny  $Y$  vzhledem k  $X$ . Dále definujeme funkci

$$E(Y|X = x) = \int_{-\infty}^{\infty} f(y|x)dy,$$

kde  $x \in M_x$ . Funkce  $E(Y|X = x)$  se nazývá *podmíněná střední hodnota* náhodné veličiny  $Y$  za podmínky  $X = x$ .

Podmíněná střední hodnota  $E(Y|X)$  má analogické vlastnosti jako střední hodnota nepodmíněná. Uvedené vztahy lze rozšířit pro náhodný vektor  $\mathbf{X}$  a dostaneme podmíněnou střední hodnotu  $E(Y|\mathbf{X} = \mathbf{x})$ . Důležitou vlastnost, kterou využijeme v další kapitole, popisuje následující věta.

**Věta 2.1.** *Nechť  $Z$  je náhodná veličina,  $\mathbf{X}$  je náhodný vektor,  $E(Z^2) < \infty$  a vektor  $(Z, \mathbf{X})'$  je diskrétního nebo spojitého typu. Potom pro každou měřitelnou funkci  $g(\mathbf{x})$  platí*

$$E[Z - g(\mathbf{X})]^2 \geq E[Z - E(Z|\mathbf{X})]^2.$$

*Rovnost nastává právě tehdy, platí-li  $g(\mathbf{X}) = E(Z|\mathbf{X})$  s pravděpodobností 1.*

Věta (2.1) říká, že ze všech funkcí vektoru  $\mathbf{X}$  je veličině  $Z$  "nejblíže" její podmíněná střední hodnota, za podmínky, že vzdálenost posuzujeme pomocí střední kvadratické chyby. Pokud chceme neznámou hodnotu  $Z$  předpovědět ze známých hodnot vektoru  $\mathbf{X}$ , nejlepší předpovědí je právě podmíněná střední hodnota  $E(Z|\mathbf{X})$ . Protože výpočet podmíněné střední hodnoty může být složitým úkolem, často hledáme nejlepší predikci veličiny  $Z$  mezi lineárními funkcemi vektoru  $\mathbf{X}$ . V případě normálního rozdělení má vektor  $(Z, \mathbf{X})'$  sdružené normální rozdělení a přesnost predikce se nezhorší, protože podmíněná střední hodnota je lineární funkcí podmínky.

### 3 Regresní model

Máme závislou proměnnou  $Z$  a nezávislé proměnné  $X_1, X_2, \dots, X_k$ , které jsou reprezentovány náhodným vektorem  $\mathbf{X} = (X_1, X_2, \dots, X_k)'$ .  $Z, X_1, X_2, \dots, X_k$  jsou náhodné veličiny na pravděpodobnostním prostoru  $(\Omega, \mathcal{A}, P)$ . Cílem regresního modelu je *predikovat*  $Z$  pomocí  $X_1, X_2, \dots, X_k$ .

#### 3.1 Regresní funkce

Z věty (2.1) plyne, že závislost  $Z$  na  $\mathbf{X}$  nejlépe vystihuje *regresní funkce*:

$$z = g(\mathbf{x}, \boldsymbol{\beta}) = E(Z|\mathbf{X} = \mathbf{x}),$$

kde  $\mathbf{x} = (x_1, \dots, x_k)$  je vektor nezávisle proměnných (hodnoty složek náhodného vektoru  $\mathbf{X}$ ),  $z$  je závisle proměnná (hodnota náhodné veličiny  $Z$ ),  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)'$  je vektor neznámých parametrů, které se nazývají regresní koeficienty a  $E(Z|\mathbf{X} = \mathbf{x})$  je podmíněná střední hodnota.

K tomu, abychom mohli odhadnout regresní koeficienty  $\beta_j$ , zavedeme nejprve veličinu, která se nazývá *reziduální součet čtverců* a je tvaru:

$$S = \sum_{i=1}^n [z_i - g(\mathbf{x}, \boldsymbol{\beta})]^2.$$

Tuto veličinu minimalizujeme a získáme tak odhady regresních koeficientů. Tato metoda se nazývá *metoda nejmenších čtverců*.

#### 3.2 Lineární regresní model

Máme náhodné veličiny  $Z, X_1, \dots, X_k$ . Pro náhodné veličiny  $X_1, \dots, X_k$  jsou dány jejich pevné hodnoty  $X_1 = x_1, \dots, X_k = x_k$ . Předpokládáme, že závislost  $Z$  na  $\mathbf{X}$  je lineární (vzhledem k regresním koeficientům). Uvažovaný model lze tedy zapsat ve tvaru:

$$Z = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \delta,$$

kde  $\delta$  je *náhodná chyba* (náhodná veličina).

Je dáno  $n$  pozorování veličin  $Z, X_1, \dots, X_k$ :

$$Z_1 = \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_k x_{1k} + \delta_1$$

$$Z_2 = \beta_1 x_{21} + \beta_2 x_{22} + \dots + \beta_k x_{2k} + \delta_2$$

⋮

$$Z_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \delta_i$$

⋮

$$Z_n = \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_k x_{nk} + \delta_n,$$

kde  $Z_i$  je  $i$ -té pozorování náhodné veličiny  $Z$ ,  $x_{ij}$  je  $i$ -té pozorování  $j$ -té složky náhodného vektoru  $\mathbf{X}$ ,  $\delta_i$  je náhodná chyba  $i$ -tého pozorování.

Zavedeme maticové označení a dostaneme  $\mathbf{Z} = (Z_1, Z_2, \dots, Z_n)'$  (tzv. vektor odezvy), vektor neznámých regresních koeficientů  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)'$ , vektor náhodných chyb  $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)$  a matici typu  $n \times k$  daných reálných čísel:

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ & & \dots & \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix}.$$

Pro náhodné chyby  $\boldsymbol{\delta}$  předpokládáme:

1.  $E(\boldsymbol{\delta}) = 0$ , tzn., že chyby  $\delta_i$  jsou nesystematické,
2.  $\text{var}(\boldsymbol{\delta}) = \sigma^2 \mathbf{I}$ ,
  - $D\delta_1 = D\delta_2 = \dots = D\delta_n = \sigma^2$ , tzn., že rozptyly chyb jsou homogenní a  $\sigma^2$  je rovněž neznámý parametr,
  - $\text{cov}(\delta_i, \delta_j) = 0, i \neq j$ , tzn., že chyby jsou nekorelované.

Dále požadujeme, aby  $\mathbf{X}$  měla lineárně nezávislé sloupce a předpokládáme  $k < n$ . Potom dostaneme  $h(\mathbf{X}) = k$  (model pak nazýváme *lineární regresní model plné hodnosti*) a matici  $\mathbf{X}'\mathbf{X}$  máme regulární.

Lineární regresní model v maticovém tvaru je:

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\delta}.$$

Když si uvědomíme, že vektor  $\mathbf{X}\boldsymbol{\beta}$  je nenáhodný, snadno vyvodíme důsledky:

- $E\mathbf{Z} = \mathbf{X}\boldsymbol{\beta}$
- $\text{var}\mathbf{Z} = \sigma^2 \mathbf{I}$ .

Regresní parametry  $\beta_1, \dots, \beta_k$  se odhadují metodou nejmenších čtverců, tzn. minimalizujeme již zmíněný reziduální součet čtverců jakožto funkci  $\boldsymbol{\beta}$ . Tyto odhady označíme  $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \dots, \hat{\beta}_k)$ . Řešíme tedy problém:

$$S = \sum_{i=1}^n \delta_i^2 = \boldsymbol{\delta}'\boldsymbol{\delta} = (\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta}) \rightarrow \min.$$

Úpravami dostaneme:

$$S = \mathbf{Z}'\mathbf{Z} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{Z} - \mathbf{Z}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{Z}'\mathbf{Z} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{Z} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta},$$

protože  $\boldsymbol{\beta}'\mathbf{X}'\mathbf{Z}$  je matice typu  $(1, 1)$ , neboli skalár a její transpozice  $(\boldsymbol{\beta}'\mathbf{X}'\mathbf{Z})^{-1} = \mathbf{Z}'\mathbf{X}\boldsymbol{\beta}$  je stejný skalár.

Hledaný odhad musí splňovat:

$$\frac{\partial S}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}} = -2\mathbf{X}'\mathbf{Z} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{0},$$

protože hledáme minimum funkce  $S$ , tedy  $S$  derivujeme a položíme derivaci rovnu nule.

Z posledního uvedeného vztahu dostaneme rovnice:

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{Z},$$

kteřé se nazývají *normální rovnice metody nejmenších čtverců* a můžeme je vynásobit výrazem  $(\mathbf{X}'\mathbf{X})^{-1}$ , protože  $(\mathbf{X}'\mathbf{X})$  je regulární a tedy existuje inverze. Dostaneme:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}.$$

Odhad  $i$ -tého pozorování náhodné veličiny  $Z$  je potom tvaru:

$$\hat{Z}_i = \sum_{j=1}^k \hat{\beta}_j x_{ij}, \quad i = 1, \dots, n.$$

Dostali jsme lineární regresní model plné hodnosti, kde odhad tvaru  $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}$  vyhovuje soustavě normálních rovnic  $\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{Z}$ .

Z výpočtu střední hodnoty  $E\hat{\boldsymbol{\beta}}$ :

$$E\hat{\boldsymbol{\beta}} = E((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E\mathbf{Z} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta}$$

vidíme, že  $\hat{\boldsymbol{\beta}}$  je *nestranným odhadem*  $\boldsymbol{\beta}$ . Ještě vyjádříme  $var(\hat{\boldsymbol{\beta}})$ :

$$\begin{aligned} var(\hat{\boldsymbol{\beta}}) &= var((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'var(\mathbf{Z})[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']' \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\sigma^2\mathbf{I})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}. \end{aligned}$$

Reziduální součet čtverců můžeme zapsat ve tvaru  $S(\hat{\boldsymbol{\beta}}) = (\mathbf{Z} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Z} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \sum_{i=1}^n (Z_i - \sum_{j=1}^k \hat{\beta}_j x_{ij})^2$  a  $S(\hat{\boldsymbol{\beta}})$  označíme  $S_e$ . Pro  $S_e$  platí:

- (i)  $S_e = \mathbf{Z}'(\mathbf{I} - \mathbf{H})\mathbf{Z} = \mathbf{Z}'\mathbf{M}\mathbf{Z}$ , kde  $\mathbf{H} = \mathbf{X}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}$  je matice projekce na  $\mathcal{M}(\mathbf{X})$ ,  $\mathbf{M} = \mathbf{I} - \mathbf{H}$  je matice projekce na  $\mathcal{M}(\mathbf{X})^\perp$ ;  $\mathcal{M}(\mathbf{X})$  je lineární prostor generovaný sloupci matice  $\mathbf{X}$ .
- (ii)  $\mathbf{H}$  a  $\mathbf{M}$  jsou idempotentní,  $h(\mathbf{H}) = k$ ,  $h(\mathbf{M}) = n - k$ .
- (iii)  $S_e = \mathbf{Z}'\mathbf{Z} - \hat{\mathbf{Z}}'\hat{\mathbf{Z}} = \sum_{i=1}^n Z_i^2 - \sum_{i=1}^n \hat{Z}_i^2$ .
- (iv)  $S_e = \mathbf{Z}'\mathbf{Z} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Z} = \mathbf{Z}'\mathbf{Z} - \hat{\boldsymbol{\beta}}'\mathbf{a}$ , kde  $\mathbf{a} = \mathbf{X}'\mathbf{Z}$  je pravá strana normálních rovnic.

Uvedené vztahy můžeme dokázat následovně:

- (ii)  $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ , pak  $\mathbf{H}^2 = (\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')'(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = \mathbf{H}$ , tedy  $\mathbf{H}$  je idempotentní,  $\mathbf{M}^2 = (\mathbf{I} - \mathbf{H})'(\mathbf{I} - \mathbf{H}) = \mathbf{I} - \mathbf{H} - \mathbf{H} + \mathbf{H}^2 = \mathbf{I} - \mathbf{H} = \mathbf{M}$ , tedy  $\mathbf{M}$  je idempotentní.

Protože  $\mathbf{H}$  je idempotentní, potom

$$h(\mathbf{H}) = Tr(\mathbf{H}) = Tr(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = Tr(\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}) = Tr(\mathbf{I}_k) = k \text{ a}$$

$$h(\mathbf{M}) = Tr(\mathbf{M}) = n - k.$$

- (i)  $(\mathbf{Z} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Z} - \mathbf{X}\hat{\boldsymbol{\beta}}) = (\mathbf{Z} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z})'(\mathbf{Z} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}) = (\mathbf{Z} - \mathbf{HZ})'(\mathbf{Z} - \mathbf{HZ}) = \mathbf{Z}'(\mathbf{I} - \mathbf{H})'(\mathbf{I} - \mathbf{H})\mathbf{Z} = \mathbf{Z}'\mathbf{M}^2\mathbf{Z} = \mathbf{Z}'\mathbf{MZ}$ , protože  $\mathbf{M}$  je idempotentní.
- (iii)  $S_e = \mathbf{Z}'(\mathbf{I} - \mathbf{H})\mathbf{Z} = \mathbf{Z}'\mathbf{Z} - \mathbf{Z}'\mathbf{HZ} = \mathbf{Z}'\mathbf{Z} - \mathbf{Z}'\mathbf{H}\mathbf{H}\mathbf{Z} = \mathbf{Z}'\mathbf{Z} - (\mathbf{HZ})'\mathbf{HZ} = \mathbf{Z}'\mathbf{Z} - \hat{\mathbf{Z}}'\hat{\mathbf{Z}}$ , protože  $\mathbf{HZ} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z} = \mathbf{X}\hat{\boldsymbol{\beta}} = \hat{\mathbf{Z}}$
- (iv)  $S_e = \mathbf{Z}'\mathbf{MZ} = \mathbf{Z}'\mathbf{Z} - \mathbf{Z}'\mathbf{HZ} = \mathbf{Z}'\mathbf{Z} - \mathbf{Z}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{Z}'\mathbf{Z} - \hat{\boldsymbol{\beta}}\mathbf{X}'\mathbf{Y}$ , tedy  $\mathbf{a} = \mathbf{X}'\mathbf{Z}$ .  $\square$

Nestranným odhadem  $\sigma^2$  je veličina  $s^2 = \frac{1}{n-k}S_e$ . Uvedenou vlastnost dokážeme výpočtem střední hodnoty:

$$E(S_e) = E(\mathbf{Z}'\mathbf{MZ}) = E(\text{Tr}(\mathbf{Z}'\mathbf{MZ})) = E[\text{Tr}(\mathbf{Z}' - \mathbf{EZ}')\mathbf{M}(\mathbf{Z} - \mathbf{EZ})],$$

protože

$$\mathbf{ME}(\mathbf{Z}) = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')(\mathbf{X}\boldsymbol{\beta}) = (\mathbf{X} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X})\boldsymbol{\beta} = \mathbf{0}\boldsymbol{\beta} = \mathbf{0}.$$

Potom

$$E(S_e) = E[\text{Tr}(\mathbf{Z} - \mathbf{EZ})(\mathbf{Z} - \mathbf{EZ})'\mathbf{M}] = \text{Tr}[E(\mathbf{Z} - \mathbf{EZ})(\mathbf{Z} - \mathbf{EZ})'\mathbf{M}],$$

protože

$$E(\boldsymbol{\delta}\boldsymbol{\delta}') = \text{var}(\boldsymbol{\delta}) = \text{var}(\mathbf{Z}),$$

dostaneme

$$E(S_e) = \text{Tr}(\text{var}(\mathbf{Z})\mathbf{M}) = \text{Tr}[(\sigma^2\mathbf{I})\mathbf{M}] = \sigma^2\text{Tr}(\mathbf{M}) = \sigma^2(n - k).$$

Z uvedeného vztahu vidíme, že

$$E(s^2) = E\left(\frac{1}{n-k}S_e\right) = \frac{1}{n-k}(n-k)\sigma^2 = \sigma^2.$$

Míru kvality lineární predikce vyjadřujeme veličinou  $R^2$ , která se nazývá *koefficient determinace* a je dána vztahem:

$$R^2 = 1 - \frac{S_e}{S_t},$$

kde  $S_e$  je reziduální součet čtverců a  $S_t = \sum_{i=1}^n (Z_i - \bar{Z})^2$  je tzv. *celkový (totální) součet čtverců*.

## 4 Prostorová data

### 4.1 Náhodný proces a jeho charakteristiky

Teorie náhodných procesů popisuje reálné přírodní, technické a jiné procesy, které mají náhodný charakter. Stochastický proces je zobecněním pojmu náhodné veličiny a při jeho definici vycházíme z pravděpodobnostního prostoru  $(\Omega, \mathcal{A}, P)$ , kde  $\Omega$  je daná. Při studování náhodných procesů a náhodných polí uvažujeme navíc neprázdnou množinu parametrů  $\mathcal{D}$ , její prvky budeme značit  $\mathbf{s}$ ,  $\mathbf{s} \in \mathcal{D}$  (resp. s indexy  $\mathbf{s}_1, \mathbf{s}_2, \dots$  apod).

Je-li  $Z(\mathbf{s})$  pro každé  $\mathbf{s} \in \mathcal{D}$  náhodnou veličinou definovanou na pravděpodobnostním prostoru  $(\Omega, \mathcal{A}, P)$ , potom systém  $\mathbf{Z} = \{Z(\mathbf{s}) | \mathbf{s} \in \mathcal{D}\}$  nazýváme *náhodným procesem*.

Je-li  $\mathcal{D} \subset \mathbf{R}$ , často interpretujeme  $\mathbf{s}$  jako čas a mluvíme o náhodném procesu. Je-li  $\mathcal{D} \subset \mathbf{R}^d$ , mluvíme o prostorovém náhodném procesu nebo o náhodném poli a  $\mathbf{s}$  může být interpretováno jako čas nebo jako souřadnice bodů v  $d$ -rozměrném prostoru.

Máme tedy náhodný proces  $\{Z(\mathbf{s})\}$  pozorovaný v jisté oblasti  $\mathcal{D} \subseteq \mathbf{R}^d$ , která má kladný  $d$ -rozměrný objem, proto je  $Z(\mathbf{s})$  pro každý bod  $\mathbf{s} \in \mathcal{D}$  náhodnou veličinou definovanou na pravděpodobnostním prostoru  $(\Omega, \mathcal{A}, P)$ . V této práci s chápeme jako prostorový index, který se spojitě mění v oblasti  $\mathcal{D}$ .

Jestliže pro každé  $\mathbf{s} \in \mathcal{D}$  existuje střední hodnota náhodné veličiny  $Z(\mathbf{s})$

$$E[Z(\mathbf{s})] = \mu(\mathbf{s}),$$

potom funkci  $\mu(\mathbf{s})$  nazveme *střední hodnotou* náhodného procesu  $\{Z(\mathbf{s})\}$ . O procesu řekneme, že je *cetrovaný*, je-li jeho střední hodnota nulová (tj. když  $E[Z(\mathbf{s})] = 0$  pro každé  $\mathbf{s} \in \mathcal{D}$ ).

Jestliže pro každé  $\mathbf{s} \in \mathcal{D}$  existuje rozptyl náhodné veličiny  $Z(\mathbf{s})$

$$D[Z(\mathbf{s})] = E[Z(\mathbf{s}) - \mu(\mathbf{s})]^2 = \sigma^2(\mathbf{s}),$$

potom funkci  $\sigma^2(\mathbf{s})$  nazveme *rozptylem* náhodného procesu  $\{Z(\mathbf{s})\}$ . Rozptyl náhodného procesu je stejně jako u náhodné veličiny nezáporná funkce a charakterizuje míru rozptýlení jednotlivých realizací kolem střední hodnoty. Z rozptylu dostaneme vztahem  $\sigma(\mathbf{s}) = \sqrt{\sigma^2(\mathbf{s})}$  *směrodatnou odchylku* náhodného procesu.

Uvedené charakteristiky nejsou pro popsání základních vlastností náhodného procesu dostačující. Zavedeme proto *kovarianční funkci* náhodného procesu

$$C(\mathbf{s}_1, \mathbf{s}_2) = \text{cov}(Z(\mathbf{s}_1), Z(\mathbf{s}_2)) = E\{[Z(\mathbf{s}_1) - \mu(\mathbf{s}_1)][Z(\mathbf{s}_2) - \mu(\mathbf{s}_2)]\}.$$

### 4.2 Stacionarita, kovariogram a variogram

#### 4.2.1 Stacionarita

Jestliže kovarianční funkce závisí pouze na rozdílu argumentů, tedy když

$$C(\mathbf{s}_1, \mathbf{s}_2) = C(\mathbf{s}_1 - \mathbf{s}_2) \quad \text{pro všechna } \mathbf{s}_1, \mathbf{s}_2 \in \mathcal{D}, \mathbf{s}_1 - \mathbf{s}_2 = \mathbf{h}, \quad (4.1)$$

potom o procesu říkáme, že je *kovariančně stacionární*. Kovarianční funkce je funkcí dvou proměnných, v našem případě  $\mathbf{s}_1, \mathbf{s}_2$  a kovarianční funkce závisující pouze na rozdílu argumentů je funkcí jedné proměnné, přesto se značí stejným písmenem  $C$ . Kovarianční funkce závisující pouze na rozdílu argumentů se často značí  $C(\mathbf{h})$  a to vyplývá z toho, že

$$C(\mathbf{s}_1 + \mathbf{h}, \mathbf{s}_1) = C(\mathbf{h}), \quad \mathbf{s}_1 \in \mathcal{D}.$$

O procesu můžeme říci, že je *izotropicky stacionární*, pokud jeho kovarianční funkce  $C$  závisí pouze na délce  $\mathbf{h}$ , tedy pokud

$$C(\mathbf{h}) = C(\|\mathbf{h}\|),$$

kde  $\|\mathbf{h}\|$  značí Euklidovskou normu vektoru  $\mathbf{h}$ .

Pokud je proces kovariančně stacionární a má konstantní střední hodnotu, tedy pokud

$$E(Z(\mathbf{s})) = \mu \quad \text{pro všechna } \mathbf{s} \in \mathcal{D},$$

mluvíme o *stacionaritě druhého řádu*.

#### 4.2.2 Variogram

Předpokládáme, že

$$\text{var}(Z(\mathbf{s}_1) - Z(\mathbf{s}_2)) = 2\gamma(\mathbf{s}_1 - \mathbf{s}_2) = 2\gamma(\mathbf{h}) \quad \text{pro všechna } \mathbf{s}_1, \mathbf{s}_2 \in \mathcal{D} \quad \text{a} \quad \mathbf{h} = \mathbf{s}_1 - \mathbf{s}_2. \quad (4.2)$$

Funkce  $2\gamma(\mathbf{h})$  se nazývá *variogram* (funkce  $\gamma(\mathbf{h})$  *semivariogram*).

Pokud pro proces platí, že má konstantní střední hodnotu

$$E(Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})) = 0$$

a existuje jeho variogram

$$\text{var}(Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})) = 2\gamma(\mathbf{h}) \quad \text{pro každé } \mathbf{s}, \mathbf{s} + \mathbf{h} \in \mathcal{D},$$

potom je proces *vnitřně stacionární*. Ze stacionarity druhého řádu plyne vnitřní stacionarita, opak ale neplatí.

Pokud variogram  $2\gamma(\mathbf{h})$  závisí pouze na délce  $\mathbf{h}$ , tedy pokud  $2\gamma(\mathbf{s}_1 - \mathbf{s}_2) = 2\gamma(\mathbf{h}) = 2\gamma(\|\mathbf{h}\|)$ , kde  $\|\mathbf{h}\|$  je Euklidovská norma, potom se variogram nazývá *izotropický*.

**Věta 4.1.** *Variogram  $2\gamma$  je **podmíněně negativně definitní** funkce, tedy splňuje podmínku*

$$\sum_{i=1}^m \sum_{j=1}^m a_i a_j 2\gamma(\mathbf{s}_i - \mathbf{s}_j) \leq 0 \quad (4.3)$$

pro každou konečnou množinu bodů  $\{\mathbf{s}_i : i = 1, \dots, m\}$  oblasti  $\mathcal{D}$  a pro libovolná reálná čísla  $\{a_i : i = 1, \dots, m\}$ , taková, že  $\sum_{i=1}^m a_i = 0$ .



Předchozí větu dokážeme následovně. Budeme předpokládat, že  $\{Z(\mathbf{s})\}$  je vnitřně stacionární proces,  $\mathbf{s}_i$  a  $a_i$  vyhovují podmínkám věty. Potom dostaneme

$$\left\{ \sum_{i=1}^m a_i Z(\mathbf{s}_i) \right\}^2 = -\frac{1}{2} \left\{ \sum_{i=1}^m \sum_{j=1}^m a_i a_j (Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2 \right\},$$

protože  $\sum_{i=1}^m a_i = 0$ . Aplikujeme-li na předchozí rovnost střední hodnotu

$$\sum_{i=1}^m \sum_{j=1}^m a_i a_j 2\gamma(\mathbf{s}_i - \mathbf{s}_j) = -2\text{var}\left(\sum_{i=1}^m a_i Z(\mathbf{s}_i)\right) \leq 0$$

a uvážíme-li, že rozptyl náhodné veličiny je nezáporný, dostaneme požadované tvrzení.  $\square$

### 4.2.3 Kovariogram a korelogram

Je-li proces kovariančně stacionární, potom funkce daná vztahem (4.1) se nazývá *kovariogram* a značí se  $C(\mathbf{s}_1 - \mathbf{s}_2)$  nebo  $C(\mathbf{h})$  (kovariogram tedy značíme stejným písmenem jako funkci kovarianční). Funkce

$$\rho(\mathbf{h}) \equiv \frac{C(\mathbf{h})}{C(\mathbf{0})}, \quad (4.4)$$

kde  $C(\mathbf{0}) > 0$ , se nazývá *korelogram* (a je také známa jako autokorelační funkce). Tato veličina bývá užívána pro analýzu časových řad, pro diagnostikování nestacionarity a určení typu stacionární závislosti, apod.

Kovariogram i korelogram jsou symetrické funkce, tedy  $C(\mathbf{h}) = C(-\mathbf{h})$ ,  $\rho(\mathbf{h}) = \rho(-\mathbf{h})$  a  $\rho(\mathbf{0}) = 1$  (což je známá vlastnost korelačního koeficientu). Kovariogram musí být pozitivně definitní, tedy musí splňovat

$$\sum_{i=1}^m \sum_{j=1}^m a_i a_j C(\mathbf{s}_i - \mathbf{s}_j) \geq 0 \quad (4.5)$$

pro nějaké pozorování v bodech  $\{\mathbf{s}_i : i = 1, \dots, m\}$  a reálná čísla  $\{a_i : i = 1, \dots, m\}$ .

Stejně jako variogram, můžeme kovariogram nazvat *izotropickým*, pokud závisí pouze na délce  $\mathbf{h}$ .

## 4.3 Lineární modely se statisticky závislou chybou

V kapitole 3 jsme představili model s homogenními rozptyly. Nyní tento model opustíme, protože u prostorových dat obvykle není  $\text{var}(\mathbf{Z}) = \sigma^2 \mathbf{I}$ . Budeme stále pracovat s daty  $Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)$ , která jsou modelována jako soubor náhodných veličin a model je stále tvaru:

$$\mathbf{Z}(\mathbf{s}) = \sum_{l=1}^k \beta_l x_l(\mathbf{s}) + \boldsymbol{\delta}(\mathbf{s}), \quad \mathbf{s} \in \mathcal{D} \subset \mathbf{R}^d, \quad (4.6)$$

kde  $\{x_l : l = 1, \dots, k\}$  je soubor  $k$  nenáhodných proměnných, které mohou (ale nemusí) záviset na umístění v prostoru a  $\boldsymbol{\delta}$  je chybový proces, který má nulovou střední hodnotu, konečný rozptyl a může být (ale nemusí) prostorově korelován.

Pozorovaná data označujeme jako  $\mathbf{Z} \equiv (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))'$  a model daný vztahem (4.6) lze maticově zapsat:

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\delta}, \quad (4.7)$$

kde  $\mathbf{X}$  je matice typu  $(n, k)$ , jejíž  $(i, j)$ -tý prvek je  $x_j(\mathbf{s}_i)$ ,  $\boldsymbol{\beta} \equiv (\beta_1, \dots, \beta_k)'$  a  $\boldsymbol{\delta} \equiv (\delta(\mathbf{s}_1), \dots, \delta(\mathbf{s}_n))'$ .

Dále budeme předpokládat, že

$$E(\mathbf{Z}) = \mathbf{X}\boldsymbol{\beta} \quad (4.8)$$

a že existuje varianční matice procesu  $\mathbf{Z}$  a označíme ji

$$\text{var}(\mathbf{Z}) = \boldsymbol{\Sigma}, \quad (4.9)$$

varianční matice nezávisí na  $\boldsymbol{\beta}$  a  $k < n$ .

Budeme hledat lineární odhad  $\boldsymbol{\beta}$ . Řekneme, že  $\hat{\boldsymbol{\beta}}$  je lineárním odhadem vektoru  $\boldsymbol{\beta}$ , jestliže existuje taková matice  $\mathbf{U}_{n \times k}$ , že  $\hat{\boldsymbol{\beta}} = \mathbf{U}'\mathbf{Z}$ .

**Věta 4.2.** *Lineární odhad je nestranný tehdy a jen tehdy, platí-li*

$$\mathbf{U}\mathbf{X} = \mathbf{I}. \quad (4.10)$$

*Důkaz:* Odhad  $\hat{\boldsymbol{\beta}}$  je nestranný tehdy a jen tehdy, je-li  $\boldsymbol{\beta} = E\hat{\boldsymbol{\beta}}$  pro libovolný vektor  $\boldsymbol{\beta}$ . Avšak rovnost  $\boldsymbol{\beta} = E(\mathbf{U}\mathbf{Z}) = \mathbf{U}E(\mathbf{Z}) = \mathbf{U}\mathbf{X}\boldsymbol{\beta}$  platí pro libovolný vektor  $\boldsymbol{\beta}$  právě tehdy, platí-li vztah (4.10).  $\square$

Z (4.10) neplyne  $\mathbf{U} = \mathbf{X}^{-1}$ , protože  $\mathbf{X}$  za předpokladu  $n < k$  není čtvercová.  $\hat{\boldsymbol{\beta}}$  nazveme nejlepším nestranným lineárním odhadem (NNLO) vektoru  $\boldsymbol{\beta}$ , platí-li:

- $\hat{\boldsymbol{\beta}}$  je lineární odhad parametru  $\boldsymbol{\beta}$ ;
- $\hat{\boldsymbol{\beta}}$  je nestranný odhad parametru  $\boldsymbol{\beta}$ ;
- je-li  $\hat{\boldsymbol{\beta}}^*$  jiný nestranný lineární odhad  $\boldsymbol{\beta}$ , pak platí  $\text{var}\hat{\boldsymbol{\beta}}^* - \text{var}\hat{\boldsymbol{\beta}} \geq 0$  (tzn. rozdíl uvedených dvou variančních matic musí být matice pozitivně semidefinitní).

### 4.3.1 Model s plnou hodností

**Věta 4.3.** *(Gaussova-Markovova) Je dán model  $\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\delta}$ . Nechť  $h(\mathbf{X}) = k$  a nechť  $\boldsymbol{\Sigma} = \text{var}(\mathbf{Z})$  je regulární. Potom NNLO pro  $\boldsymbol{\beta}$  je roven*

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Z} \quad (4.11)$$

a má varianční matici

$$\text{var}\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}. \quad (4.12)$$

*Důkaz:* Z předpokladů plyne, že  $\boldsymbol{\Sigma}$  je pozitivně definitní. Proto existuje matice  $\boldsymbol{\Sigma}^{\frac{1}{2}}$  typu  $n \times n$  hodnosti  $k$  tak, že  $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{\frac{1}{2}}(\boldsymbol{\Sigma}^{\frac{1}{2}})'$ , kde  $\boldsymbol{\Sigma}^{\frac{1}{2}}$  je regulární. Proto existuje  $(\boldsymbol{\Sigma}^{\frac{1}{2}})^{-1}$  a tuto matici označíme  $\boldsymbol{\Sigma}^{-\frac{1}{2}}$ . Potom je  $h(\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{X}) = k$  a platí, že  $h(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}) = h(\mathbf{X}'\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{X}) = k$ , takže  $\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}$  je regulární. Odhad (4.11) je lineární. Je také nestranný, protože

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}E\mathbf{Z} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta}.$$

Pokud je  $\hat{\beta}^*$  nějaký jiný lineární nestranný odhad. Potom je

$$\text{var}\hat{\beta} = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}, \quad \text{var}\hat{\beta}^* = \mathbf{U}\Sigma^{-1}\mathbf{U}', \quad \text{cov}(\hat{\beta}, \hat{\beta}^*) = \text{var}\hat{\beta}.$$

Z toho plyne, že

$$0 \leq \text{var}(\hat{\beta}^* - \hat{\beta}) = \text{var}\hat{\beta}^* - \text{cov}(\hat{\beta}^*, \hat{\beta}) - \text{cov}(\hat{\beta}, \hat{\beta}^*) + \text{var}\hat{\beta} = \text{var}\hat{\beta}^* - \text{var}\hat{\beta}. \quad \square$$

Odhad  $\hat{\beta}$  je odhad zobecněnou metodou nejmenších čtverců (generalized-least-squares estimator), proto se tento odhad označuje  $\hat{\beta}_{GLS}$ .

## 4.4 Příklady prostorových dat

Součástí této práce je programová implementace metody představené v kapitole 5, její demonstrace bude předvedena na 3 příkladech. Příslušná data a jejich grafické znázornění budou uvedena v následujícím odstavci.

### 4.4.1 Simulovaná data

#### *Regresní rovina*

Prvním příkladem je regresní funkce  $\mathbf{Z} = \mathbf{x} + 2\mathbf{y} + \boldsymbol{\delta}$ , napočítaná pro  $\mathbf{x} = \mathbf{y} = 1 : 1 : 15$  (hodnoty  $x$  a  $y$  jsou od 1 s krokem 1 do 15). Získané hodnoty  $\mathbf{Z}$  jsou navíc zatíženy náhodnou chybou  $\boldsymbol{\delta}$ , což je normálně rozdělený chybový proces s nulovou střední hodnotou a se směrodatnou odchylkou 0.2. V Matlabu takovýto proces zapíšeme jako  $\boldsymbol{\delta} = \text{normrnd}(0, 0.2, n_x, n_y)$ , kde  $n_x$  je velikost vektoru  $\mathbf{x}$  a  $n_y$  je velikost vektoru  $\mathbf{y}$ .

Tabulka hodnot  $\mathbf{Z}$  pro případ roviny je na Obr.(1) a vykreslené hodnoty vidíme na Obr.(3).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	3.1188	5.0131	7.2617	9.2597	11.2455	13.131	15.2353	16.6178	18.9051	21.2234	23.1454	25.172	26.8171	28.5841	31.185
2	4.2855	6.3195	8.2322	9.763	11.8951	14.163	16.1673	17.9943	19.8345	21.4271	23.9803	25.8563	27.8727	30.0935	31.6219
3	5.2402	7.0142	8.6596	11.1455	13.1128	15.0545	16.9941	18.9253	20.9411	23.4414	24.7511	26.9449	28.932	31.4339	32.7237
4	6.1046	7.9968	9.8767	11.8588	13.7323	15.8724	17.9219	20.0817	22.3249	23.954	25.6633	28.1764	30.074	31.8405	34.004
5	6.9269	9.0454	10.9333	12.9764	15.0348	17.2939	19.1521	20.6237	22.743	24.4477	27.0063	29.3987	31.2127	33.0556	35.1424
6	8.0775	10.2787	12.1103	13.8993	15.8427	17.9062	20.0236	22.2008	23.8509	25.7421	28.0323	29.7354	31.7659	34.1521	36.3468
7	8.914	11.167	12.9297	14.9546	17.179	18.8308	21.2872	22.9809	24.9714	26.6533	28.9522	30.8435	32.6485	35.2924	37.1278
8	10.1841	11.8022	14.1441	16.1595	17.8172	19.6641	22.1799	23.0089	26.1314	28.0054	30.1254	31.8694	33.8912	35.9247	37.7634
9	10.8979	13.0201	14.749	17.1094	18.8863	20.9889	23.24	25.0639	26.7361	28.8367	31.1263	32.9569	35.0674	37.2648	39.4143
10	11.9504	14.1952	15.9756	17.9801	20.1249	21.9729	24.1307	25.825	28.0746	29.8598	32.0556	34.384	36.0501	38.0805	40.1974
11	13.3582	15.1266	16.9983	19.2873	20.9087	22.5239	24.6078	26.6673	28.9089	31.1423	33.5115	34.9334	37.0144	39.3313	41.1706
12	13.7233	15.9282	17.9759	19.964	22.0946	23.6149	25.7514	27.8604	29.8129	31.9709	34.1801	35.952	37.6676	39.9326	42.1293
13	15.3165	16.9086	18.8243	21.2169	23.0333	25.2516	26.7388	29.0299	31.0118	33.0096	35.4016	36.9305	38.6575	41.2896	42.8005
14	15.9166	18.0255	19.8462	21.6515	24.2241	26.0692	27.95	29.9393	32.152	33.823	36.1007	38.2796	40.0615	41.9424	44.1015
15	16.8332	19.0737	21.3771	23.0455	25.3513	27.1926	28.8317	31.0735	32.8843	35.1087	36.9081	38.9263	40.6444	42.9259	44.9938

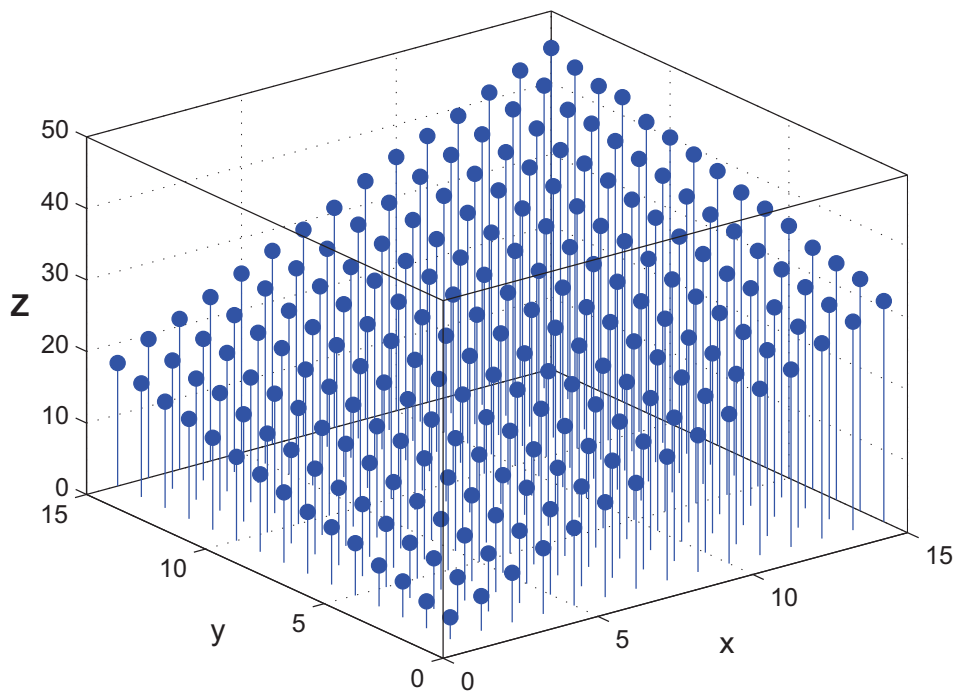
Obrázek 1: Data  $\mathbf{Z}$  pro případ regresní roviny

#### *Obecná lineární regresní plocha*

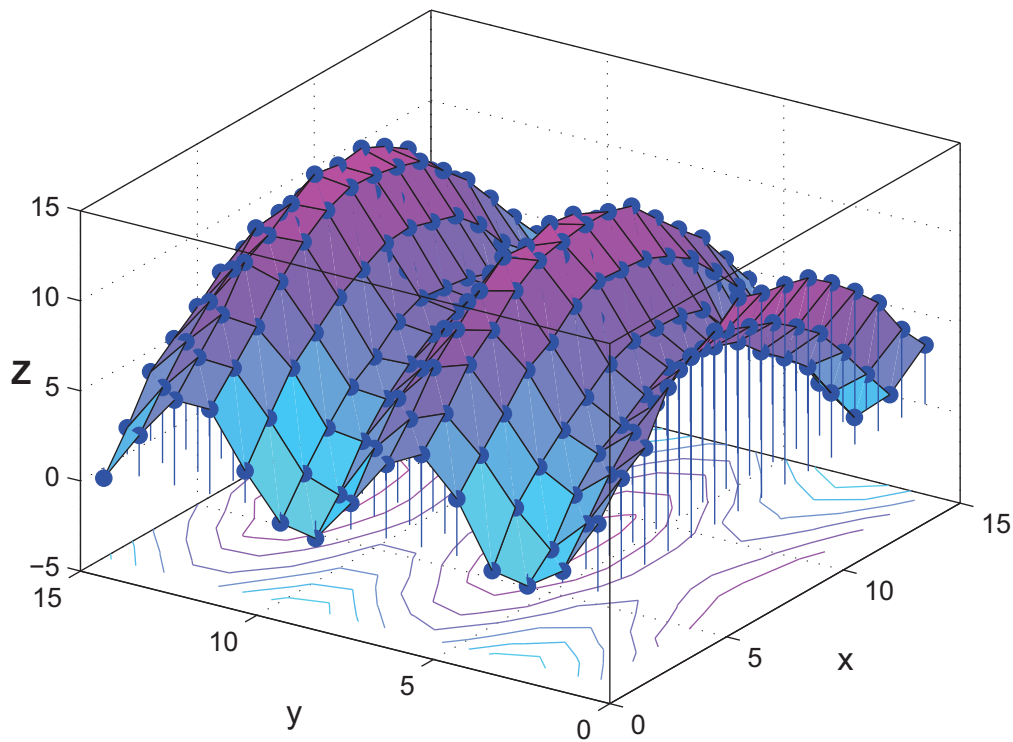
Druhým řešeným příkladem je regresní funkce  $\mathbf{Z} = 3 \cos(\mathbf{x}) + 10 \sin(0.2\mathbf{y}) + \boldsymbol{\delta}$ , kde je opět  $\mathbf{x} = \mathbf{y} = 1 : 1 : 15$  a data jsou opět zatížena chybou  $\boldsymbol{\delta} = \text{normrnd}(0, 0.2, n_x, n_y)$ . Tabulka hodnot  $\mathbf{Z}$  pro případ zvlněné plochy je na Obr.(2). Napočítané hodnoty jsem pro lepší viditelnost tvaru dat proložila plochou - viz. Obr. (4).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	3.5981	5.7564	7.4304	8.7245	9.768	10.7221	11.5474	11.6294	11.6626	10.8393	9.9914	8.2923	7.0208	4.8564	3.2229
2	1.0719	2.4768	4.4237	5.9689	7.0628	8.3408	9.1464	8.9564	8.9579	7.8619	7.156	5.8512	3.8817	2.0256	0.0291
3	-1.1339	1.1136	2.4963	4.5264	5.3156	6.2736	7.0173	6.9436	6.9962	6.3737	4.8751	3.7215	1.967	0.1702	-1.4383
4	0.116	1.804	3.588	5.2413	6.6765	7.2713	7.3809	7.9482	7.6535	7.432	5.9813	5.0848	3.3894	1.4031	-0.3929
5	2.9149	4.5976	6.65	8.1494	9.3999	9.8826	10.7194	10.6796	10.3807	10.1738	8.843	7.4931	6.1973	4.2298	2.4489
6	4.7534	6.5711	8.5183	10.1963	11.477	12.0531	12.6541	12.5921	12.8372	11.8174	10.924	9.6483	7.9744	6.0926	4.1406
7	4.4259	6.228	7.8557	9.3677	10.8452	11.8429	12.1872	12.1335	12.2489	11.5558	10.2917	8.7077	7.3763	5.9048	3.5018
8	1.822	3.3807	5.3534	6.6764	7.6255	9.0121	9.5192	9.3225	9.1017	8.5005	7.9649	6.3522	4.8052	2.9507	0.9025
9	-0.9462	1.5749	2.8558	4.7025	5.4648	6.6376	6.943	7.1131	6.9801	6.2755	5.1192	4.1668	2.9694	0.8597	-1.1233
10	-0.5092	1.1612	2.9456	4.5686	5.8997	6.7991	7.3458	7.4125	7.4495	6.5897	5.5722	4.4687	2.7436	1.056	-0.9912
11	1.8052	3.9931	5.4363	7.0222	8.6088	9.4259	9.8978	9.9126	9.6259	9.0989	8.1367	6.5514	5.0765	3.3436	1.6973
12	4.4687	6.5085	8.4038	9.8709	11.0104	12.2471	12.3685	12.6806	12.2828	11.4903	10.712	9.6733	7.8138	5.7813	3.8003
13	4.844	6.3605	8.2133	10.1975	11.388	12.0073	12.4754	12.6294	12.6637	11.8094	10.7845	9.4672	8.1123	5.6528	4.4416
14	2.5245	4.5028	6.0234	7.374	8.6421	9.4847	10.3942	10.3924	10.5277	9.4529	8.3441	7.238	5.2536	4.016	1.7679
15	-0.2352	1.8734	3.3806	4.9446	6.0357	6.8461	7.4452	7.9086	7.6605	6.6552	5.7818	4.2405	3.2166	0.9849	-0.8958

Obrázek 2: Data  $\mathbf{Z}$  pro případ obecné lineární regresní plochy



Obrázek 3: Vykreslené hodnoty procesu  $Z$  pro případ regresní roviny

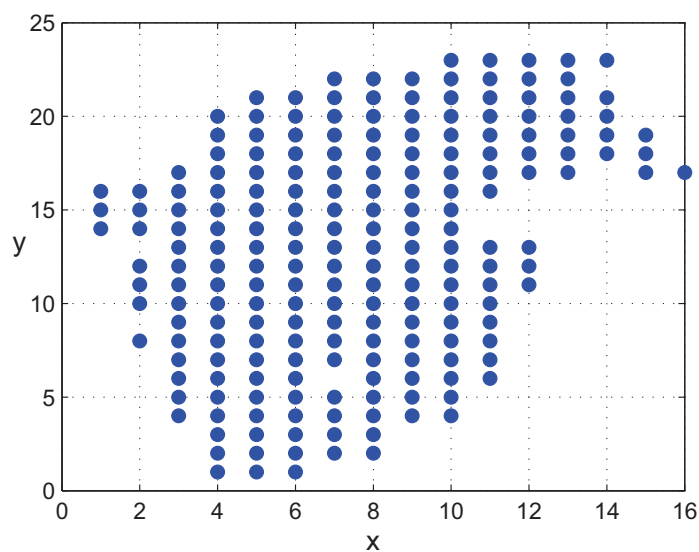


Obrázek 4: Vykreslené hodnoty procesu  $Z$  pro případ obecné lineární regresní plochy

#### 4.4.2 Reálná data

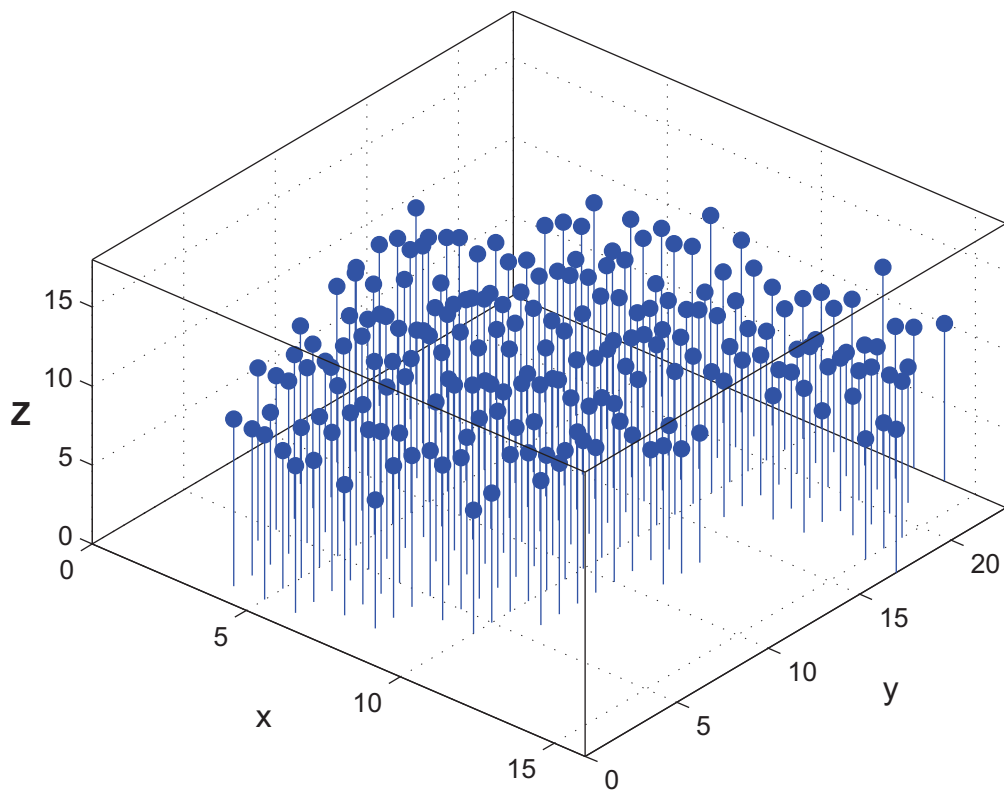
##### *Uhelný popel v Pensylvánii (Coal-Ash Data)*

Tento příklad je převzatý z [1]. Jde o hodnoty uhelného popelu naměřené v Pensylvánii. Vektor  $\mathbf{x} = 1 : 1 : 16$  a  $\mathbf{y} = 1 : 1 : 23$  s tím rozdílem, že tentokrát nemáme hodnotu  $Z$  naměřenou ve všech bodech  $\mathbf{s} = [x, y]$ . Body, ve kterých je známá hodnota  $Z$  jsou znázorněny na Obr.(5). Hodnoty procesu  $\mathbf{Z}$  jsou dány na Obr.(6)(tabulka vyjadřuje procentuální hodnoty uhelného popelu v bodech  $\mathbf{s} = [x, y]$ ) a prostorové vykreslení naměřených hodnot je na Obr.(7)



									8.59	9.00	11.86	8.91	9.99						
23						11.62	10.91	8.76	8.89	9.10	7.62	9.65							
22					10.39	10.65	10.36	9.58	10.66	8.92	7.80	7.84	9.03	8.60					
21																			
20			9.79	9.06	10.70	11.21	8.98	9.27	8.19	7.88	7.61	8.20	8.77						
19			10.74	12.80	10.03	9.36	8.57	9.01	9.04	7.28	9.58	9.69	9.96	9.91					
18			11.21	9.89	10.34	8.20	9.82	10.06	8.58	8.89	8.64	7.04	8.81	7.95					
17			9.97	9.70	9.84	10.29	9.84	10.01	9.01	7.68	9.25	7.83	9.14	7.63	9.07				
16	11.17	10.14	9.93	10.27	10.21	11.09	10.63	8.82	10.18	9.34	8.61								
15	9.92	10.82	11.65	8.96	9.88	8.90	10.18	9.34	10.56	9.06									
14	10.21	10.73	9.46	9.35	9.78	10.38	9.79	8.91	9.22	11.43									
13			12.50	9.63	10.82	10.12	9.40	9.48	10.99	9.92	7.85	8.21							
12			9.92	11.05	10.11	11.46	10.41	8.45	8.90	8.07	7.96	7.00	7.90						
11			11.31	9.41	9.37	11.21	9.93	10.70	9.27	9.28	10.13	8.61	8.78						
10			11.15	9.91	10.17	10.55	11.61	9.16	10.04	11.19	8.10	11.30							
9			10.82	11.75	9.78	11.00	9.79	10.19	9.15	8.15	9.20								
8		10.01	8.23	11.04	10.28	13.07	10.47	11.58	9.46	8.54	10.87								
7			10.39	11.11	10.96	10.83	10.09	8.69	11.17	9.39	9.56								
6			10.41	10.82	17.61	10.87		13.06	11.41	9.96	9.15								
5			9.76	11.10	10.80	8.86	9.48	9.22	9.61	8.20									
4			10.93	10.94	9.53	10.61	10.27	9.59	9.82	7.81									
3				9.64	9.52	10.06	12.65	9.63											
2				9.29	8.75	8.96	8.27	8.14											
1				10.59	10.43	9.32													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16		

Obrázek 6: Data  $\mathbf{Z}$  pro příklad o uhelném popelu (viz[1])



Obrázek 7: Vykreslené hodnoty procesu  $Z$  pro data o uhelném popelu

Data uvedená v popsaných příkladech budou analyzována v kapitolách 8, 9 a 10.

## 5 Kriging metoda

### 5.1 Univerzální kriging model

Nejprve popíšeme model, který se nazývá *univerzální kriging*. Budeme předpokládat, že jsou dána pozorování uspořádaná do vektoru

$$\mathbf{Z} \equiv (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))', \quad (5.1)$$

pozorování jsou provedeny ve známých bodech

$$\{\mathbf{s}_1, \dots, \mathbf{s}_n\}, \mathbf{s} \in \mathcal{D} \subset \mathbf{R}^d \quad (5.2)$$

a chceme z těchto dat predikovat hodnotu  $Z(\mathbf{s}_0)$ . U tohoto modelu předpokládáme proces  $\mathbf{Z}(\mathbf{s})$  ve tvaru:

$$\mathbf{Z}(\mathbf{s}) = \boldsymbol{\mu}(\mathbf{s}) + \boldsymbol{\delta}(\mathbf{s}), \mathbf{s} \in \mathcal{D}, \quad (5.3)$$

kde  $\boldsymbol{\mu}(\mathbf{s})$  je střední hodnotou  $\mathbf{Z}(\mathbf{s})$  a  $\boldsymbol{\delta}(\mathbf{s})$  je náhodný korelovaný chybový proces. Tedy  $\boldsymbol{\mu}(\mathbf{s})$  je deterministická složka procesu  $\mathbf{Z}$ , která popisuje tzv. *large-scale variability* procesu a  $\boldsymbol{\delta}(\mathbf{s})$  popisuje směs tzv. *smooth-scale variability*, *mikro-scale variability* a *chyb měření*. O náhodném procesu  $\boldsymbol{\delta}(\mathbf{s})$  budeme dále předpokládat, že má nulovou střední hodnotu ( $E(\boldsymbol{\delta}(\mathbf{s})) = 0$ ), je vnitřně stacionární a má variogram  $2\gamma$ .

O střední hodnotě  $\boldsymbol{\mu}(\mathbf{s})$  procesu  $\mathbf{Z}$  dále předpokládáme, že je lineární kombinací známých funkcí  $f_0(\mathbf{s}), \dots, f_p(\mathbf{s}), \mathbf{s} \in \mathcal{D}$ . Ačkoliv každá  $f_j(\mathbf{s})$  je psána jako funkce polohy  $\mathbf{s}$ , uvažujeme i konstantní funkce.

Tedy model předpokládáme ve tvaru

$$\mathbf{Z}(\mathbf{s}) = \sum_{j=1}^{p+1} f_{j-1}(\mathbf{s})\beta_{j-1} + \boldsymbol{\delta}(\mathbf{s}), \quad \mathbf{s} \in \mathcal{D}, \quad (5.4)$$

kde  $\boldsymbol{\beta} \equiv (\beta_0, \dots, \beta_p)' \in \mathbf{R}^{p+1}$  je neznámý vektor parametrů, který chceme odhadnout a  $\boldsymbol{\delta}$  je vnitřně stacionární náhodný proces s nulovou střední hodnotou a variogramem  $2\gamma$ . Označíme-li

$$\mathbf{X} = \begin{pmatrix} f_0(\mathbf{s}_1) & \dots & f_p(\mathbf{s}_1) \\ \vdots & \ddots & \vdots \\ f_0(\mathbf{s}_n) & \dots & f_p(\mathbf{s}_n) \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix}, \boldsymbol{\delta} = \begin{pmatrix} \delta(\mathbf{s}_1) \\ \vdots \\ \delta(\mathbf{s}_n) \end{pmatrix}$$

můžeme rovnici (5.4) přepsat do maticového tvaru

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\delta}, \quad (5.5)$$

matice  $\mathbf{X}$  je tedy typu  $(n, p+1)$ , jejíž  $(i, j)$ -tý prvek je  $f_{j-1}(\mathbf{s}_i)$ .

Pomocí metody universální kriging chceme nalézt nejlepší lineární prediktor

$$Z^*(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i) \quad (5.6)$$



procesu  $\mathbf{Z}(\mathbf{s})$  v bodě  $\mathbf{s} = \mathbf{s}_0$ . Hledáme tedy reálná čísla  $\lambda_1, \dots, \lambda_n$  taková, aby střední kvadratická chyba predikce  $Z^*$  v bodě  $\mathbf{s}_0$

$$MSE(\mathbf{s}_0) = E(Z^*(\mathbf{s}_0) - Z(\mathbf{s}_0))^2 = E\left(Z(\mathbf{s}_0) - \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i)\right)^2 \quad (5.7)$$

byla minimální. Navíc požadujeme, aby tento odhad byl nestranný, tedy minimalizujeme (5.7) za podmínky

$$EZ^*(\mathbf{s}_0) = EZ(\mathbf{s}_0). \quad (5.8)$$

S ohledem na předchozí uvedené vztahy lze podmínku (5.8) přepsat do tvaru:

$$\mu(\mathbf{s}_0) = \sum_{j=1}^{p+1} \beta_{j-1} f_{j-1}(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i \mu(\mathbf{s}_i) = \sum_{i=1}^n \lambda_i \sum_{j=1}^{p+1} \beta_{j-1} f_{j-1}(\mathbf{s}_i). \quad (5.9)$$

Označíme-li

$$\boldsymbol{\lambda} = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}, \mathbf{x} = \begin{pmatrix} f_0(\mathbf{s}_0) \\ \vdots \\ f_p(\mathbf{s}_0) \end{pmatrix},$$

můžeme (5.9) přepsat do tvaru

$$\mathbf{x}'\boldsymbol{\beta} = \boldsymbol{\lambda}'\mathbf{X}\boldsymbol{\beta} \quad (5.10)$$

Z podmínky nestrannosti (5.8) prediktoru  $Z^*(\mathbf{s})$  pak plyne, že (5.10) musí platit pro každé  $\boldsymbol{\beta} \in \mathbf{R}^{p+1}$  a tedy ze (5.10) plyne  $\mathbf{x} = \mathbf{X}'\boldsymbol{\lambda}$ . Když množina  $\mathcal{U} = \{\boldsymbol{\lambda} \in \mathbf{R}^n : \mathbf{x} = \mathbf{X}'\boldsymbol{\lambda}\}$  je neprázdná, potom univerzální kriging estimátor  $\mathbf{Z}^*$  může být nalezen minimalizací střední čtvercové chyby (5.7) za podmínky, že  $\boldsymbol{\lambda} \in \mathcal{U}$ .

## 5.2 Kriging odhady

Chceme predikovat hodnotu procesu  $Z$  v bodě  $\mathbf{s}_0$  a pomocí zavedeného označení dostaneme pro  $Z(\mathbf{s}_0)$  vztah

$$Z(\mathbf{s}_0) = \mu(\mathbf{s}_0) + \delta(\mathbf{s}_0) = \mathbf{x}'\boldsymbol{\beta} + \delta(\mathbf{s}_0). \quad (5.11)$$

Nejlepší lineární nestranný prediktor, který jsme označili  $Z^*$ , minimalizuje (5.7) pomocí  $\lambda_1, \dots, \lambda_n$  za podmínky  $\boldsymbol{\lambda}'\mathbf{X} = \mathbf{x}'$ . Prívlastek univerzální užil poprvé Matheron(1969), aby zdůraznil nestrannost prediktoru, ačkoliv  $\mu(\mathbf{s})$  je neznámou lineární kombinací známých funkcí. Po zavedení Lagrangeových multiplikátorů je minimalizována funkce

$$E\left(Z(\mathbf{s}_0) - \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i)\right)^2 - 2 \sum_{j=1}^{p+1} m_{j-1} \left\{ \sum_{i=1}^n \lambda_i f_{j-1}(\mathbf{s}_i) - f_{j-1}(\mathbf{s}_0) \right\} \quad (5.12)$$

s ohledem na  $\lambda_1, \dots, \lambda_n$  a  $m_0, \dots, m_p$ , kde  $m_0, \dots, m_p$  jsou Lagrangeovy multiplikátory, které zaručí splnění podmínky  $\boldsymbol{\lambda}'\mathbf{X} = \mathbf{x}'$ . Nyní budeme předpokládat, že  $f_0(\mathbf{s}) \equiv 1$ , což zaručí, že  $\sum_{i=1}^n \lambda_i = 1$  (jedna z podmínek nestrannosti).

Z rovnic (5.5), (5.11), (5.6) a užitím vztahu  $\sum_{i=1}^n \lambda_i = 1$  dostaneme

$$\begin{aligned} \left( Z(\mathbf{s}_0) - \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i) \right)^2 &= \left( \mathbf{x}'\boldsymbol{\beta} + \delta(\mathbf{s}_0) - \boldsymbol{\lambda}'\mathbf{X}\boldsymbol{\beta} - \sum_{i=1}^n \lambda_i \delta(\mathbf{s}_i) \right)^2 \\ &= \left( \delta(\mathbf{s}_0) - \sum_{i=1}^n \lambda_i \delta(\mathbf{s}_i) \right)^2 \\ &= - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \frac{(\delta(\mathbf{s}_i) - \delta(\mathbf{s}_j))^2}{2} + 2 \sum_{i=1}^n \lambda_i \frac{(\delta(\mathbf{s}_0) - \delta(\mathbf{s}_i))^2}{2}. \end{aligned} \quad (5.13)$$

Dále předpokládáme, že

$$2\gamma(\mathbf{h}) = \text{var}(Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})). \quad (5.14)$$

Funkci (5.12), kterou máme minimalizovat, dostaneme ve tvaru

$$- \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(\mathbf{s}_i - \mathbf{s}_j) + 2 \sum_{i=1}^n \lambda_i \gamma(\mathbf{s}_0 - \mathbf{s}_i) - 2 \sum_{j=1}^{p+1} m_{j-1} \left\{ \sum_{i=1}^n \lambda_i f_{j-1}(\mathbf{s}_i) - f_{j-1}(\mathbf{s}_0) \right\}. \quad (5.15)$$

Při odvození těchto rovnic jsme předpokládali, že  $f_0(\mathbf{s}) \equiv 1$ , v případě, že žádná z  $\{f_{j-1}(\mathbf{s}) : j = 1, \dots, p+1\}$  není identicky rovna 1, potom minimalizování (5.7) v závislosti na  $\boldsymbol{\lambda}'\mathbf{X} = \mathbf{x}'$  nemusí být dosaženo minimalizováním (5.15).

### 5.2.1 Rovnice univerzálního krigingu

Vztah (5.15) derivujeme podle proměnných  $\lambda_1, \dots, \lambda_n, m_0, \dots, m_p$  a výsledek položíme roven 0, potom optimální váhy získáme z rovnice

$$\boldsymbol{\lambda}_U = \boldsymbol{\Gamma}_U^{-1} \boldsymbol{\gamma}_U, \quad (5.16)$$

kde

$$\boldsymbol{\lambda}_U \equiv (\lambda_1, \dots, \lambda_n, m_0, \dots, m_p)', \quad (5.17)$$

$$\boldsymbol{\gamma}_U \equiv (\gamma(\mathbf{s}_0 - \mathbf{s}_1), \dots, \gamma(\mathbf{s}_0 - \mathbf{s}_n), 1, f_1(\mathbf{s}_0), \dots, f_p(\mathbf{s}_0))' \quad (5.18)$$

a  $\boldsymbol{\Gamma}_U$  je symetrická matice typu  $(n + p + 1) \times (n + p + 1)$

$$\boldsymbol{\Gamma}_U = \begin{cases} \gamma(\mathbf{s}_i - \mathbf{s}_j), & i = 1, \dots, n; j = 1, \dots, n, \\ f_{j-1-n}(\mathbf{s}_i), & i = 1, \dots, n; j = n + 1, \dots, n + p + 1, \\ 0, & i = n + 1, \dots, n + p + 1; j = n + 1, \dots, n + p + 1, \end{cases} \quad (5.19)$$

kde  $f_0 \equiv 1$ . A tedy koeficienty  $\boldsymbol{\lambda}$  jsou dány vztahem

$$\boldsymbol{\lambda}' = \{\boldsymbol{\gamma} + \mathbf{X}(\mathbf{X}'\boldsymbol{\Gamma}^{-1}\mathbf{X})^{-1}(\mathbf{x} - \mathbf{X}'\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma})\}'\boldsymbol{\Gamma}^{-1} \quad (5.20)$$

a

$$\mathbf{m}' = -(\mathbf{x} - \mathbf{X}'\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma})'(\mathbf{X}'\boldsymbol{\Gamma}^{-1}\mathbf{X})^{-1},$$

kde  $\boldsymbol{\gamma} \equiv (\gamma(\mathbf{s}_0 - \mathbf{s}_1), \dots, \gamma(\mathbf{s}_0 - \mathbf{s}_n))'$  a  $\boldsymbol{\Gamma}$  je matice typu  $n \times n$ , jejíž  $(i, j)$ -tý prvek je  $\gamma(\mathbf{s}_i - \mathbf{s}_j)$ , tedy

$$\boldsymbol{\Gamma} \equiv \begin{pmatrix} \gamma(\mathbf{0}) & \gamma(\mathbf{s}_1 - \mathbf{s}_2) & \dots & \gamma(\mathbf{s}_1 - \mathbf{s}_n) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(\mathbf{s}_n - \mathbf{s}_1) & \gamma(\mathbf{s}_n - \mathbf{s}_2) & \dots & \gamma(\mathbf{0}) \end{pmatrix}$$

Dosud jsme předpokládali, že proces  $\mathbf{Z}$  je vnitřně stacionární. Nyní předpoklad vnitřní stacionarity zesílíme a odhady upravíme pro proces, který je stacionární druhého řádu. Pro takový proces je definován jeho kovariogram  $C(\mathbf{h}) \equiv \text{cov}(Z(\mathbf{s} + \mathbf{h}), Z(\mathbf{s}))$ . Rovnice analogická rovnici (5.16) je rovnice

$$(\lambda_1, \dots, \lambda_n, -m_0, \dots, -m_p)' = \Sigma_U^{-1} \mathbf{c}_U,$$

kde  $\mathbf{c}_U \equiv (C(\mathbf{s}_1 - \mathbf{s}_0), \dots, C(\mathbf{s}_n - \mathbf{s}_0), f_0(\mathbf{s}_0), f_1(\mathbf{s}_0), f_p(\mathbf{s}_0))'$  a  $\Sigma_U$  je symetrická matice typu  $(n + p + 1) \times (n + p + 1)$

$$\Sigma_U = \begin{cases} C(\mathbf{s}_i - \mathbf{s}_j), & i = 1, \dots, n; j = 1, \dots, n, \\ f_{j-1-n}(\mathbf{s}_i), & i = 1, \dots, n; j = n + 1, \dots, n + p + 1, \\ 0, & i = n + 1, \dots, n + p + 1; j = n + 1, \dots, n + p + 1, \end{cases}$$

a již nevyžadujeme, aby  $f_0 \equiv 1$ .

Koeficienty  $\boldsymbol{\lambda}$  minimalizující (5.15) jsou dány (viz [1]) vztahem

$$\boldsymbol{\lambda}' = \{\mathbf{c} + \mathbf{X}(\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}(\mathbf{x} - \mathbf{X}'\mathbf{C}^{-1}\mathbf{c})\}'\mathbf{C}^{-1},$$

kde  $\mathbf{c} = (C(\mathbf{s}_1 - \mathbf{s}_0), \dots, C(\mathbf{s}_n - \mathbf{s}_0))'$  a matice  $\mathbf{C}$  je matice typu  $n \times n$ , kde

$$\mathbf{C} = \begin{pmatrix} C(\mathbf{0}) & C(\mathbf{s}_1 - \mathbf{s}_2) & \dots & C(\mathbf{s}_1 - \mathbf{s}_n) \\ \vdots & \vdots & \ddots & \vdots \\ C(\mathbf{s}_n - \mathbf{s}_1) & C(\mathbf{s}_n - \mathbf{s}_2) & \dots & C(\mathbf{0}) \end{pmatrix}.$$

Navíc

$$\mathbf{m}' = (\mathbf{x} - \mathbf{X}'\mathbf{C}^{-1}\mathbf{c})'(\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}.$$

### 5.2.2 Kriging rozptyl a intervaly predikce

Kriging rozptyl (predikční rozptyl, minimální střední kvadratická chyba predikce) je

$$\begin{aligned} \sigma_k^2(\mathbf{s}_0) &= \min_{\boldsymbol{\lambda} \in \mathcal{U}} MSE(\mathbf{s}_0) \\ &= \boldsymbol{\lambda}'_U \boldsymbol{\gamma}_U = \sum_{i=1}^n \lambda_i \gamma(\mathbf{s}_0 - \mathbf{s}_i) + \sum_{j=1}^{p+1} m_{j-1} f_{j-1}(\mathbf{s}_0) \\ &= \boldsymbol{\gamma}' \boldsymbol{\Gamma}^{-1} \boldsymbol{\gamma} - (\mathbf{x} - \mathbf{X}'\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma})'(\mathbf{X}'\boldsymbol{\Gamma}^{-1}\mathbf{X})^{-1}(\mathbf{x} - \mathbf{X}'\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}). \end{aligned} \quad (5.21)$$

Kromě toho,

$$\sigma_k^2(\mathbf{s}_0) = 2 \sum_{i=1}^n \lambda_i \gamma(\mathbf{s}_0 - \mathbf{s}_i) - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(\mathbf{s}_i - \mathbf{s}_j) \quad (5.22)$$

Pokud je proces  $\mathbf{Z}$  stacionární druhého řádu a jeho kovariogram existuje, potom kriging rozptyl můžeme vyjádřit jako

$$\begin{aligned} \sigma_k^2(\mathbf{s}_0) &= C(\mathbf{0}) - \sum_{i=1}^n \lambda_i C(\mathbf{s}_0 - \mathbf{s}_i) + \sum_{j=1}^{p+1} m_{j-1} f_{j-1}(\mathbf{s}_0) \\ &= C(\mathbf{0}) - \mathbf{c}'\mathbf{C}^{-1}\mathbf{c} + (\mathbf{x} - \mathbf{X}'\mathbf{C}^{-1}\mathbf{c})'(\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}(\mathbf{x} - \mathbf{X}'\mathbf{C}^{-1}\mathbf{c}) \end{aligned}$$

a také

$$\sigma_k^2(\mathbf{s}_0) = C(\mathbf{0}) - 2 \sum_{i=1}^n \lambda_i C(\mathbf{s}_0 - \mathbf{s}_i) + \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(\mathbf{s}_i - \mathbf{s}_j)$$

Z rovnic (5.6), (5.16) a (5.21) sestrojíme asymptotický predikční interval

$$A \equiv (Z^*(\mathbf{s}_0) - u_{0.975} \sigma_k(\mathbf{s}_0), Z^*(\mathbf{s}_0) + u_{0.975} \sigma_k(\mathbf{s}_0)),$$

který je 95% interval predikce pro  $Z(\mathbf{s}_0)$ .

### 5.2.3 Odhad střední hodnoty

Ještě nám zbývá odhadnout koeficienty  $\boldsymbol{\beta}$  z rovnice (5.4). Z (5.5) vidíme, že data  $\mathbf{Z}$  vyhovují obecnému lineárnímu modelu, kde  $E(\mathbf{Z}) = \mathbf{X}\boldsymbol{\beta}$  a  $\text{var}(\mathbf{Z}) = \mathbf{C}$ . Odhad zobecněnou metodou nejmenších čtverců je tedy

$$\hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}^{-1}\mathbf{Z} \quad (5.23)$$

a hodnota procesu  $\mathbf{Z}$  v bodě  $\mathbf{s}_0$  je dána vztahem

$$Z^*(\mathbf{s}_0) = \mathbf{Z}'\boldsymbol{\lambda} = \mathbf{c}'\mathbf{C}^{-1}\mathbf{Z} + (\mathbf{x} - \mathbf{X}'\mathbf{C}^{-1}\mathbf{c})'\hat{\boldsymbol{\beta}}_{GLS}. \quad (5.24)$$

## 5.3 Další kriging modely

Kriging modelů existuje celá řada. Tyto modely se většinou liší předpokladem tvaru střední hodnoty procesu nebo v požadavcích na stacionaritu. Mezi nejzákladnější kriging modely patří již zmíněný univerzální kriging model a dále potom např. simple kriging, ordinary kriging, robustní kriging apod.

### 5.3.1 Ordinární kriging

Při ordinárním krigingu předpokládáme opět model ve tvaru

$$\mathbf{Z}(\mathbf{s}) = \mu + \delta(\mathbf{s}), \quad \mathbf{s} \in \mathbf{D},$$

ale tentokrát máme  $\mu \in \mathbf{R}$  neznámou. Prediktor očekáváme ve tvaru

$$Z^*(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i), \quad \sum_{i=1}^n \lambda_i = 1.$$

Druhá podmínka zaručuje nestrannost prediktoru, tzn.  $E(Z^*(\mathbf{s}_0)) = \mu = E(Z(\mathbf{s}_0))$ .

### 5.3.2 Simple kriging

V případě, že střední hodnota  $\mu(\mathbf{s}) = E\mathbf{Z}(\mathbf{s})$  je známá, pak mluvíme o prosté metodě kriging (simple kriging). U tohoto modelu se nevyžaduje, aby  $\sum_{i=1}^n \lambda_i = 1$ . V tomto případě je známý vektor  $\boldsymbol{\beta}$  a vzorec pro predikci  $Z(\mathbf{s}_0)$  prostou metodou kriging lze zapsat ve tvaru

$$Z^*(\mathbf{s}_0) = \mathbf{c}'\mathbf{C}^{-1}\mathbf{Z} + (\mathbf{x} - \mathbf{X}'\mathbf{C}^{-1}\mathbf{c})'\boldsymbol{\beta}.$$

## 6 Odhad variogramu

Pokud prakticky aplikujeme kriging metodu, setkáváme se s problémem, že neznáme kovariogram  $\mathbf{C}$  nebo variogram  $2\gamma(\mathbf{h})$  procesu  $\mathbf{Z}$ . Potom potřebujeme nalézt kvalitní odhad  $\mathbf{C}$  nebo  $2\gamma(\mathbf{h})$ , což je obtížný úkol, protože obvykle nelze chybový proces  $\delta$  přímo pozorovat. Potom se k odhadu kovariogramu nebo variogramu chybového procesu  $\delta$  užívá nějakých specifických rysů uvažovaného modelu. Často se vychází z předpokladu izotropie a skutečnosti, že funkce  $\mu$  je v nějakém směru konstantní apod.

Při popisu korelovanosti sledovaného procesu  $\mathbf{Z}$  dáme přednost variogramu před kovariogramem, protože třída procesů druhého řádu stacionárních je zahrnuta mezi vnitřně stacionárními procesy. Variogram budeme odhadovat za předpokladu, že sledovaný proces je vnitřně stacionární.

### 6.1 Klasický odhad variogramu

Jednou z možností, jak odhadnout kovarianční strukturu procesu, je využít klasický odhad variogramu. Klasický odhad, který navrhl Matheron (1962), je dán vztahem

$$2\hat{\gamma}(\mathbf{h}) = \frac{1}{|N(\mathbf{h})|} \sum_{N(\mathbf{h})} (Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2, \quad (6.1)$$

kde  $N(\mathbf{h}) = \{(\mathbf{s}_i, \mathbf{s}_j) : \mathbf{s}_i - \mathbf{s}_j = \mathbf{h}; i, j = 1, 2, \dots, n\}$  a  $|N(\mathbf{h})|$  je počet prvků v množině  $N(\mathbf{h})$ . Tento odhad je nestranný, ale je značně ovlivněn atypickými pozorováními, protože ve sčítanci (6.1) se vyskytuje výraz  $(\cdot)^2$ . Atypická pozorování mohou značně zkreslit výsledný odhadovaný variogram.

### 6.2 Robustní odhady variogramu

Právě z toho důvodu, že klasický odhad je velmi citlivý na odlehlá pozorování, navrhli Cressie a Hawkins (viz.[3]) dvě varianty robustního odhadu variogramu. První je tvaru

$$2\gamma(\mathbf{h}) = \frac{[\frac{1}{|N(\mathbf{h})|} \sum_{N(\mathbf{h})} |Z(\mathbf{s}_i) - Z(\mathbf{s}_j)|^{\frac{1}{2}}]^4}{0.457 + \frac{0.494}{|N(\mathbf{h})|}}.$$

Druhá varianta je založena na mediánu a příslušný odhad má při asymptotické korekci vychýlení tvar

$$2\gamma(\mathbf{h}) = \frac{[\text{med}\{|Z(\mathbf{s}_i) - Z(\mathbf{s}_j)|^{\frac{1}{2}} : (\mathbf{s}_i - \mathbf{s}_j) \in N(\mathbf{h})\}]^4}{0.457}.$$

Robustní varianty dávají mnohem kvalitnější odhady než klasický odhad, což lze ověřit v praxi nebo pomocí simulací. Robustní varianty odhadu a klasický odhad se souhrnně označují jako *empirické*.

### 6.3 Základní typy modelových variogramů

Variogram  $2\gamma(\mathbf{h})$  je podmíněně negativně definitní funkce (viz. (4.3)). Proto i na odhad variogramu klademe požadavek, aby byl podmíněně negativně definitní. Potom za předpokladu, že sledovaný proces je izotropický, můžeme popsat *základní typy* modelových variogramů a jim odpovídajících semivariogramů  $\gamma(\mathbf{h})$ , které tomuto požadavku vyhovují a zároveň mají při praktických aplikacích nejčastější uplatnění.

- Lineární model pro  $d \geq 1$ :

$$\gamma(\mathbf{h}) = \begin{cases} 0, & \mathbf{h} = 0, \\ c_0 + b\|\mathbf{h}\|, & \mathbf{h} \neq 0, \end{cases}$$

kde  $c_0 \geq 0$  a  $b \geq 0$  jsou parametry. Na Obr. 8 jsou vykresleny lineární modely pro různé hodnoty parametrů  $c_0$  a  $b$ , aby bylo viditelné, jak volba parametrů může ovlivnit model.

- Sférický model pro  $1 \leq d \leq 3$ :

$$\gamma(\mathbf{h}) = \begin{cases} 0, & \mathbf{h} = 0, \\ c_0 + c_s \left\{ \frac{3}{2} \frac{\|\mathbf{h}\|}{a_s} - \frac{1}{2} \left( \frac{\|\mathbf{h}\|}{a_s} \right)^3 \right\}, & 0 < \|\mathbf{h}\| \leq a_s, \\ c_0 + c_s, & \|\mathbf{h}\| \geq a_s, \end{cases}$$

kde  $c_0 \geq 0, c_s \geq 0$  a  $a_s \geq 0$  jsou parametry. Na Obr. 9 opět můžeme pozorovat vliv volby parametrů na tvar sférického modelu.

- Exponenciální model pro  $d \geq 1$ :

$$\gamma(\mathbf{h}) = \begin{cases} 0, & \mathbf{h} = 0, \\ c_0 + c_e [1 - \exp\{-\frac{\|\mathbf{h}\|}{a_e}\}], & \mathbf{h} \neq 0, \end{cases}$$

kde  $c_0 \geq 0, c_e \geq 0$  a  $a_e \geq 0$  jsou parametry. Na Obr. 10 jsou nakresleny exponenciální modely pro různé hodnoty parametrů  $c_0, c_e$  a  $a_e$ .

- Racionální kvadratický model pro  $d \geq 1$ :

$$\gamma(\mathbf{h}) = \begin{cases} 0, & \mathbf{h} = 0, \\ c_0 + \frac{c_r \|\mathbf{h}\|^2}{1 + \frac{\|\mathbf{h}\|^2}{a_r}}, & \mathbf{h} \neq 0, \end{cases}$$

kde  $c_0 \geq 0, c_r \geq 0$  a  $a_r \geq 0$  jsou parametry. Na Obr. 11 jsou znázorněny racionální kvadratické modely pro různé hodnoty  $c_0, c_r$  a  $a_r$ .

- Mocninný model pro  $d \geq 1$ :

$$\gamma(\mathbf{h}) = \begin{cases} 0, & \mathbf{h} = 0, \\ c_0 + b_m \|\mathbf{h}\|^\lambda, & \mathbf{h} \neq 0, \end{cases}$$

kde  $c_0 \geq 0, b_m \geq 0$  a  $0 \leq \lambda < 2$  jsou parametry. Vliv volby parametrů  $c_0, b_m$  a  $\lambda$  na mocninný model je zviditelněn na Obr.12.

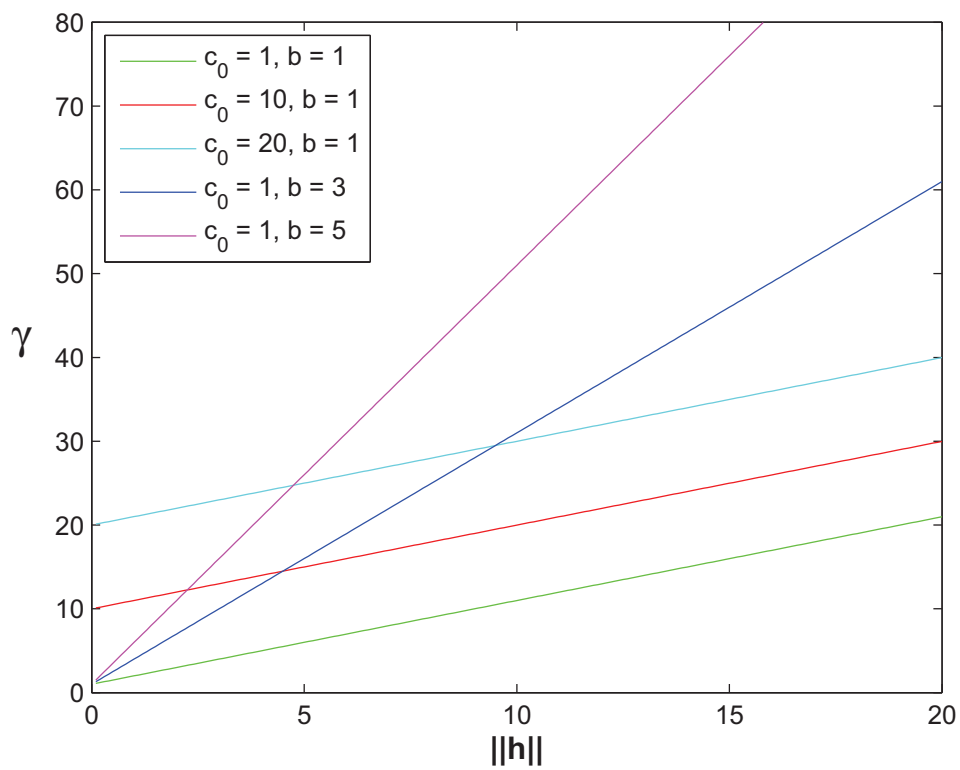
- Vlnový model pro  $d \geq 1$ :

$$\gamma(\mathbf{h}) = \begin{cases} 0, & \mathbf{h} = 0, \\ c_0 + c_w \left\{ 1 - a_w \frac{\sin(\frac{\|\mathbf{h}\|}{a_w})}{\|\mathbf{h}\|} \right\}, & \mathbf{h} \neq 0, \end{cases}$$

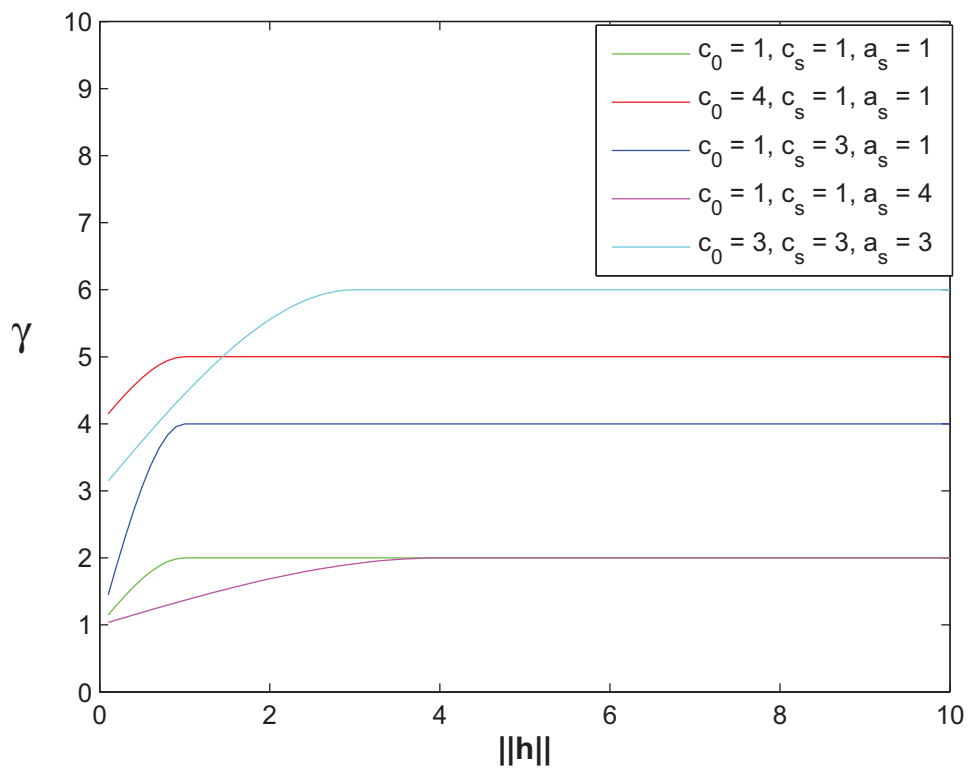
kde  $c_0 \geq 0, c_w \geq 0$  a  $a_w \geq 0$  jsou parametry. Na Obr. 13 jsou vlnové modely pro různé hodnoty parametrů  $c_0, c_w$  a  $a_w$ .

Příslušný odhad semivariogramu dostaneme tak, že proložíme některý z vybraných modelů empirickým variogramem vhodnou statistickou procedurou. Často se pro odhad parametrů uvedených teoretických modelů používá nelineární metoda nejmenších čtverců. Další možností je postupně volit parametry základních typů modelů a takto vzniklé modely vizuálně porovnávat s empirickým odhadem.

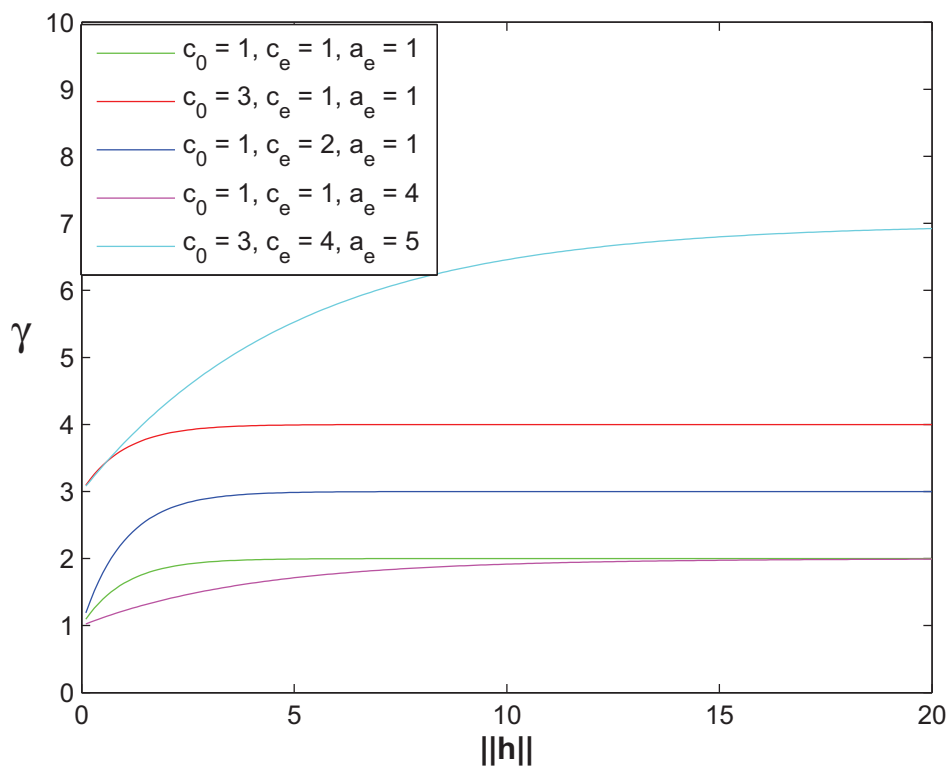
Pokud tedy k příkladu sestrojíme některý z empirických odhadů, potom z něho můžeme odhadnout parametry nějakého ze základních modelů. V mé implementaci používám klasický odhad variogramu a všechny typy uvedených základních modelových variogramů.



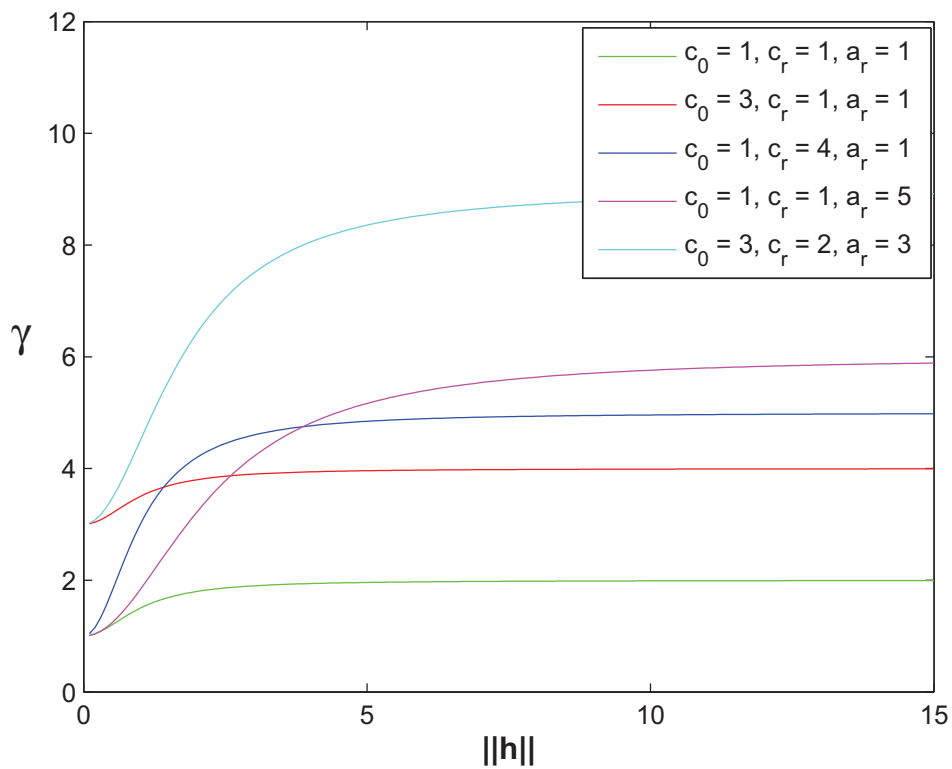
Obrázek 8: Lineární model pro různé hodnoty parametrů  $c_0$  a  $b$



Obrázek 9: Sférický model pro různé hodnoty parametrů  $c_0$ ,  $c_s$  a  $a_s$

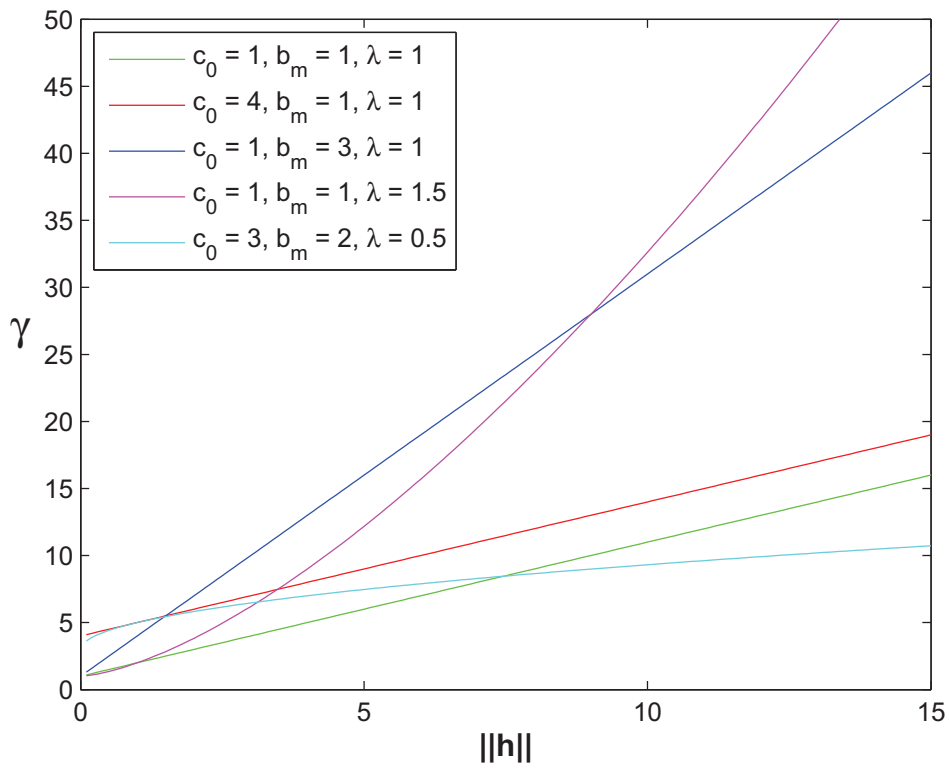


Obrázek 10: Exponenciální model pro různé hodnoty parametrů  $c_0$ ,  $c_e$  a  $a_e$

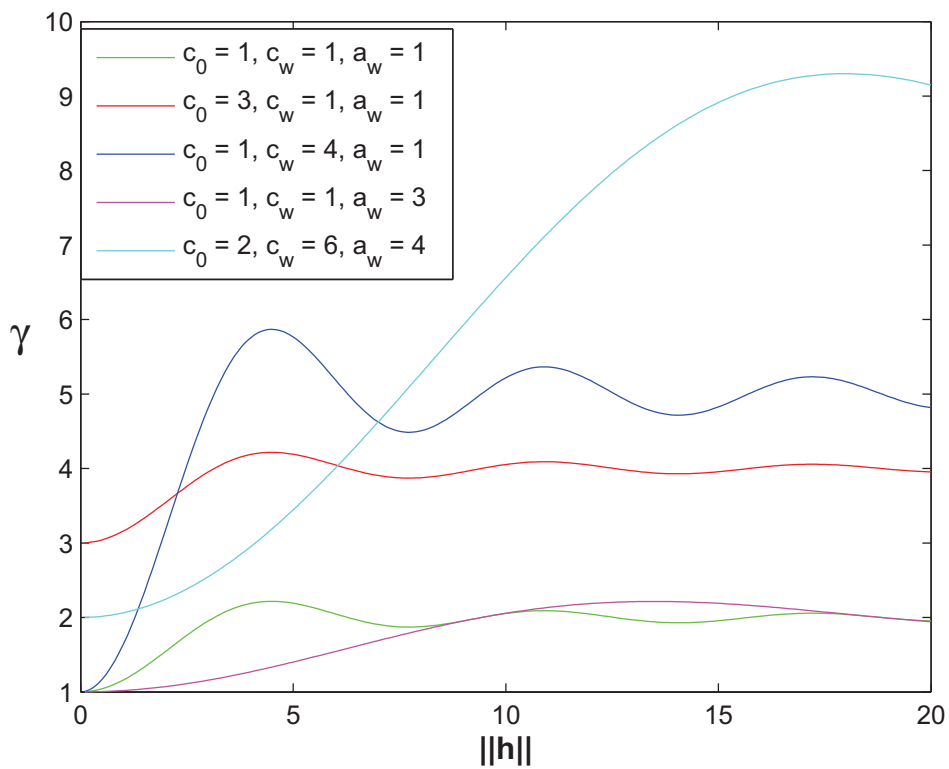


Obrázek 11: Racionální kvadratický model pro různé hodnoty parametrů  $c_0$ ,  $c_r$  a  $a_r$





Obrázek 12: Mocinný model pro různé hodnoty parametrů  $c_0$ ,  $b_m$  a  $\lambda$



Obrázek 13: Vlnový model pro různé hodnoty parametrů  $c_0$ ,  $c_w$  a  $a_w$

## 6.4 Jiný přístup k modelování variogramu

Jiný přístup k modelování variogramu je popsán v článku [2].

**Definice 6.1.** Řekneme, že reálná funkce  $g(\mathbf{h})$ ,  $\mathbf{h} \in \mathbf{R}^d$  je *přípustný variogram*, jestliže

- $g(\mathbf{h})$  je spojitá v  $\mathbf{R}^d$  s případnou výjimkou počátku,
- $g(\mathbf{h})$  je nezáporná, tedy  $g(\mathbf{h}) \geq 0$  pro všechna  $\mathbf{h} \in \mathbf{R}^d$ ,
- $g(\mathbf{h})$  je symetrická, tedy  $g(\mathbf{h}) = g(-\mathbf{h})$  pro všechna  $\mathbf{h} \in \mathbf{R}^d$ ,
- $g(\mathbf{h})$  je podmíněně negativně definitní funkce.

Přípustný variogram  $g(\mathbf{h})$  můžeme získat pomocí funkce  $f(\mathbf{h})$ , pokud  $f(\mathbf{h})$  je vhodná pozitivně semidefinitní funkce, vztahem

$$g(\mathbf{h}) = -f(\mathbf{h}) + c,$$

kde  $c$  je vhodná konstanta. Volbou  $f$  s uvedenými vlastnostmi dostaneme speciální typy variogramů.

Například pro  $d = 2$  dostaneme přípustný model semivariogramu  $2\gamma(\mathbf{h})$ , kde  $\mathbf{h} = (h_1, h_2) \in \mathbf{R}^2$ , pomocí funkce

$$f(h_1, h_2) = \sum_{j_1=-m_1}^{m_1} \sum_{j_2=-m_2}^{m_2} a_{j_1 j_2} \cos(h_1 t_{j_1} + h_2 t_{j_2}),$$

kde  $m_1$  a  $m_2$  jsou daná přirozená čísla,  $t_{j_1}, t_{j_2}, a_{j_1 j_2}$  jsou pro  $j_1 = -m_1, \dots, m_1; j_2 = -m_2, \dots, m_2$  vhodné reálné konstanty vyhovující podmínce

$$c - \sum_{j_1=-m_1}^{m_1} \sum_{j_2=-m_2}^{m_2} a_{j_1 j_2} \geq 0.$$

Jejich doporučená volba vychází ze simulací (viz [2]). Výhodou těchto modelů je, že umožňují modifikovat odhad variogramu na reálná data a že zahrnují také třídu modelů, které nevycházejí z předpokladů izotropie.

## 7 Popis programu

V této kapitole popíšu program, který jsem vytvořila v programovém prostředí MATLAB. Program počítá odhady metodou kriging pro příklady uvedené v kapitole 4, tedy pro regresní rovinu, obecnou lineární regresní plochu a Coal-Ash Data. Výsledky pro jednotlivé příklady budou uvedeny v následujících kapitolách. Program se skládá z řady funkcí, se kterými se nyní seznámíme. Nejprve popíšeme funkce pro simulovaná data.

### 7.1 Zadání vstupních dat

První funkcí je funkce "ZadaniVstupnichDat". V tomto M-filu zadáváme:

- vektory  $x$  a  $y$ , které určují souřadnice bodů, ke kterým u simulovaných dat dopočítáme pomocí zadaného vztahu hodnoty sledovaného procesu;
- souřadnice  $x_0$  a  $y_0$ , ve kterých se má odhadnout (případně predikovat) hodnota procesu (tyto hodnoty lze ve funkci "CelkovyOdhad" změnit nebo rozšířit na vektory souřadnic bodů predikce).

Pomocí zadaných  $x, y, x_0, y_0$  funkce vypočítá:

- $s_0 = [x_0, y_0]$ , což je matice souřadnic predikovaných bodů;
- $p_p = \text{size}(x)$ , čili  $p_p$  je rozměr vektoru  $x$  (resp.  $y$ , protože vektory  $x$  a  $y$  mají podle zadání v případě simulovaných dat stejný rozměr),  $p_p = (\text{počet řádků}, \text{počet sloupců})$ ;
- $n_n$  je počet různých  $x$ -ových ( $y$ -ových) souřadnic;
- $Z$  vypočítá hodnoty procesu, které by mohly být teoreticky naměřeny (zadokumentováním řádku volíme mezi regresní rovinou a obecnou lineární regresní plochou, která je lineární kombinací funkcí sin a cos), způsob zadání v MATLABu je na Obr.14;

```
for i = 1:nn
    for j = 1:nn
        Z(i,j) = x(i)+2*y(j); %%ROVINA
        %Z(i,j) = (3*cos(x(i))+10*sin(0.2*y(j))); %%SIN&COS
    end
end
```

Obrázek 14: Způsob zadání výpočtu hodnot  $Z$

- $\delta = \text{normrnd}(0, 0.2, n_n, n_n)$  je normálně rozdělený náhodný chybový vektor s nulovou střední hodnotou, se směrodatnou odchylkou 0.2 o rozměrech  $nn \times nn$ ;
- $ZS = Z + \delta$ , kde  $ZS$  jsou hodnoty získané tak, že teoreticky vypočítané hodnoty  $Z$  zatížíme náhodnou chybou  $\delta$ ;
- matici  $data = [xS, yS, ZSS]$ , kde sloupce  $xS$  a  $yS$  vyjadřují všechny možné kombinace prvků vektorů  $x$  a  $y$  a  $ZSS$  jsou k nim příslušné hodnoty  $ZS$ ;

- $p = \text{size}(xS)$ , čili  $p = (\text{počet řádků}, \text{počet sloupců})$  vektoru  $xS$  (a i  $yS$ , protože vektory  $xS$  a  $yS$  mají stejný rozměr);
- $n = \text{počet různých bodů vzniklých kombinací prvků vektoru } x \text{ a vektoru } y (n = n_n^2)$ ;
- $h$  je matice vzdáleností jednotlivých bodů pozorování,  $h$  je napočítáno Euklidovskou metrikou (tedy jestliže máme body  $s_1 = [xS_1, yS_1]$  a  $s_2 = [xS_2, yS_2]$ , potom  $h(s_1, s_2) = \sqrt{(xS_1 - xS_2)^2 + (yS_1 - yS_2)^2}$ ). Příklad zápisu tohoto výpočtu je na Obr.(15).

```

for i = 1:n
    for j = 1:n
        h(i,j) = sqrt((xS(i)-xS(j))^2 +(yS(i)-yS(j))^2);
    end
end

```

Obrázek 15: Zápis výpočtu matice  $h$  v MATLABu

## 7.2 Funkce Mezikruží

Funkce "Mezikruzi" přebírá od funkce "ZadaniVstupnichDat" hodnoty  $xS, yS, x0, y0, p, n, data, ZSS$  a  $h$ . Pro výpočet odhadu variogramu v bodě  $s0 = [x0, y0]$  používaný klasický odhad variogramu, který je daný vztahem (6.1), kde počítáme s množinou  $N(\mathbf{h})$ . Oblast pozorování tedy potřebujeme rozdělit podle vzdáleností bodů  $s = [xS, yS]$  od bodu  $s0$ . Ve vztahu (6.1) je  $N(\mathbf{h}) = \{(\mathbf{s}_i, \mathbf{s}_j) : \mathbf{s}_i - \mathbf{s}_j = \mathbf{h}; i, j = 1, 2, \dots, n\}$ , v této funkci ale oblast rozdělíme pomocí kružnic tak, že

$$N_k(\mathbf{h}) = \{(\mathbf{s}_i, \mathbf{s}_j) : \|\mathbf{h}_{k-1}\| \leq \|\mathbf{s}_i - \mathbf{s}_j\| \leq \|\mathbf{h}_k\|; i, j = 1, 2, \dots, n\},$$

kde  $N_k(\mathbf{h})$  je  $k$ -té mezikruží rozdělující oblast.

V této funkci se zadává:

- *PocKruznic*, což je konstanta udávající počet kružnic, kterými chceme danou oblast rozdělit;
- *MaxPolomer* = maximální vzdálenost od bodu  $s0$  taková, že pokud body leží v kružnici  $k(s0, \text{MaxPolomer})$ , tak je ještě zahrneme do výpočtu.

Pomocí funkce získáme:

- $hS0$ , vektor vzdáleností jednotlivých bodů  $s = [xS, yS]$  od bodu  $s0$ ;
- *Krok* =  $\text{MaxPolomer}/\text{PocKruznic}$  udává krok, se kterým rozdělíme uvažovanou oblast;
- $r = 0:\text{Krok}:\text{MaxPolomer}$  jsou poloměry kružnic, kterými danou oblast rozdělíme na požadovaná mezikruží;
- matici  $M$ , jejíž první tři sloupce jsou shodné s maticí  $data$  a navíc jsou přidány sloupce vyjadřující jednotlivá mezikruží, hodnoty v těchto sloupcích jsou buď 1 nebo 0 podle toho, zda daný bod leží v mezikruží, či nikoliv;
- $q$  = počet sloupců matice  $M$ , tedy  $q - 3$  vyjadřuje počet mezikruží dané oblasti.

91	7	1	8.659	0	0	1	0	0	0
92	7	2	10.94	0	0	1	0	0	0
93	7	3	13.2875	0	0	1	0	0	0
94	7	4	14.9166	0	1	0	0	0	0
95	7	5	17.0034	0	1	0	0	0	0
96	7	6	19.3363	0	1	0	0	0	0
97	7	7	20.7991	0	1	0	0	0	0
98	7	8	23.1855	0	1	0	0	0	0
99	7	9	25.1002	0	1	0	0	0	0
100	7	10	27.2108	0	1	0	0	0	0
101	7	11	28.8858	0	0	1	0	0	0
102	7	12	30.7699	0	0	1	0	0	0
103	7	13	32.9634	0	0	1	0	0	0
104	7	14	35.0082	0	0	0	1	0	0
105	7	15	36.9466	0	0	0	1	0	0
106	8	1	10.0257	0	0	1	0	0	0
107	8	2	12.1016	0	0	1	0	0	0

Obrázek 16: Ukázka, jak může vypadat matice  $M$

Na Obr.(16) je část matice  $M$  a tedy vidíme, že například bod  $s = [7, 1]$  leží ve třetím mezikruží, bod  $s = [7, 4]$  leží ve druhém mezikruží, bod  $s = [7, 14]$  leží ve čtvrtém mezikruží apod.

### 7.3 Klasický odhad variogramu

Funkce "KlasickyOdhadVariogramu" vypočítává samozřejmě hodnoty klasického odhadu variogramu v závislosti na  $r$ . Z předchozích funkcí si přebírá hodnoty  $n, q, M, ZSS$  a výstupem této funkce je:

- *KvadratHodnotProcesu*, který ve vztahu (6.1) vypočítá  $(Z(s_i) - Z(s_j))^2$ , kde  $N(\mathbf{h}) = \{(s_i, s_j) : s_i - s_j \leq \mathbf{h}; i, j = 1, 2, \dots, n\}$ ;
- *Suma*, která je součtem získaných hodnot *KvadratHodnotProcesu* a dostaneme tak  $\sum_{N(h)} (Z(s_i) - Z(s_j))^2$  ze vztahu (6.1);
- *PocetPrvkuNH* udává, kolik bodů leží v uvažovaných mezikružích;
- *KOV*, což jsou právě hodnoty klasického odhadu variogramu v závislosti na  $r$ .

### 7.4 Základní typy modelových variogramů

V kapitole 6 jsem uvedla 6 základních typů modelových variogramů. Každý je implementován jako samostatná funkce a k volbě modelu přistupujeme až ve funkci "KrigingOdhad". Pomocí základních typů modelových variogramů získáme  $\mathbf{c}$  a  $\mathbf{C}$ , které jsou potřebné pro výpočet (5.23) a (5.24). Všechny funkce využívají hodnoty  $n, h, MaxPolomer, hS0$  z předchozích funkcí.

- Lineární model se využívá ve funkci "LinearniOdhadVariogramu". Výstupy této funkce jsem označila *GammaLM*, *GammaLMS0*, kde *GammaLMS0* se v následující funkci dosadí za  $c$  a *GammaLM* za  $C$ . V této funkci je také potřeba zadat hodnoty parametrů  $c_0$  a  $b$ , které získáme pomocí klasického odhadu variogramu.

- Funkce "SferickyOdhadVariogramu" opět vypočítá hodnoty  $GammaSMS0$  a  $GammaSM$ , které lze postupně dosadit za  $c$  a  $C$ , pomocí sférického modelu. Pomocí klasického odhadu je zapotřebí odhadnout parametry  $c_0$ ,  $c_s$  a  $a_s$  a zadat je do této funkce.
- Výstupem funkce "ExponencialniOdhadVariogramu" jsou proměnné  $GammaEMSO$  a  $GammaEM$ , které využijeme ve funkci "KrigingOdhad" za  $c$  a  $C$  a opět je nutné dosadit parametry exponenciálního modelu  $c_0$ ,  $c_e$  a  $a_e$ .
- Pro racionální kvadratický model máme funkci "RacionalniKvadratickyOdhadVariogramu", díky které dostaneme hodnoty  $GammaRKM$  a  $GammaRKMS0$ . Parametry tohoto modelu jsou  $c_0$ ,  $c_r$  a  $a_r$  a je potřeba je zadat v této funkci.
- $GammaMM$  a  $GammaMMS0$  jsou výstupy funkce "MocninnyOdhadVariogramu". Parametry mocninného modelu jsou  $c_0$ ,  $b_m$  a  $\lambda$ .
- Vlnový model využívá funkce "VlnovyOdhadVariogramu", jejíž výstupy jsou  $GammaVMS0$  a  $GammaVM$  a zadávají se zde parametry  $c_0$ ,  $c_w$  a  $a_w$ .

## 7.5 Kriging odhad

Výstupem funkce "KrigingOdhad" je požadovaná predikce  $Z^*(s_0)$  daná vztahem (5.24). Funkce z předešlých funkcí využívá  $x0, y0, n, xS, yS, ZSS, Gamma, GammaS0$ , kde za  $Gamma$  a  $GammaS0$  se dosadí příslušné veličiny podle volby typu modelu. Ve funkci je také ještě nutno zadat  $X$  a  $x$ , kde

$$\mathbf{X} = \begin{pmatrix} f_0(\mathbf{s}_1) & \dots & f_p(\mathbf{s}_1) \\ \vdots & \ddots & \vdots \\ f_0(\mathbf{s}_n) & \dots & f_p(\mathbf{s}_n) \end{pmatrix}, \mathbf{x} = \begin{pmatrix} f_0(\mathbf{s}_0) \\ \vdots \\ f_p(\mathbf{s}_0) \end{pmatrix}.$$

podle kapitoly 5. Protože v programu již  $x$  máme, jsou tyto proměnné označeny  $F$  a  $f$ .

## 7.6 Celkový odhad

Pokud bychom chtěli predikovat pouze v bodě  $s_0$ , vystačili bychom si s dosud uvedenými funkcemi. Funkce "CelkovyOdhad" rozšiřuje  $x0$  a  $y0$  na vektory a používá dosud uvedené funkce v cyklu, dokud je nepoužije pro všechny zadané body  $s_0$ . Nakonec je ze všech  $Z^*(s_0)$  sestavena matice a je označena  $ZhvS$ .

U reálného příkladu je vypuštěna simulační část. U reálných dat se funkce, přes kterou se zadávají vstupní data nazývá "CoalAshData" a rozdíl oproti funkci "ZadaniVstupnich-Dat" je v tom, že nemusíme zadávat žádný vztah pro  $Z$  a ani náhodnou chybu  $\delta$ . Matice *data* je zadána přímo a je opět pro další výpočty rozdělena na  $xS, yS$  a  $ZSS$ . Ostatní funkce uvedené pro simulovaná data se používají stejně i u dat Coal-Ash.

Moje programová implementace má dvě varianty, pro simulovaná a reálná data, které jsou uloženy samostatně. CD přiložené k této práci obsahuje právě tyto programy.

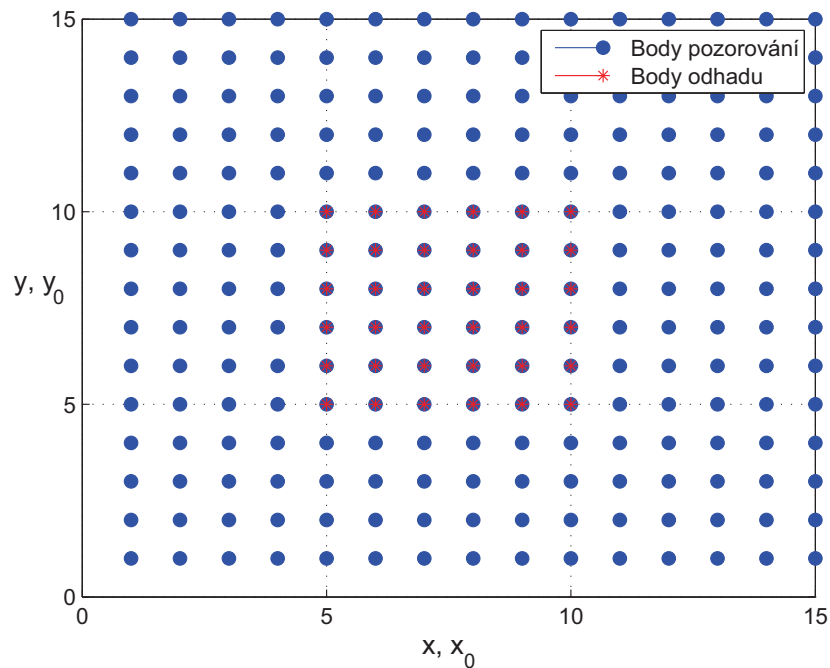
## 8 Kriging metoda pro regresní rovinu

Prvním příkladem, který jsem řešila mým programem je příklad regresní roviny představený v kapitole 4. Po využití funkcí "ZadaniVstupnichDat" a "Mezikruzi", jejichž pomocí vytvoříme hodnoty potřebné pro další výpočty a uvažovanou oblast rozdělíme na mezikruží, použijeme funkci "KlasickyOdhadVariogramu". Dostaneme hodnoty, ze kterých odhadneme parametry základních typů modelových variogramů. V tomto případě nám stačí získat klasický odhad variogramu pro jeden bod  $s_0$ , odhadnout parametry základních modelů a tyto parametry potom využít pro všechny body predikce.

Na základě grafického průběhu empirického odhadu variogramu byly vybrány následující parametry jednotlivých základních typů modelových variogramů:

- Lineární model:  $c_0 = 1, b = 2$ ;
- Sférický model:  $c_0 = 10, c_s = 15$  a  $a_s = 4$ ;
- Exponenciální model:  $c_0 = 3, c_e = 7$  a  $a_e = 5$ ;
- Racionální kvadratický model:  $c_0 = 8, c_r = 5$  a  $a_r = 3$ ;
- Mocninný model:  $c_0 = 3, b_m = 2$  a  $\lambda = 1.8$ ;
- Vlnový model:  $c_0 = 1, c_w = 8$  a  $a_w = 2$ .

Pro regresní rovinu a pro jednotlivé základní typy modelů provedeme odhad v 36-ti bodech, které leží uvnitř pozorované oblasti, jak je znázorněno na Obr.17. Ve funkci "CelkovyOdhad" tedy rozšíříme  $x_0$  a  $y_0$  na vektory.



Obrázek 17: Znázornění bodů, ve kterých známe hodnoty procesu  $Z$  a bodů, ve kterých budeme odhadovat  $Z^*$

Jak je viditelné na Obr.17, body, ve kterých budeme odhadovat  $Z^*$  se kryjí s body, ke kterým hodnotu  $Z$  známe. Je to z toho důvodu, abychom mohli pro každý typ základního modelu variogramu vypočítat rezidua vztahem  $r(s_i) = Z(s_i) - Z^*(s_i)$ , kde  $s_i = [x(i), y(i)]$ . Odhady  $Z^*$  provádíme pro  $x = y = 5 : 1 : 10$ . Na Obr.18 jsou připomenuty hodnoty  $ZS$  pro regresní rovinu a je zvýrazněna oblast bodů, ke kterým odhadujeme  $Z^*$ .

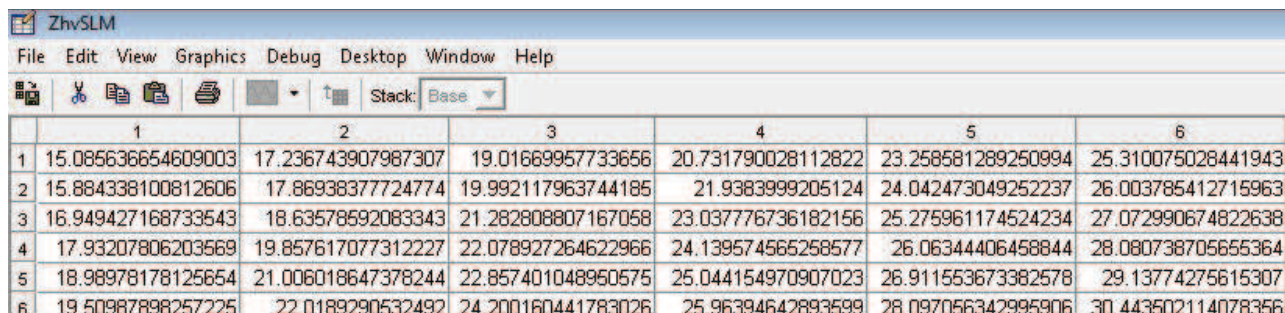
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	3.1057	4.6684	7.2358	9.0964	11.0237	12.8667	15.0629	16.8572	18.7948	21.3807	22.889	25.0582	27.2684	29.373	31.1478
2	4.4578	6.0116	8.0409	9.9508	12.0621	14.1666	15.9382	18.1651	19.5851	21.989	23.9278	25.9658	28.011	30.2183	32.3089
3	5.0246	6.7853	9.1267	10.8568	13.0415	15.1724	17.3486	19.4762	21.2432	22.8217	25.1126	27.0164	29.0751	30.9463	33.1314
4	5.5767	8.0335	10.1887	12.1444	14.1568	16.1314	17.9189	20.3156	21.9964	23.7975	26.1189	27.8813	29.8752	32.1494	33.6101
5	6.8958	9.4712	11.1249	13.0204	15.1142	17.1148	18.8608	20.8738	23.0638	24.7808	27.0341	29.1895	30.9351	33.3415	35.1012
6	7.996	10.3121	12.0694	13.9876	16.4271	18.1206	19.8843	21.6559	23.8764	26.4949	28.0568	30.4406	32.167	33.9257	35.4787
7	8.9103	10.9364	12.8574	15.1109	16.8877	19.2347	21.0734	22.6445	24.9411	26.848	28.6992	30.9132	33.4088	35.0561	37.0911
8	10.0635	12.0268	14.0349	16.0532	17.8037	20.053	22.0866	24.0559	25.8027	27.8558	30.1703	31.9311	34.2763	36.0373	38.4253
9	11.2419	13.3697	15.0468	17.0121	18.8626	21.3146	23.3665	25.1431	27.1574	28.9442	30.8919	33.142	34.9257	36.9521	38.7437
10	11.8043	13.6857	16.3845	18.2879	20.1006	22.0216	23.9196	26.4403	28.2065	29.8335	31.7087	34.1743	36.0052	37.8223	39.8774
11	13.1128	15.2611	16.8814	18.9529	20.8639	22.5893	24.9094	26.7552	29.2242	31.0294	32.6968	34.6359	37.0376	39.2218	40.9649
12	13.7335	15.8591	18.0648	19.9494	22.165	23.8696	26.1318	27.854	30.122	31.882	34.2688	36.0497	37.9183	39.886	41.7484
13	14.6172	16.8362	18.8833	20.9284	22.7222	24.9366	26.9594	29.0846	31.3483	33.113	34.7944	37.3839	38.8807	40.9122	43.3917
14	16.0316	18.1097	20.3231	22.0633	24.1809	25.9787	28.2688	30.1749	31.9306	33.8406	36.0063	38.1302	39.9885	42.1293	44.093
15	16.8465	19.0935	21.34	22.8442	24.9183	27.1319	28.9776	31.0682	33.3749	35.2009	36.9884	38.7788	41.3213	42.9767	45.2807

Obrázek 18: Znázornění hodnot  $Z$ , které budeme srovnávat s odhadnutými hodnotami

## 8.1 Odhady regresní roviny

### 8.1.1 Lineární variogram

Pomocí lineárního modelu získáme  $Z^*$ , které jsem v MATLABu označila jako ZhvSLM (viz. Obr.19).



	1	2	3	4	5	6
1	15.085636654609003	17.236743907987307	19.01669957733656	20.731790028112822	23.258581289250994	25.310075028441943
2	15.884338100812606	17.86938377724774	19.992117963744185	21.9383999205124	24.042473049252237	26.003785412715963
3	16.949427168733543	18.63578592083343	21.282808807167058	23.037776736182156	25.275961174524234	27.072990674822638
4	17.932078062035669	19.857617077312227	22.078927264622966	24.139574565258577	26.06344406458844	28.080738705655364
5	18.98978178125654	21.006018647378244	22.857401048950575	25.044154970907023	26.911553673382578	29.13774275615307
6	19.50987898257225	22.0189290532492	24.200160441783026	25.96394642893599	28.097056342995906	30.443502114078356

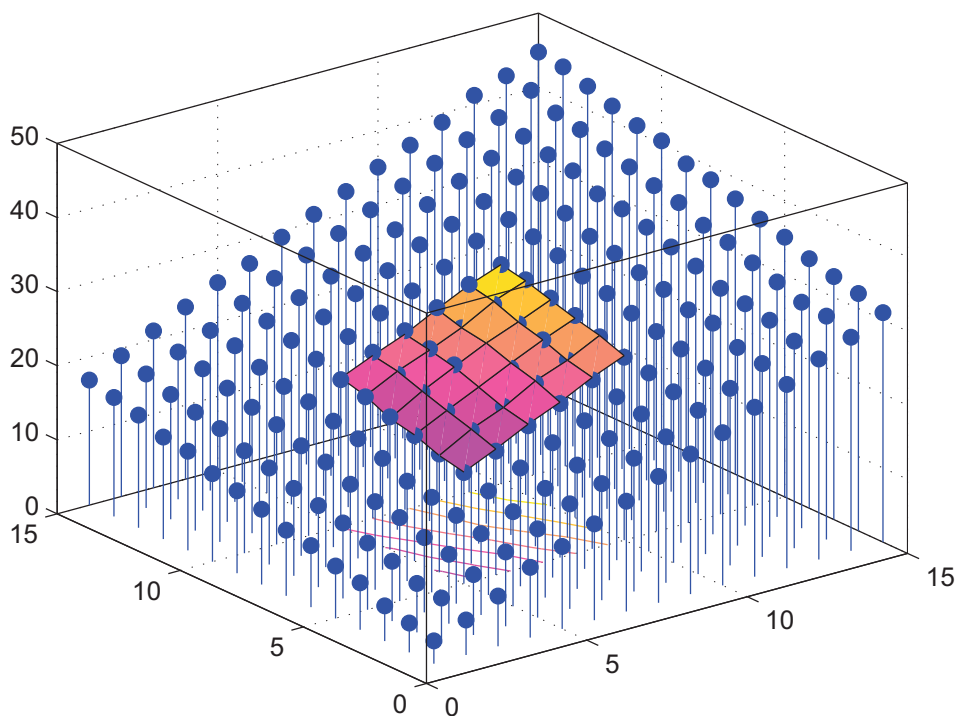
Obrázek 19: Hodnoty  $Z^*$  získané pomocí lineárního modelu variogramu

Vykreslené hodnoty  $Z^*$  pro uvažované hodnoty a hodnoty pozorované jsou společně vykresleny na Obr.20. Odhadnuté hodnoty jsou pro lepší vizualizaci proloženy plochou.

Již zmíněná rezidua jsou pro lineární model v MATLABu označena *reziduaLM* a jejich hodnoty jsou na Obr.21.

Stejné výpočty a vizualizace provedeme i provedeme i pro ostatní základní typy variogramů, proto uvedu pro následující odhady pouze obrázky s popisky.





Obrázek 20: Vykreslení hodnot  $Z^*$  pro lineární model variogramu

	1	2	3	4	5	6
1	0.081197436187859	-0.046778152511973	0.028995925842434	0.391765878946355	0.024775588500301	-0.672145507407837
2	-0.056861590814371	0.390664583788883	-0.1139008087662	0.124498615202889	-0.25161753133861	-0.010766183749073
3	0.026619077833086	0.119333123486562	-0.098155891674633	0.187037996870391	-0.476866096981823	0.062161926830601
4	0.491658485840706	0.096639425887499	0.291314950998654	0.086245540956909	-0.211123293803578	-0.258911768645213
5	-0.069657977228733	-0.047221314548636	0.035947906236473	-0.130932591878253	-0.100452489853041	-0.51600990425074
6	0.440329529382286	0.133708766696966	-0.307897037907694	0.201187447982786	-0.124137684735789	4.156675004196586e-013

Obrázek 21: Rezidua odpovídajících si bodů pro lineární model

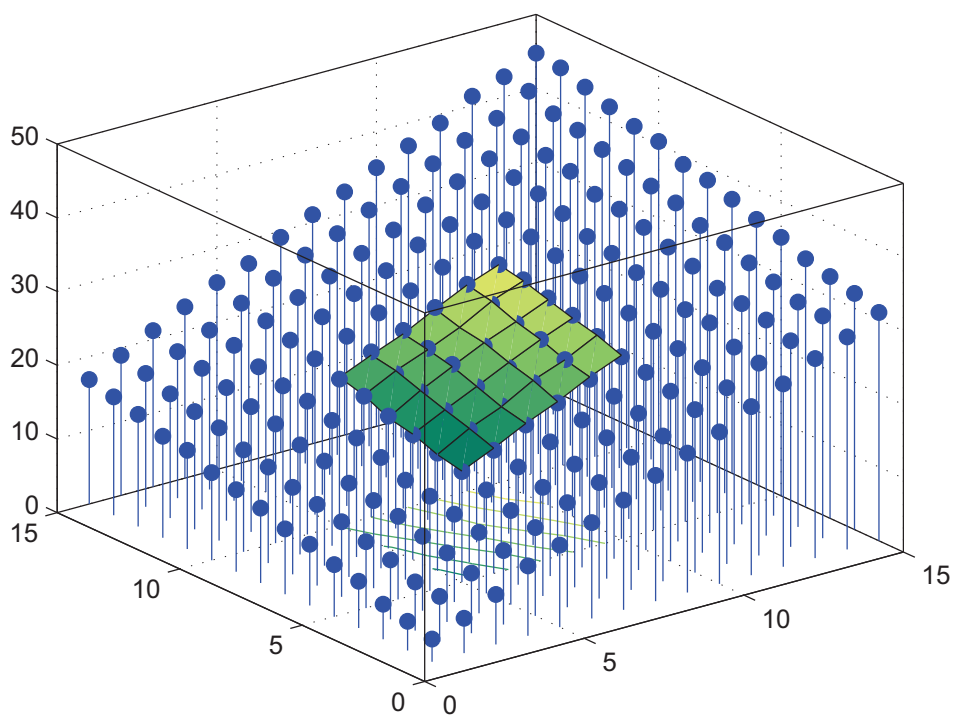
### 8.1.2 Sférický variogram

	1	2	3	4	5	6
1	15.085636654609123	17.23674390798749	19.016699577336396	20.731790028113622	23.2585812892508	25.310075028442228
2	15.884338100812801	17.869383777248032	19.992117963745788	21.93839992051283	24.042473049255353	26.003785412716454
3	16.94942716873355	18.635785920836476	21.28280880716632	23.037776736177882	25.275961174524305	27.072990674828468
4	17.93207806203691	19.857617077311833	22.078927264619352	24.13957456525836	26.063444064584484	28.080738705655683
5	18.98978178125658	21.006018647384817	22.857401048949654	25.044154970904394	26.911553673383207	29.137742756159945
6	19.50987898257228	22.01892905324952	24.20016044179042	25.963946428935735	28.09705634299938	30.44350211407887

Obrázek 22: Hodnoty  $Z^*$  získané pomocí sférického modelu variogramu

	1	2	3	4	5	6
1	0.081197436187738	-0.046778152512154	0.028995925842597	0.391765878945556	0.024775588500496	-0.672145507408121
2	-0.056861590814567	0.390664583788592	-0.113900808767802	0.124498615202459	-0.251617531341726	-0.010766183749563
3	0.026619077833079	0.119333123483518	-0.098155891673894	0.187037996874665	-0.476866096981894	0.062161926824771
4	0.491658485840485	0.096639425887894	0.291314951002267	0.086245540957126	-0.21112329379962	-0.258911768645532
5	-0.069657977228772	-0.047221314555209	0.035947906237393	-0.130932591875624	-0.100452489853669	-0.516009904257615
6	0.440329529382257	0.133708766696646	-0.307897037915087	0.201187447983042	-0.124137684739264	-9.947598300641403e-014

Obrázek 23: Rezidua odpovídajících si bodů pro sférický model



Obrázek 24: Vykreslení hodnot  $Z^*$  pro sférický model variogramu

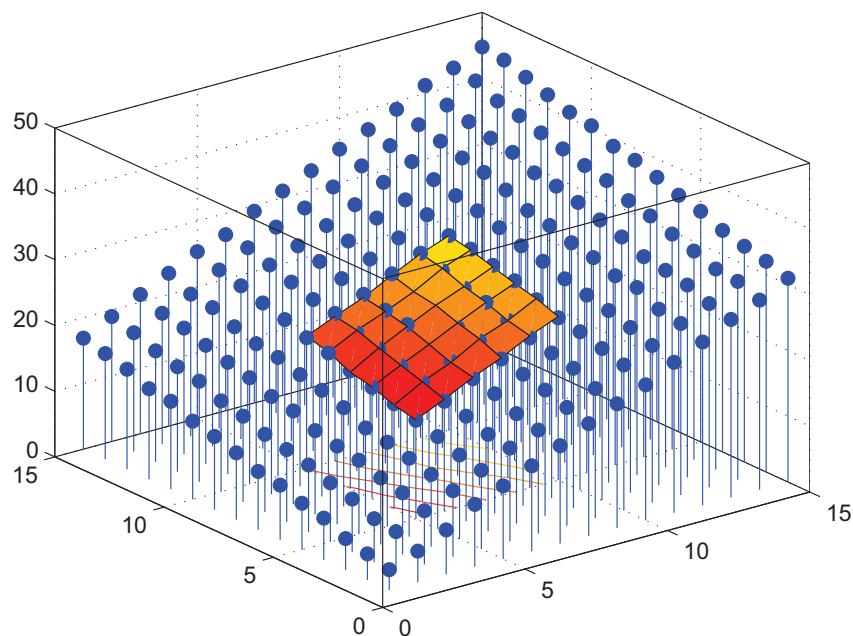
### 8.1.3 Racionální kvadratický variogram

	1	2	3	4	5	6
1	15.052161679775995	17.008830213038124	19.16521852379547	20.962064783072513	23.143539613961593	24.917539641264998
2	15.749832501545876	18.143860116244134	19.916263477241795	22.31847394641832	23.93454778389816	26.12095569993781
3	17.027730442159527	19.153253161880855	21.15236658883255	23.23717110329279	25.166185642783315	26.886691529926807
4	17.710834970292304	20.2267316653795	22.317678611925082	24.37833091076197	26.03757954542468	27.857592305470057
5	19.25945127110691	21.392100045380513	23.07325439762821	24.917965346275444	27.320170242688867	28.90129103545582
6	20.129915724364675	22.000351510309905	23.681783867145008	25.884594452354445	27.972918658261804	29.563470126008664

Obrázek 25: Hodnoty  $Z^*$  získané pomocí racionálního kvadratického modelu variogramu

	1	2	3	4	5	6
1	0.081197436187651	-0.046778152511759	0.028995925842473	0.391765878945961	0.024775588500368	-0.672145507407567
2	-0.056861590814	0.390664583788816	-0.113900808767301	0.124498615201944	-0.251617531339779	-0.010766183749261
3	0.026619077832926	0.119333123485369	-0.09815589167583	0.187037996870622	-0.476866096984192	0.062161926829141
4	0.491658485840809	0.096639425887297	0.291314950998089	0.086245540957179	-0.211123293804587	-0.258911768645934
5	-0.069657977228964	-0.047221314549919	0.035947906234995	-0.130932591879414	-0.100452489855133	-0.516009904252194
6	0.440329529382691	0.133708766696884	-0.307897037909076	0.201187447982605	-0.124137684737359	3.232969447708456e-013

Obrázek 26: Rezidua odpovídajících si bodů pro racionální kvadratický model



Obrázek 27: Vykreslení hodnot  $Z^*$  pro racionální kvadratický model variogramu

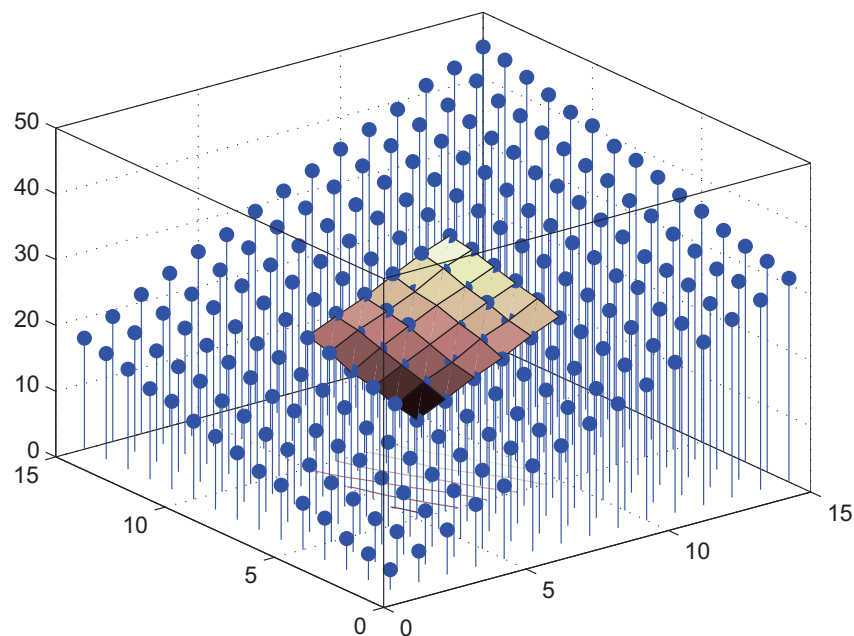
### 8.1.4 Exponenciální variogram

	1	2	3	4	5	6
1	15.052161679775885	17.008830213038415	19.16521852379555	20.962064783072794	23.143539613961625	24.917539641265364
2	15.749832501546267	18.143860116244163	19.91626347724071	22.3184739464181	23.93454778389705	26.120955699937905
3	17.02773044215954	19.1532531618797	21.1523665886325	23.237171103292425	25.166185642783045	26.88669152992535
4	17.71083497029249	20.226731665379365	22.317678611923487	24.37833091076223	26.037579545423338	27.857592305469844
5	19.25945127110689	21.392100045379543	23.07325439762796	24.917965346274453	27.3201702426867	28.901291035454317
6	20.129915724364903	22.00035151031004	23.68178386714377	25.884594452354257	27.972918658260415	29.56347012600913

Obrázek 28: Hodnoty  $Z^*$  získané pomocí exponenciálního modelu variogramu

	1	2	3	4	5	6
1	0.081197436187763	-0.046778152512054	0.028995925842402	0.391765878945652	0.02477558850034	-0.67214550740794
2	-0.056861590814398	0.390664583788784	-0.113900808766232	0.124498615202178	-0.251617531338532	-0.010766183749354
3	0.026619077832922	0.119333123486477	-0.098155891675852	0.187037996870966	-0.476866096983944	0.062161926830633
4	0.491658485840631	0.096639425887414	0.291314950999183	0.086245540956902	-0.211123293803258	-0.258911768645735
5	-0.069657977228918	-0.047221314548899	0.035947906235105	-0.130932591877734	-0.10045248985487	-0.516009904250733
6	0.440329529382417	0.133708766696753	-0.307897037907811	0.201187447982804	-0.124137684735899	-1.385558334732195e-013

Obrázek 29: Rezidua odpovídajících si bodů pro exponenciální model



Obrázek 30: Vykreslení hodnot  $Z^*$  pro exponenciální model variogramu

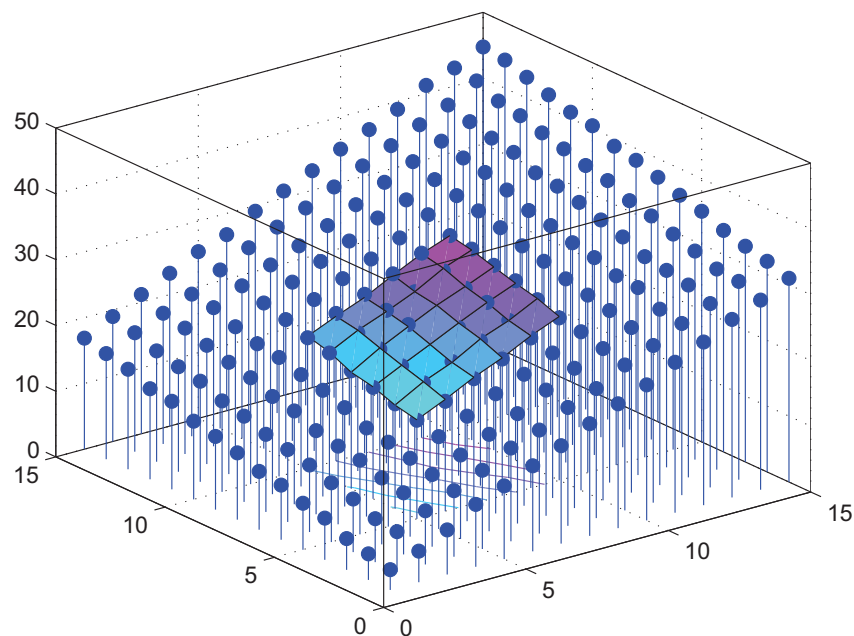
### 8.1.5 Mocninný variogram

	1	2	3	4	5	6
1	15.052161679775836	17.008830213038205	19.16521852379552	20.962064783071668	23.143539613961828	24.91753964126513
2	15.749832501546175	18.14386011624443	19.916263477240367	22.31847394641829	23.934547783897262	26.120955699938172
3	17.02773044215955	19.153253161879153	21.15236658883259	23.237171103293612	25.166185642783024	26.886691529925702
4	17.71083497029022	20.226731665379983	22.317678611924606	24.378330910761882	26.037579545424816	27.857592305470018
5	19.25945127110691	21.39210004537875	23.073254397628045	24.917965346275974	27.320170242686796	28.901291035454552
6	20.129915724364775	22.000351510310193	23.681783867143213	25.884594452355223	27.97291865826032	29.563470126009122

Obrázek 31: Hodnoty  $Z^*$  získané pomocí mocninného modelu variogramu

	1	2	3	4	5	6
1	0.081197436187765	-0.046778152511745	0.028995925842413	0.391765878946782	0.024775588500152	-0.672145507407688
2	-0.056861590814369	0.390664583788528	-0.11390080876588	0.124498615201933	-0.25161753133877	-0.010766183749606
3	0.026619077832862	0.11933312348717	-0.098155891675972	0.187037996869854	-0.476866096983912	0.06216192683025
4	0.49165848584272	0.096639425886728	0.291314950997975	0.086245540957222	-0.211123293804516	-0.25891176864604
5	-0.069657977228985	-0.047221314548068	0.036947906235069	-0.130932591879123	-0.100452489854945	-0.516009904250971
6	0.440329529382421	0.133708766696579	-0.307897037907207	0.201187447981923	-0.12413768473586	-1.527666881884215e-013

Obrázek 32: Rezidua odpovídajících si bodů pro mocninný model



Obrázek 33: Vykreslení hodnot  $Z^*$  pro mocninný model variogramu

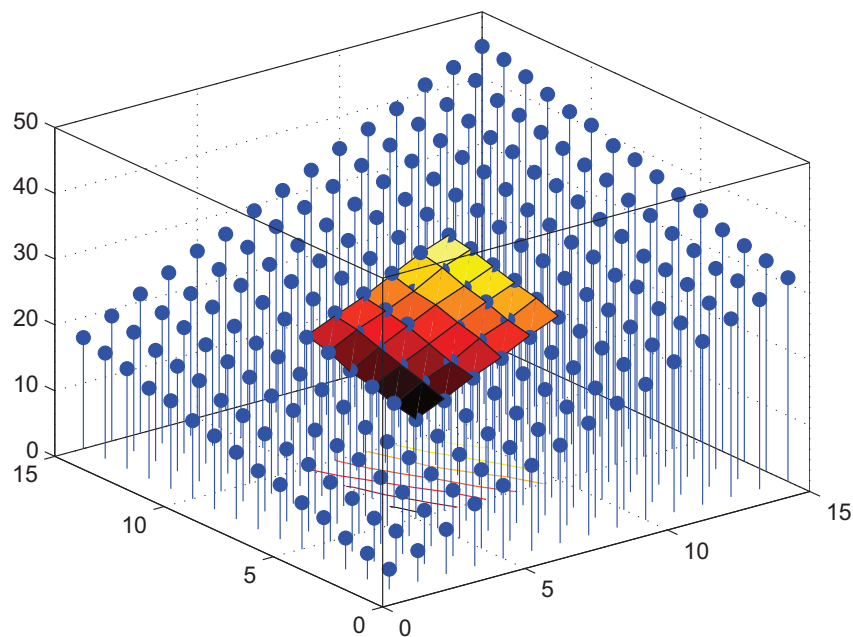
### 8.1.6 Vlnový variogram

	1	2	3	4	5	6
1	15.052161679775685	17.008830213037896	19.165218523795335	20.962064783072524	23.1435396139615	24.917539641265606
2	15.749832501544983	18.143860116244834	19.916263477243554	22.318473946418134	23.934547783898896	26.12095569993853
3	17.027730442159225	19.153253161883956	21.15236658883275	23.237171103292383	25.166185642783834	26.88669152992632
4	17.710834970292574	20.226731665379333	22.317678611923174	24.37833091076203	26.03757954542326	27.857592305469915
5	19.25945127110658	21.392100045385256	23.073254397628205	24.917965346274265	27.32017024268739	28.901291035455465
6	20.12991572436291	22.000351510310576	23.681783867147924	25.88459445235425	27.972918668262774	29.56347012600938

Obrázek 34: Hodnoty  $Z^*$  získané pomocí vlnového modelu variogramu

	1	2	3	4	5	6
1	0.081197436187919	-0.046778152511965	0.028995925842601	0.391765878945961	0.024775588600439	-0.672145507407681
2	-0.056861590813236	0.390664583788102	-0.11390080876907	0.124498615202132	-0.251617531340145	-0.010766183749961
3	0.026619077833256	0.119333123482313	-0.098155891675983	0.18703799687102	-0.476866096984882	0.062161926829575
4	0.491658485840521	0.096639425887467	0.291314950999946	0.086245540957094	-0.211123293803155	-0.258911768645788
5	-0.06965797722863	-0.047221314554179	0.03594790623476	-0.130932591877542	-0.100452489855279	-0.516009904251735
6	0.440329529384282	0.133708766696284	-0.307897037912159	0.201187447982811	-0.12413768473839	-4.334310688136611e-013

Obrázek 35: Rezidua odpovídajících si bodů pro vlnový model



Obrázek 36: Vykreslení hodnot  $Z^*$  pro vlnový model variogramu

## 8.2 Srovnání výsledků

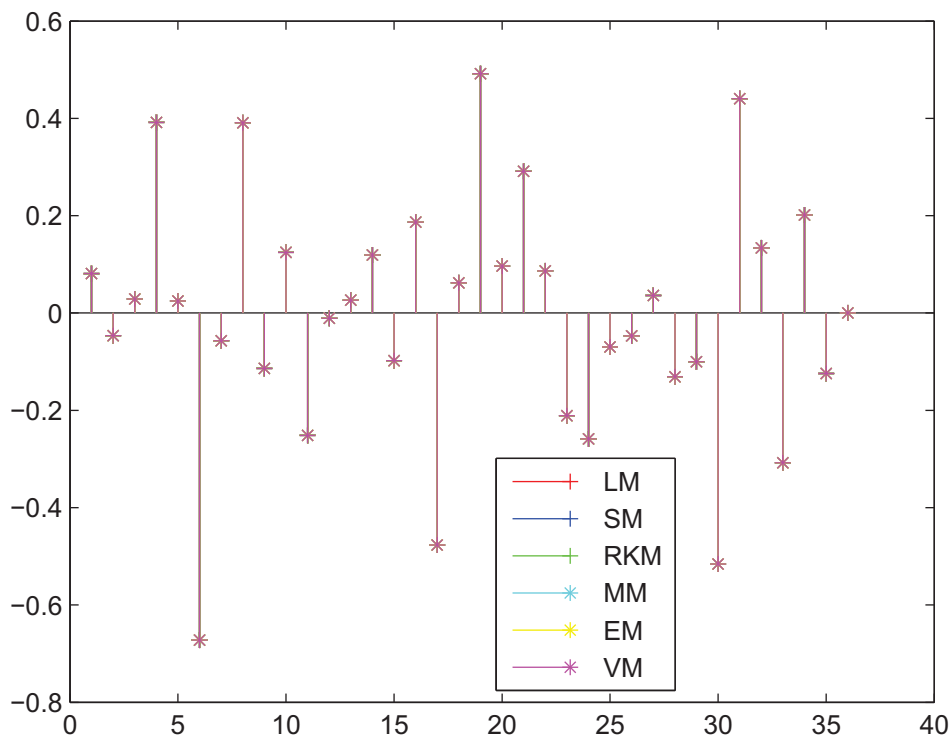
Vliv volby základního typu variogramu jsem chtěla zhodnotit pomocí reziduí, kritériem, které dostaneme vztahem

$$\sum_i r_i^2.$$

Výsledky tohoto výpočtu jsou na Obr.37. Jelikož tyto hodnoty jsou skoro stejné pro všechny typy modelů, nemohu říci, který model by byl nejvhodnější pro případ regresní roviny. Dále jsem chtěla znázornit jednotlivá  $r_i$  odpovídajících si bodů pro všechny základní typy modelových variogramů (viz. Obr.38), ale hodnoty reziduí jsou si tak blízké, že se vykreslené body překrývají i při velmi malém měřítku. Body jsou vykresleny barvou 'm', protože to byla poslední zadávaná barva.

Model	$\sum r_i^2$
Lineární model	2,273388853173621
Sférický model	2,273388853188223
Racionální kvadratický model	2,273388853179874
Mocninný model	2,273388853177787
Exponenciální model	2,27338885317607
Vlnový model	2,273388853183021

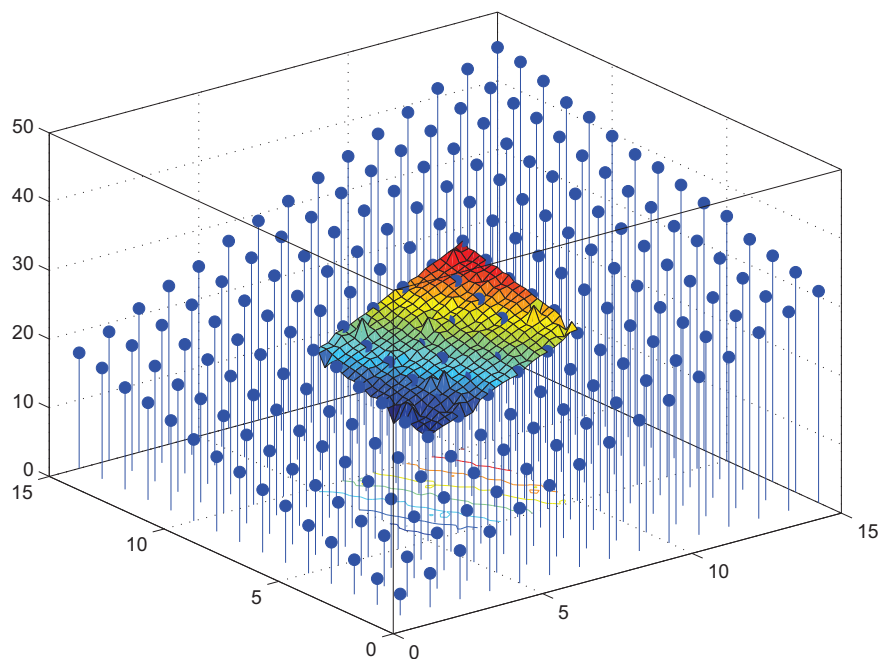
Obrázek 37: Vyhodnocení pomocí reziduí



Obrázek 38: Rezidua odpovídajících si bodů pro všechny základní modely

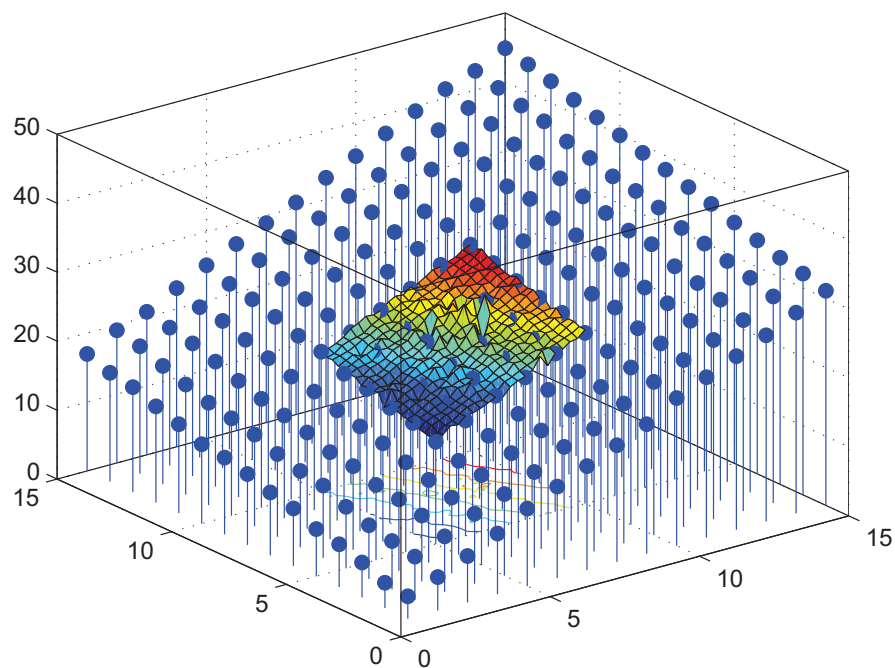
### 8.3 Zjemnění $x$ -ové a $y$ -ové škály měřítka souřadnic odhadu

Vektory, ve kterých se má odhadovat hodnota  $Z$ , lze zjemnit ( $x_0 = y_0 = 5 : 0.25 : 10$ ) a získat tak hodnoty  $Z^*$  v bodech, ve kterých nemáme  $Z$  pozorované. Vzniklé plochy jsou znázorněny na Obr.39 - Obr.44.

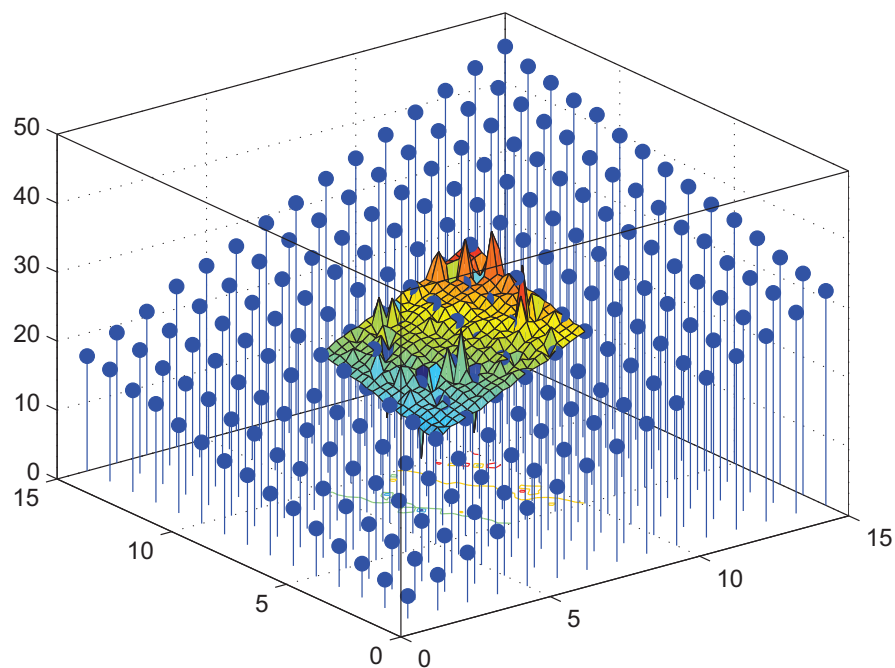


Obrázek 39: Hodnoty  $Z^*$  získané pomocí lineárního modelu variogramu pro zjemněné vektory  $x_0$  a  $y_0$

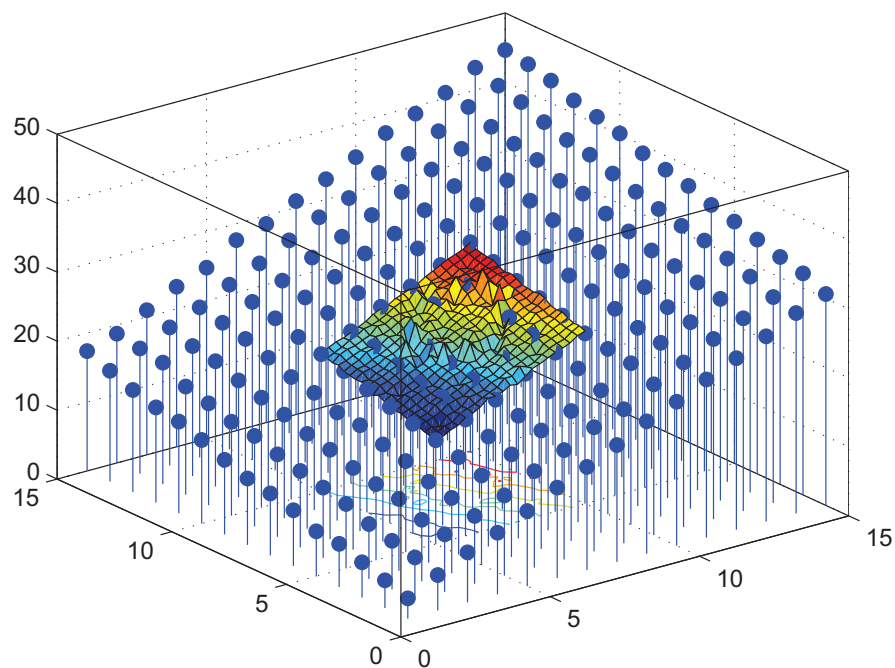




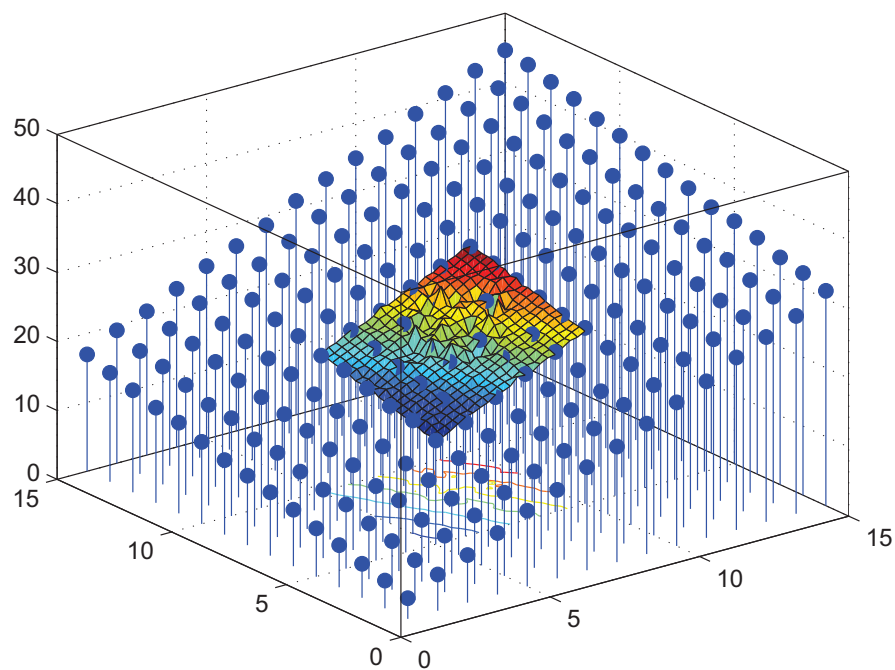
Obrázek 40: Hodnoty  $Z^*$  získané pomocí sférického modelu variogramu pro zjemněné vektory  $x_0$  a  $y_0$



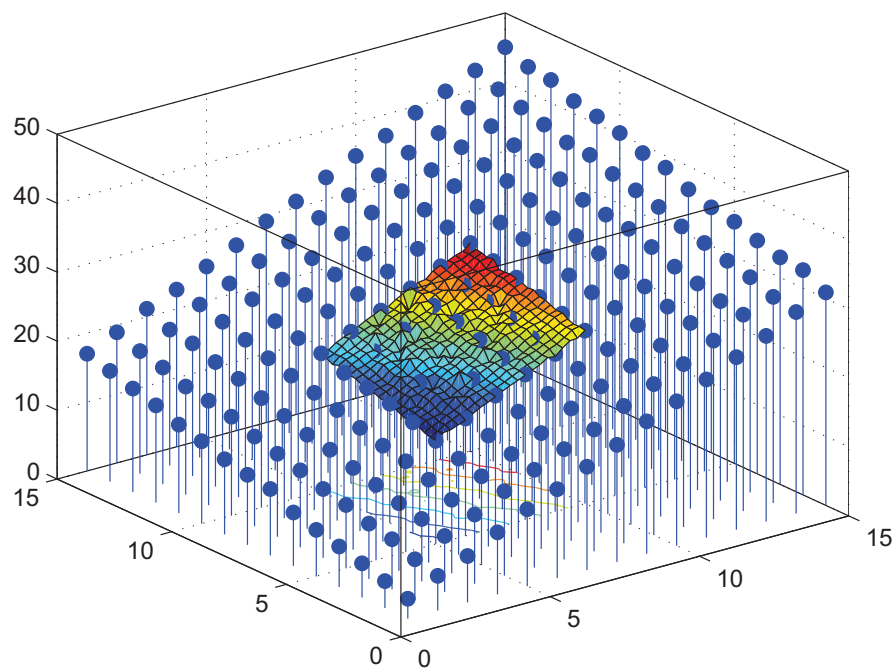
Obrázek 41: Hodnoty  $Z^*$  získané pomocí racionálního kvadratického modelu variogramu pro zjemněné vektory  $x_0$  a  $y_0$



Obrázek 42: Hodnoty  $Z^*$  získané pomocí exponenciálního modelu variogramu pro zjemněné vektory  $x_0$  a  $y_0$



Obrázek 43: Hodnoty  $Z^*$  získané pomocí mocninného modelu variogramu pro zjemněné vektory  $x_0$  a  $y_0$



Obrázek 44: Hodnoty  $Z^*$  získané pomocí vlnového modelu variogramu pro zjemněné vektory  $x_0$  a  $y_0$

Pomocí obrázků pro jemnější měřítko souřadnic odhadů můžeme také zhodnotit vliv výběru modelového variogramu. Na Obr.41 je viditelné, že racionální kvadratický model vykazuje největší vychýlení od očekávaných hodnot tohoto příkladu, zatímco jako nejlepší se jeví vlnový a lineární model (Obr.44 a Obr.39).

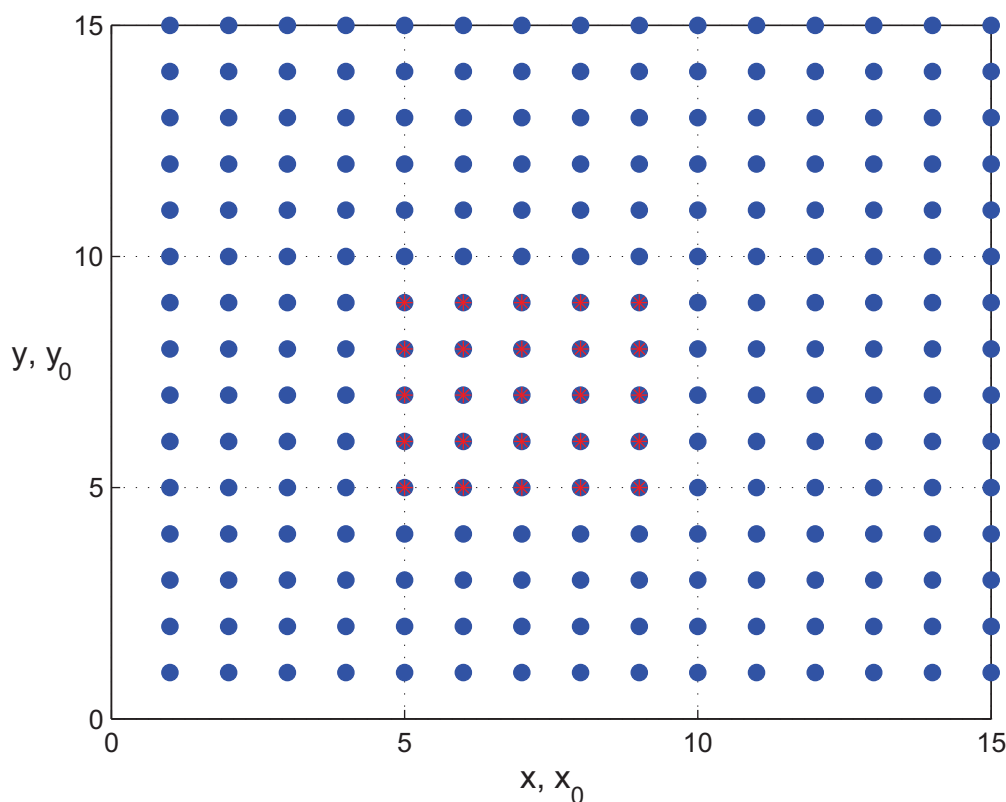
## 9 Kriging metoda pro obecnou lineární regresní plochu

Druhým příkladem je příklad obecné lineární regresní plochy. Opět zavoláme funkce "ZadaniVstupnichDat" a "Mezikruzi" a pomocí funkce "KlasickyOdhadVariogramu" dostaneme klasický odhad variogramu v jednom bodě a pomocí tohoto odhadu určíme parametry základních typů modelových variogramů.

Odhadnuté parametry mají v tomto příkladě hodnoty:

- Lineární model:  $c_0 = 0.01$ ,  $b = 0.025$ ;
- Sférický model:  $c_0 = 0.1$ ,  $c_s = 0.07$  a  $a_s = 4$ ;
- Exponenciální model:  $c_0 = 0.05$ ,  $c_e = 0.17$  a  $a_e = 2$ ;
- Racionální kvadratický model:  $c_0 = 0.015$ ,  $c_r = 0.05$  a  $a_r = 4$ ;
- Mocninný model:  $c_0 = 0.012$ ,  $b_m = 0.12$  a  $\lambda = 0.1$ ;
- Vlnový model:  $c_0 = 0.01$ ,  $c_w = 0.2$  a  $a_w = 2$ .

Pro obecnou lineární regresní plochu a pro jednotlivé základní typy modelů provedeme odhad v 25-ti bodech, které leží uvnitř pozorované oblasti, jak je znázorněno na Obr.45. Ve funkci "CelkovyOdhad" tedy rozšíříme  $x_0$  a  $y_0$  na vektory.



Obrázek 45: Znázornění bodů, ve kterých známe hodnoty procesu  $Z$  a bodů, ve kterých budeme odhadovat  $Z^*$

Body, ve kterých budeme odhadovat  $Z^*$  jsou úmyslně zvoleny tak, aby k nim existovala i hodnota pozorovaná. Potom opět pomocí vztahu  $r(s_i) = Z(s_i) - Z^*(s_i)$ , kde  $s_i = [x(i), y(i)]$ , vypočítáme rezidua jednotlivých bodů pro různé základní typy modelových variogramů a pokusíme se porovnat jednotlivé modely právě pomocí těchto reziduí. Odhady  $Z^*$  provádíme pro  $x = y = 5 : 1 : 9$ . Na Obr.46 jsou připomenuty hodnoty  $ZS$  pro zvolenou plochu a je zvýrazněna oblast bodů, ke kterým odhadujeme  $Z^*$ .

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	3.6611	5.9459	7.0475	8.8986	9.9898	11.1819	11.3231	11.7319	11.4093	10.8282	9.8252	8.2332	6.4989	4.6793	2.9325
2	0.7274	2.5477	4.4092	6.1768	7.118	8.4683	8.7365	8.6744	8.0594	7.9055	6.697	5.3909	4.0329	1.9499	0.4485
3	-0.9518	0.9549	2.5029	4.334	5.4663	6.4494	7.0092	7.4076	6.9031	6.057	4.981	4.0711	2.4484	0.1358	-1.7816
4	0.0674	2.155	3.7854	5.2162	6.5618	7.476	7.371	8.159	7.6047	6.8584	5.9608	5.2045	3.133	1.3742	-0.3597
5	3.1257	4.6761	6.622	7.7984	9.3877	9.9904	10.7867	10.5887	10.9292	10.0419	9.0094	7.7781	6.1593	3.8644	2.0092
6	4.5936	6.9133	8.2575	10.3915	11.4304	12.016	12.5828	13.1984	12.6916	11.4737	10.9696	9.6756	7.7909	6.3813	4.1016
7	3.7721	6.1338	8.2296	9.5339	10.3674	11.6263	12.0143	12.3587	12.14	11.1159	10.6181	9.1209	7.3835	5.6595	3.4704
8	1.4592	3.4143	5.3413	6.8311	7.8967	9.0375	9.4266	9.5048	9.4575	8.6655	7.7225	6.5594	4.8205	2.9622	1.0138
9	-0.7955	1.1967	2.9951	4.5476	5.8819	6.6509	6.7235	7.2539	6.9032	6.6166	5.4694	4.1023	2.8056	0.7393	-1.2288
10	-1.0364	1.1775	3.2272	4.589	5.7906	6.7646	7.1619	7.3205	7.2633	6.1444	5.2018	3.9835	2.6337	0.8041	-1.1461
11	1.9165	4.1348	5.5859	7.2015	8.2693	9.2405	9.6932	9.8356	9.487	9.0354	8.0602	6.5878	5.2076	3.3857	1.0897
12	4.4628	6.3836	8.14	9.4289	10.836	11.9464	12.6383	12.2505	11.5191	10.4875	8.848	7.6223	6.1063	4.1771	
13	4.9381	6.6743	8.5743	10.1614	11.3226	12.2053	12.4175	12.3754	12.5537	11.7639	10.5261	9.6123	7.6824	5.9225	4.3278
14	2.4675	4.3391	6.06	7.5941	8.7902	9.9273	10.0753	10.236	10.1276	9.3528	8.4679	7.2392	5.6777	3.6059	1.7771
15	-0.4917	1.4384	3.431	4.784	6.0703	7.0423	7.6238	8.0793	7.2679	6.6638	5.926	4.6313	3.0048	1.2654	-1.1447

Obrázek 46: Znázornění hodnot  $Z$ , které budeme srovnávat s odhadnutými hodnotami

## 9.1 Kriging odhady

Nyní upravíme podkapitoly oproti předchozí kapitole, abychom mohli lépe porovnat jednotlivé modely.

### 9.1.1 Odhady $Z^*$ regresní funkce

Uvedeme tabulky hodnot  $Z^*$  všech 6-ti typů modelových variogramů. Hodnoty  $Z^*$  jsou si pro odpovídající body opět hodně blízké.

	1	2	3	4	5
1	9.448894541983101	10.136863231626101	10.572348565435917	11.165470093105988	10.604180270815183
2	11.614516916189684	12.017355469846134	12.67501042665558	12.650147018979329	12.520232078861708
3	10.48431171771136	11.398819027186388	11.938946966535113	12.403659474950866	11.870499504147901
4	8.03508704307649	9.026152198657357	9.220573953264404	9.705705498790904	9.180796717505478
5	5.775250061129921	6.179201364825718	7.237520749033342	7.513787284700178	6.903180030627249

Obrázek 47: Hodnoty  $Z^*$  získané pomocí lineárního modelu variogramu

	1	2	3	4	5
1	9.448894541983606	10.136863231626297	10.572348565435894	11.16547009311382	10.604180270815107
2	11.614516916189872	12.017355469843235	12.675010426655597	12.650147018978108	12.52023207886172
3	10.48431171771139	11.398819027186388	11.938946966534923	12.403659474950029	11.870499504147707
4	8.035087043085548	9.026152198556385	9.22057395326356	9.70570549879142	9.180796717504752
5	5.775250061129975	6.179201364825702	7.237520749033166	7.513787284699539	6.903180030627118

Obrázek 48: Hodnoty  $Z^*$  získané pomocí sférického modelu variogramu

	1	2	3	4	5
1	9.448894541983053	10.136863231626176	10.57234856543607	11.16547009310956	10.604180270815304
2	11.614516916189732	12.017355469843265	12.675010426655478	12.650147018978526	12.520232078861552
3	10.484311717711591	11.39881902718632	11.938946966535106	12.403659474954493	11.8704995041479
4	8.035087043080454	9.026152198556838	9.220573953267266	9.705705498791032	9.180796717510754
5	5.775250061130206	6.179201364825621	7.237520749033338	7.513787284704979	6.903180030627282

Obrázek 49: Hodnoty  $Z^*$  získané pomocí racionálního kvadratického modelu variogramu

	1	2	3	4	5
1	9.448894541982868	10.136863231616543	10.57234856543599	11.165470093109775	10.604180270815197
2	11.6145169161797	12.01735546984338	12.675010426654548	12.650147018978167	12.52023207886077
3	10.484311717711492	11.398819027185656	11.938946966533484	12.40365947495027	11.870499504146574
4	8.035087043080635	9.026152198556414	9.220573953263912	9.705705498809694	9.180796717504974
5	5.775250061130073	6.179201364824752	7.237520749031888	7.513787284699874	6.90318003062756

Obrázek 50: Hodnoty  $Z^*$  získané pomocí exponenciálního modelu variogramu

	1	2	3	4	5
1	9.448894541982948	10.13686323162594	10.572348565438418	11.165470093109729	10.604180270818496
2	11.614516916189524	12.017355469832982	12.67501042665538	12.65014701897855	12.52023207886144
3	10.484311717714428	11.398819027186168	11.938946966534923	12.403659474944165	11.87049950414766
4	8.035087043080646	9.02615219855677	9.220573953259725	9.70570549879098	9.180796717499188
5	5.77525006113321	6.179201364825514	7.237520749032876	7.513787284696522	6.903180030626842

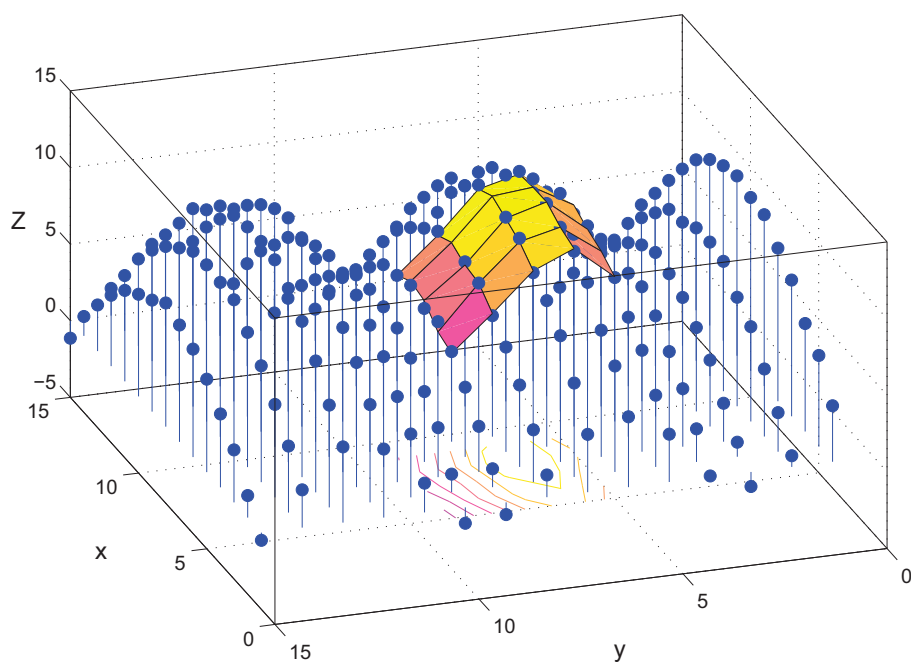
Obrázek 51: Hodnoty  $Z^*$  získané pomocí mocinného modelu variogramu

	1	2	3	4	5
1	9.44889454198351	10.136863231625425	10.572348565436343	11.165470093108588	10.604180270815126
2	11.614516916189142	12.0173554698447	12.675010426655446	12.650147018978132	12.520232078861671
3	10.484311717711995	11.398819027186128	11.938946966535793	12.403659474950086	11.870499504152017
4	8.035087043078843	9.026152198557542	9.22057395326366	9.705705498791186	9.180796717504844
5	5.775250061130195	6.17920136482546	7.23752074902918	7.513787284699633	6.903180030630062

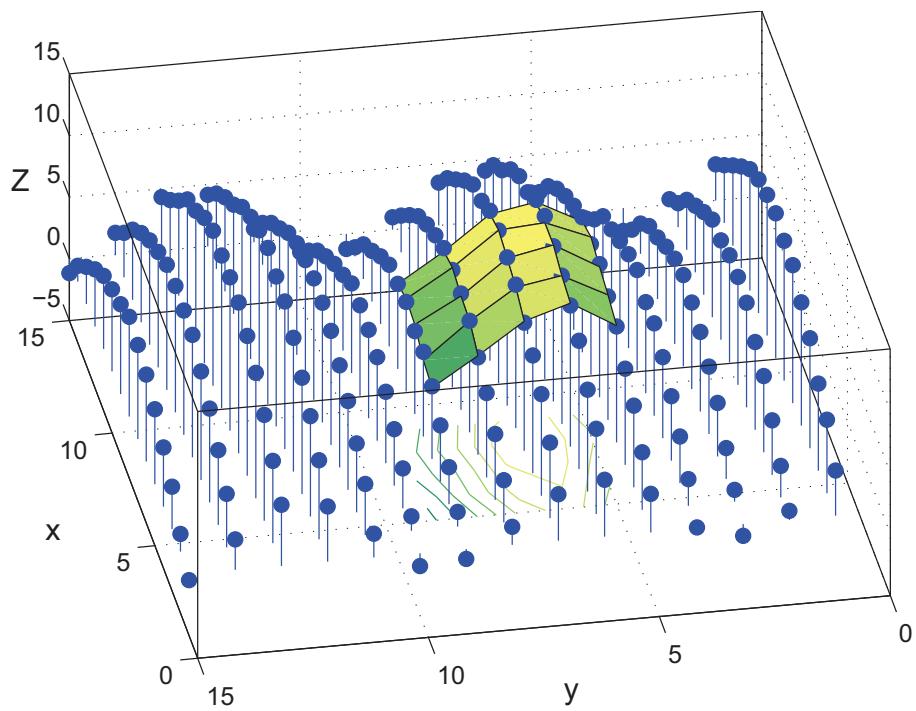
Obrázek 52: Hodnoty  $Z^*$  získané pomocí vlnového modelu variogramu

### 9.1.2 Grafické znázornění výsledků

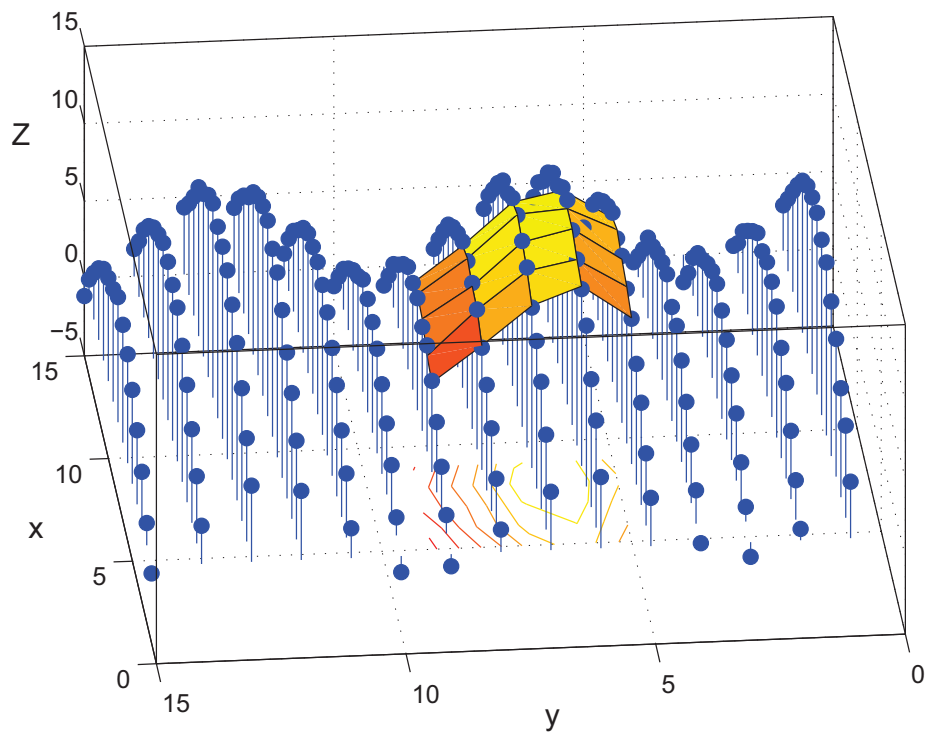
Vypočítané hodnoty jsem opět pro lepší znázornění proložila plochou a navíc bylo nutné zvolit vhodný úhel náhledu.



Obrázek 53: Vykreslení hodnot  $Z^*$  pro lineární model variogramu

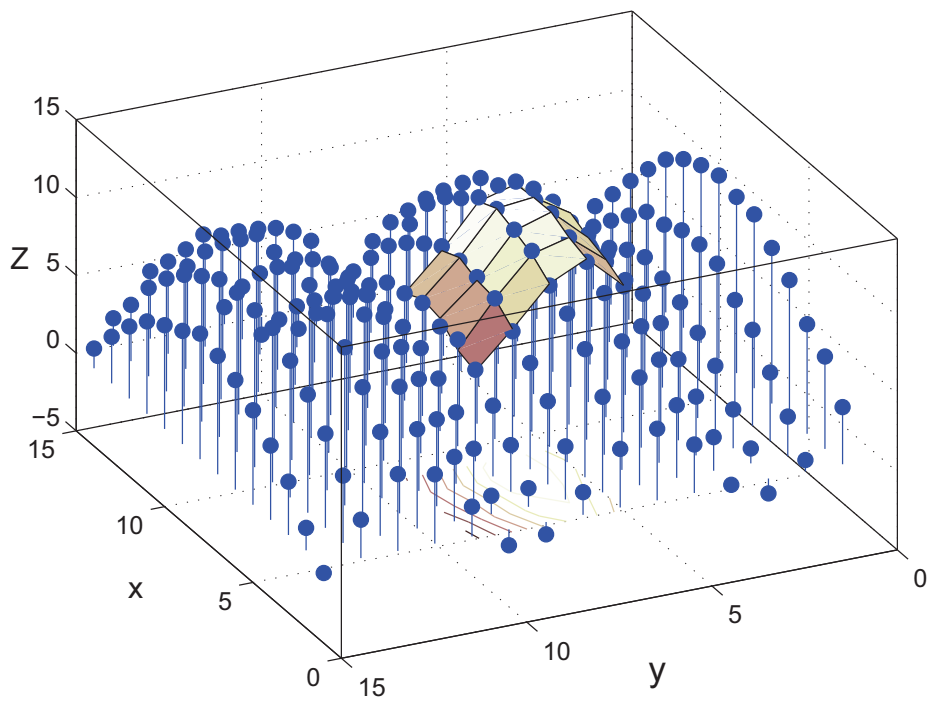


Obrázek 54: Vykreslení hodnot  $Z^*$  pro sférický model variogramu

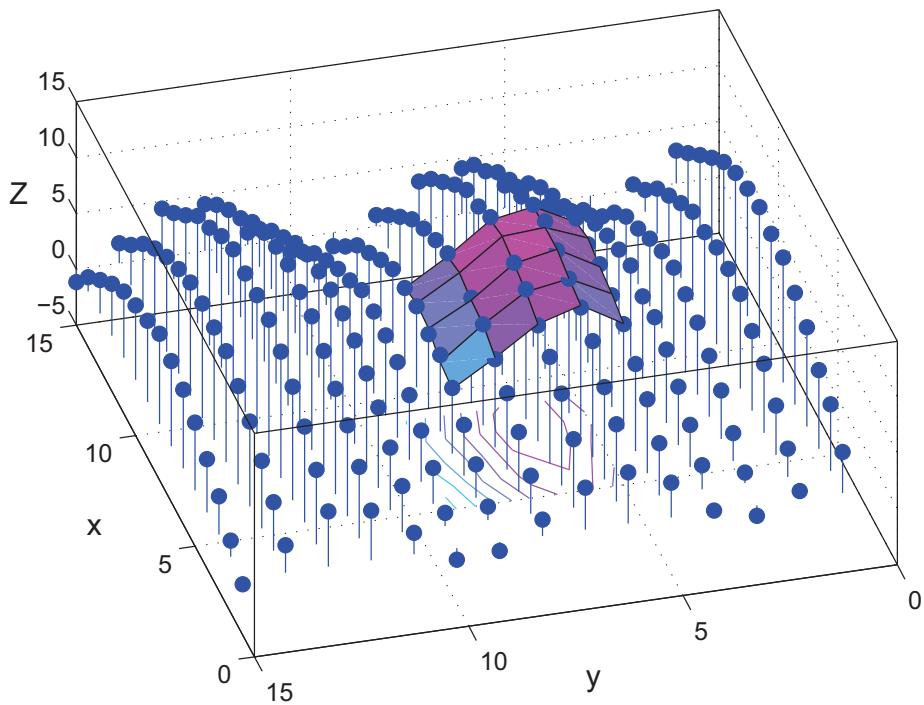


Obrázek 55: Vykreslení hodnot  $Z^*$  pro racionální kvadratický model variogramu

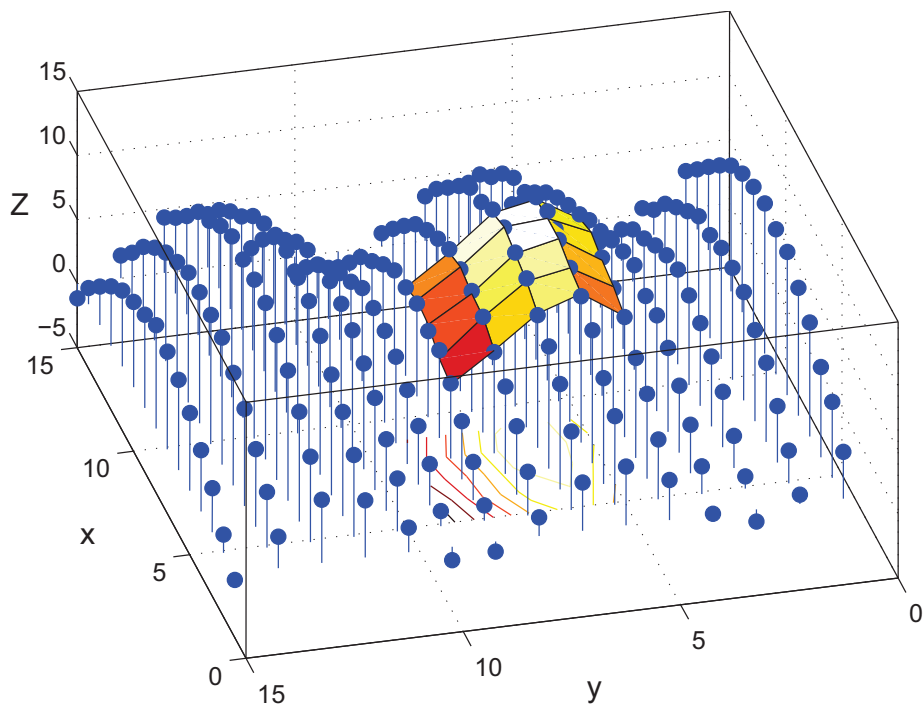




Obrázek 56: Vykreslení hodnot  $Z^*$  pro exponenciální model variogramu



Obrázek 57: Vykreslení hodnot  $Z^*$  pro mocninný model variogramu



Obrázek 58: Vykreslení hodnot  $Z^*$  pro vlnový model variogramu

### 9.1.3 Reziduální analýza

Jak je vidět na Obr.59 - Obr.64, hodnoty reziduí jsou si opět velmi blízké a opět se nedaří vykreslení jednotlivých reziduí pro všechny základní typy modelových variogramů, protože se body překrývají. U obecné lineární regresní funkce nelze stejně jako u příkladu regresní roviny říci, který základní typ modelového variogramu je pro tento příklad nejvhodnější.

	1	2	3	4	5
1	-0.061180221683454	-0.146417082485536	0.214323962728514	-0.576765432731012	0.324998012498613
2	-0.184119095070136	-0.001393778411485	-0.092164530125451	0.548245033345914	0.171356036171851
3	-0.116870243235178	0.227502438898817	0.075389942864694	-0.044955374575467	0.269501079169336
4	-0.138377169700449	0.011334618166741	0.208002395857427	-0.201065811751104	0.276723142715097
5	0.10663332709396	0.671696297316128	-0.514003657631172	-0.25993509471366	-9.50350909079134e-014

Obrázek 59: Rezidua odpovídajících si bodů pro lineární model

	1	2	3	4	5
1	-0.061180221683959	-0.146417082485732	0.214323962728537	-0.576765432738844	0.324998012498689
2	-0.184119095070324	-0.001393778408586	-0.092164530125467	0.548245033347134	0.171356036171838
3	-0.116870243235208	0.227502438898817	0.075389942864884	-0.044955374574631	0.26950107916953
4	-0.138377169709507	0.011334618167712	0.208002395858271	-0.20106581175162	0.276723142715824
5	0.106633327093907	0.671696297316144	-0.514003657630996	-0.259935094713022	3.641531520770514e-014

Obrázek 60: Rezidua odpovídajících si bodů pro sférický model

	1	2	3	4	5
1	-0.061180221683406	-0.146417082485611	0.214323962728361	-0.576765432734584	0.324998012498492
2	-0.184119095070184	-0.001393778408616	-0.092164530125348	0.548245033346717	0.171356036172007
3	-0.116870243235409	0.227502438898885	0.075389942864701	-0.044955374579095	0.269501079169338
4	-0.138377169704412	0.011334618167259	0.208002395854566	-0.201065811751231	0.276723142709821
5	0.106633327093676	0.671696297316225	-0.514003657631169	-0.259935094718462	-1.27897692436818e-013

Obrázek 61: Rezidua odpovídajících si bodů pro racionální kvadratický model

	1	2	3	4	5
1	-0.061180221683221	-0.146417082475978	0.214323962728441	-0.576765432734799	0.324998012498598
2	-0.184119095060151	-0.001393778408731	-0.092164530124419	0.548245033347076	0.171356036172789
3	-0.11687024323531	0.227502438899549	0.075389942866323	-0.044955374574872	0.269501079170663
4	-0.138377169704594	0.011334618167684	0.208002395857919	-0.201065811769894	0.276723142715602
5	0.106633327093808	0.671696297317094	-0.514003657629718	-0.259935094713357	-4.058975378029572e-013

Obrázek 62: Rezidua odpovídajících si bodů pro exponenciální model

	1	2	3	4	5
1	-0.061180221683301	-0.146417082485375	0.214323962726013	-0.576765432734753	0.3249980124953
2	-0.184119095069976	-0.001393778398333	-0.092164530125251	0.548245033346692	0.171356036172119
3	-0.116870243238246	0.227502438899037	0.075389942864884	-0.044955374568767	0.269501079169578
4	-0.138377169704604	0.011334618167327	0.208002395862106	-0.20106581175118	0.276723142721387
5	0.106633327090671	0.671696297316332	-0.514003657630706	-0.259935094710005	3.11750625314744e-013

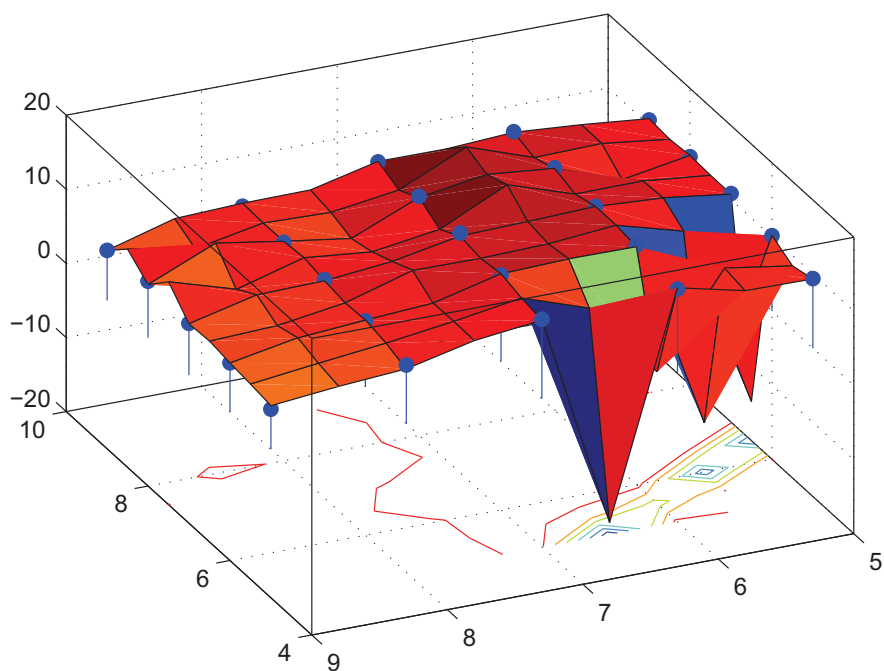
Obrázek 63: Rezidua odpovídajících si bodů pro mocninový model

	1	2	3	4	5
1	-0.061180221683863	-0.146417082484859	0.214323962728088	-0.576765432733612	0.32499801249867
2	-0.184119095069594	-0.001393778410051	-0.092164530125316	0.548245033347111	0.171356036171888
3	-0.116870243235812	0.227502438899076	0.075389942864014	-0.044955374574688	0.26950107916522
4	-0.138377169702801	0.011334618166556	0.208002395858172	-0.201065811751386	0.276723142715731
5	0.106633327093687	0.671696297316386	-0.51400365762701	-0.259936094713116	-2.90789614609821e-012

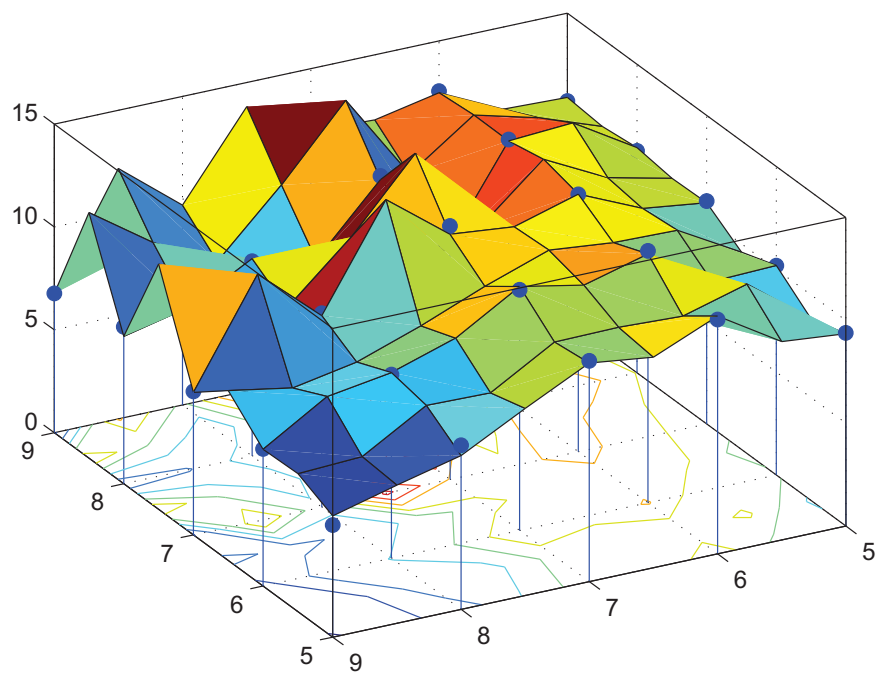
Obrázek 64: Rezipua odpovídajících si bodů pro vlnový model

## 9.2 Zjemnění měřítek vektorů $x_0$ a $y_0$

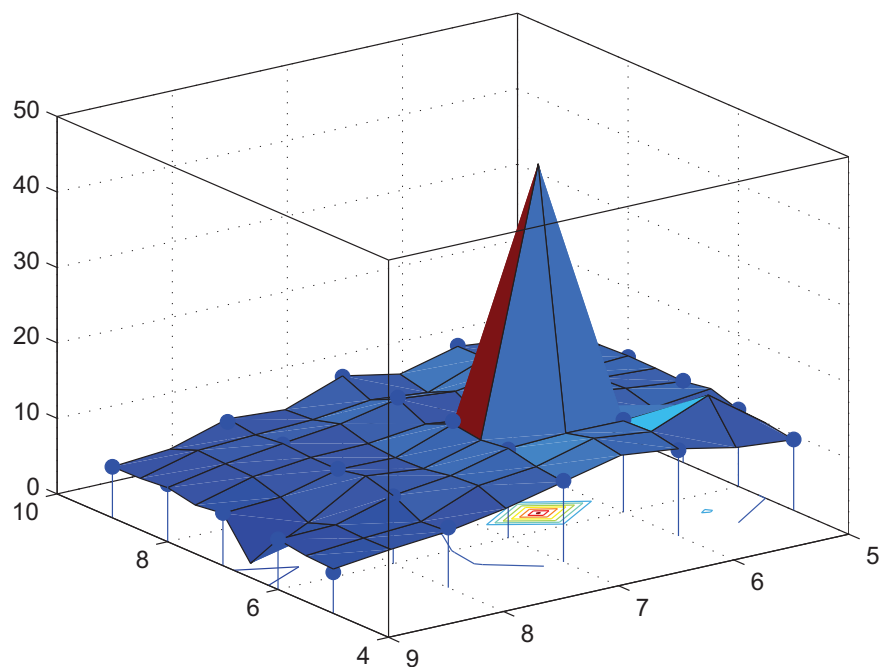
Vektory  $x_0$  a  $y_0$  zjemníme na  $x_0 = y_0 = 5 : 0.5 : 9$  a pokusíme se posoudit vliv volby modelu variogramu posoudit pro případ zjemněných vektorů.



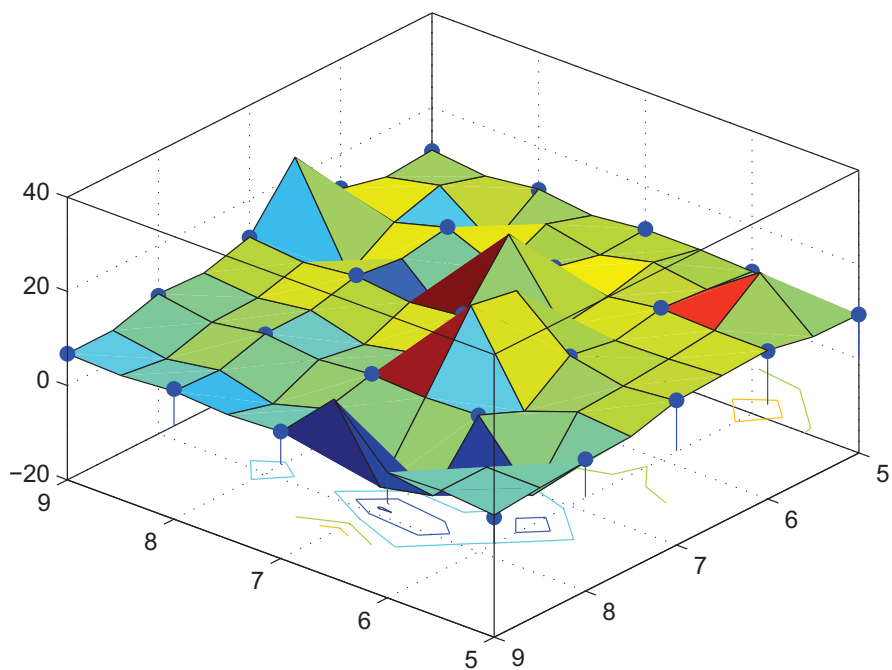
Obrázek 65: Vykreslená plocha odhadů získaných pomocí lineárního modelu pro zjemněné vektory  $x_0$  a  $y_0$



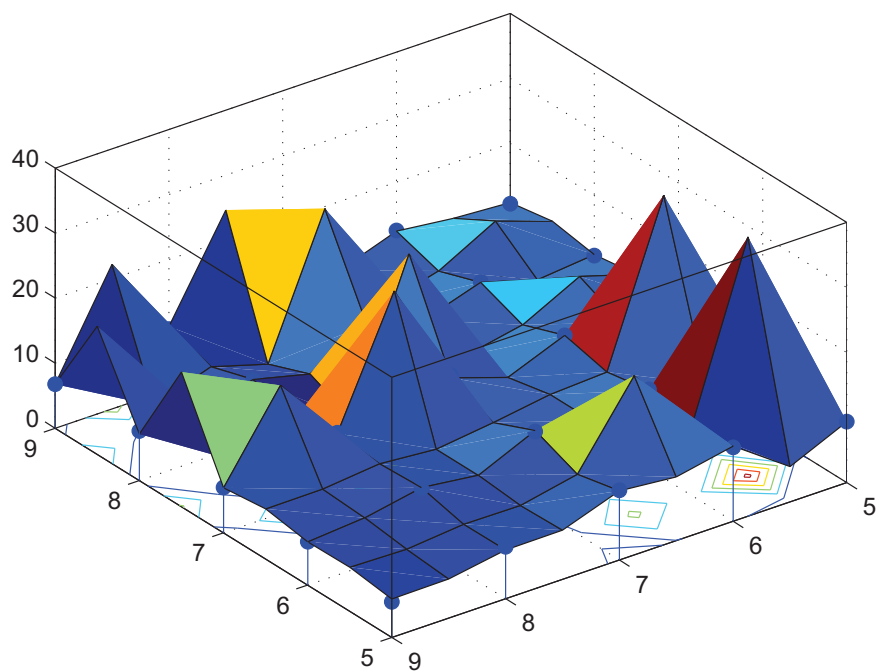
Obrázek 66: Vykreslená plocha odhadů získaných pomocí sférického modelu pro zjemněné vektory  $x_0$  a  $y_0$



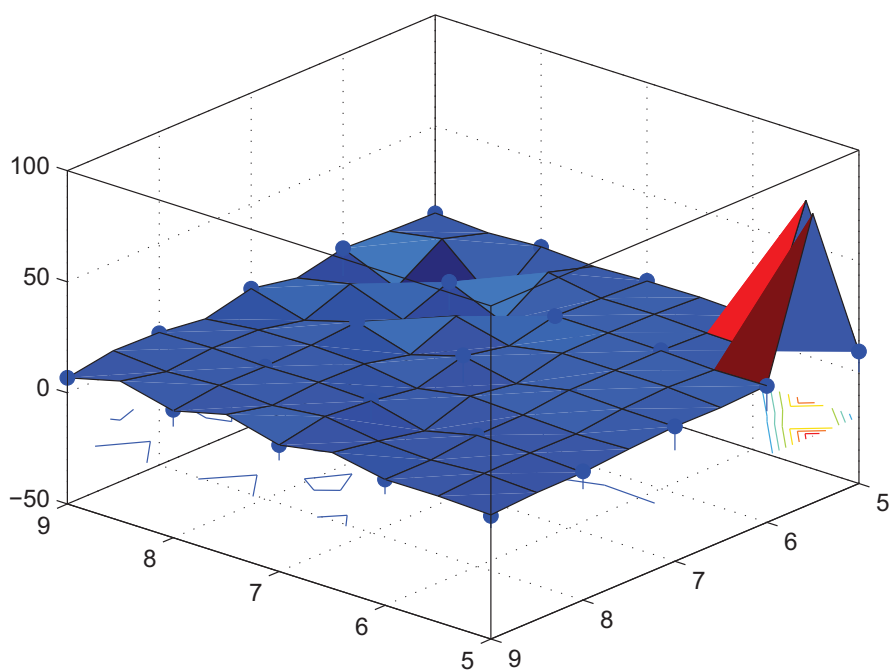
Obrázek 67: Vykreslená plocha odhadů získaných pomocí racionálního kvadratického modelu pro zjemněné vektory  $x_0$  a  $y_0$



Obrázek 68: Vykreslená plocha odhadů získaných pomocí exponenciálního modelu pro zjemněné vektory  $x_0$  a  $y_0$



Obrázek 69: Vykreslená plocha odhadů získaných pomocí mocinného modelu pro zjemněné vektory  $x_0$  a  $y_0$



Obrázek 70: Vykreslená plocha odhadů získaných pomocí vlnového modelu pro zjemněné vektory  $x_0$  a  $y_0$

Jak vidíme (Obr.65 - Obr.70), pro obecnou lineární regresní plochu nejsou odhady v nepozorovaných bodech dobré a vykazují velké rozptyly. Nejlepšími se zdají být sférický a exponenciální model.

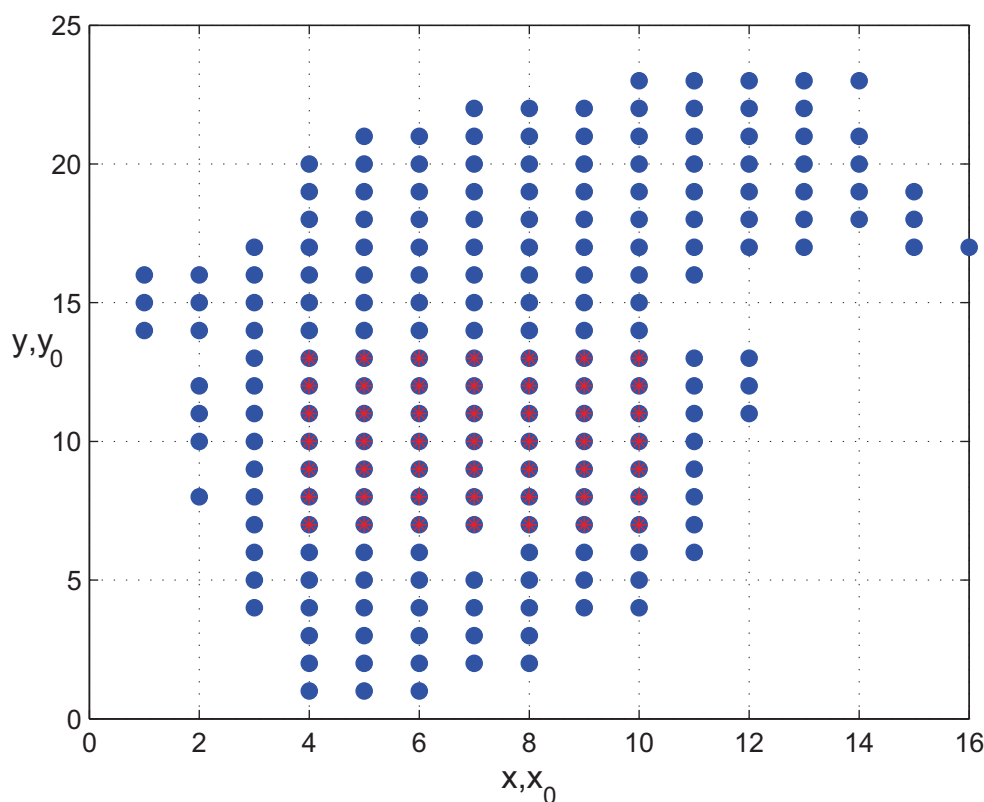
## 10 Reálný příklad pro Coal-Ash Data

Posledním příkladem je příklad dat pro uhelný popel v Pensylvánii. Zavoláme funkce "CoalAshData" a "Mezikruží" a pomocí funkce "KlasickyOdhadVariogramu" dostaneme klasický odhad variogramu v jednom bodě, pomocí něhož odhadneme parametry základních typů modelových variogramů.

Odhadnuté parametry mají v tomto příkladě hodnoty:

- Lineární model:  $c_0 = 0.05$ ,  $b = 0.002$ ;
- Sférický model:  $c_0 = 0.01$ ,  $c_s = 0.04$  a  $a_s = 5$ ;
- Racionální kvadratický model:  $c_0 = 0.015$ ,  $c_r = 0.5$  a  $a_r = 0.1$ ;
- Mocninný model:  $c_0 = 0.003$ ,  $b_m = 0.002$  a  $\lambda = 1.3$ ;
- Exponenciální model:  $c_0 = 0.05$ ,  $c_e = 0.02$  a  $a_e = 5$ ;
- Vlnový model:  $c_0 = 0.04$ ,  $c_w = 0.6$  a  $a_w = 2$ .

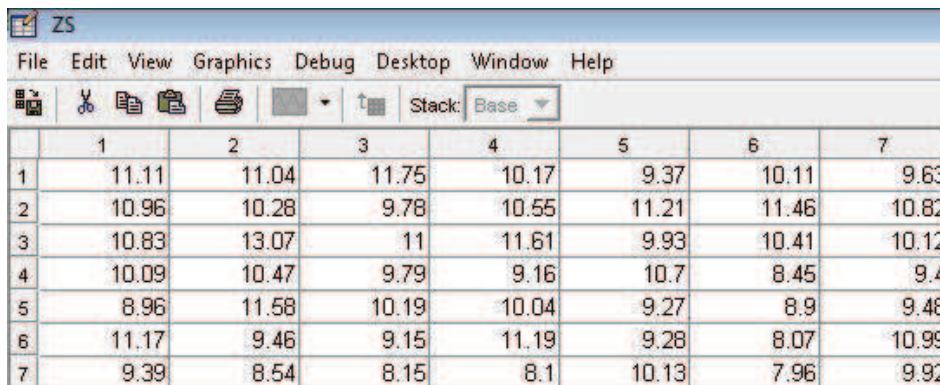
Pro tyto data a pro jednotlivé základní typy modelů provedeme odhad v 49-ti bodech, které leží uvnitř pozorované oblasti, jak je znázorněno na Obr.(71). Ve funkci "Celkovy-Odhad" tedy rozšíříme  $x_0$  a  $y_0$  na vektory.



Obrázek 71: Znázornění bodů, ve kterých známe hodnoty procesu  $Z$  a bodů, ve kterých budeme odhadovat  $Z^*$



K bodům, ve kterých budeme odhadovat  $Z^*$ , opět existuje i hodnota pozorovaná. Opět budeme vyšetřovat stav  $r(s_i) = Z(s_i) - Z^*(s_i)$ , kde  $s_i = [x(i), y(i)]$ , tedy rezidua jednotlivých bodů pro různé základní typy modelových variogramů a pokusíme se porovnat jednotlivé modely právě pomocí těchto reziduí. Odhady  $Z^*$  provádíme pro  $x = 4 : 1 : 10, y = 7 : 1 : 13$ . Na Obr.(72) jsou připomenuty hodnoty  $ZS$  pro body, ve kterých budeme predikovat  $Z^*$ .



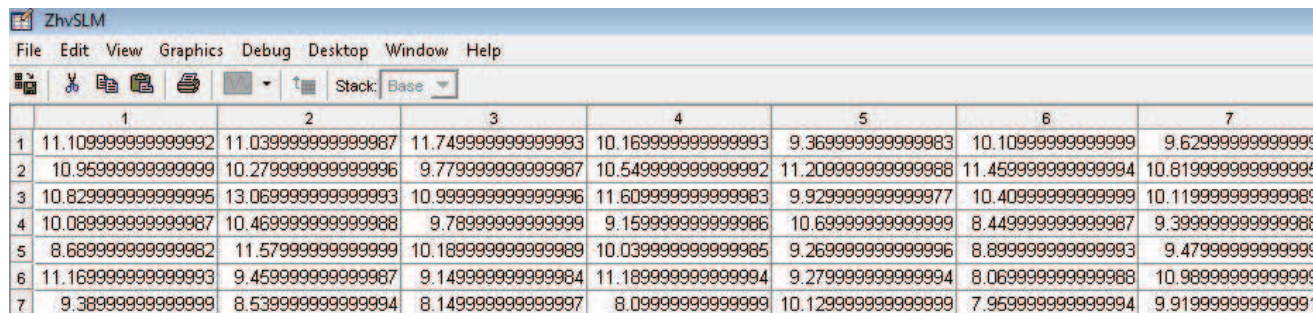
	1	2	3	4	5	6	7
1	11.11	11.04	11.75	10.17	9.37	10.11	9.63
2	10.96	10.28	9.78	10.55	11.21	11.46	10.82
3	10.83	13.07	11	11.61	9.93	10.41	10.12
4	10.09	10.47	9.79	9.16	10.7	8.45	9.4
5	8.96	11.58	10.19	10.04	9.27	8.9	9.48
6	11.17	9.46	9.15	11.19	9.28	8.07	10.99
7	9.39	8.54	8.15	8.1	10.13	7.96	9.92

Obrázek 72: Znázornění hodnot  $Z$ , které budeme srovnávat s odhadnutými hodnotami

## 10.1 Kriging odhady

### 10.1.1 Odhady regresní funkce

Uvedeme tabulky hodnot  $Z^*$  všech 6-ti typů modelových variogramů. Hodnoty  $Z^*$  jsou si pro odpovídající body opět hodně blízké.



	1	2	3	4	5	6	7
1	11.109999999999992	11.039999999999987	11.749999999999993	10.169999999999993	9.369999999999983	10.109999999999999	9.629999999999999
2	10.959999999999999	10.279999999999996	9.779999999999987	10.549999999999992	11.209999999999988	11.459999999999994	10.819999999999995
3	10.829999999999995	13.069999999999993	10.999999999999996	11.609999999999983	9.929999999999977	10.409999999999999	10.119999999999989
4	10.089999999999987	10.469999999999988	9.789999999999999	9.159999999999986	10.699999999999999	8.449999999999987	9.399999999999988
5	8.689999999999982	11.579999999999999	10.189999999999989	10.039999999999985	9.269999999999996	8.899999999999993	9.479999999999999
6	11.169999999999993	9.459999999999987	9.149999999999984	11.189999999999994	9.279999999999994	8.069999999999988	10.969999999999999
7	9.389999999999999	8.539999999999994	8.149999999999997	8.099999999999999	10.129999999999999	7.969999999999994	9.919999999999991

Obrázek 73: Hodnoty  $Z^*$  získané pomocí lineárního modelu variogramu

	1	2	3	4	5	6	7
1	11.109999999999989	11.039999999999997	11.749999999999986	10.170000000000007	9.37	10.109999999999989	9.629999999999983
2	10.959999999999992	10.279999999999998	9.779999999999992	10.549999999999988	11.209999999999987	11.459999999999994	10.82
3	10.829999999999993	13.069999999999993	10.999999999999998	11.609999999999999	9.929999999999987	10.409999999999993	10.119999999999992
4	10.089999999999999	10.469999999999999	9.789999999999988	9.159999999999979	10.699999999999982	8.449999999999999	9.399999999999995
5	8.689999999999998	11.579999999999999	10.189999999999996	10.039999999999999	9.269999999999994	8.899999999999983	9.479999999999984
6	11.169999999999987	9.459999999999985	9.149999999999999	11.189999999999994	9.279999999999983	8.069999999999983	10.989999999999972
7	9.389999999999981	8.539999999999987	8.149999999999979	8.099999999999996	10.129999999999988	7.959999999999986	9.919999999999984

Obrázek 74: Hodnoty  $Z^*$  získané pomocí sférického modelu variogramu

	1	2	3	4	5	6	7
1	11.11	11.039999999999996	11.749999999999979	10.169999999999993	9.369999999999985	10.110000000000003	9.630000000000001
2	10.959999999999999	10.279999999999994	9.779999999999989	10.550000000000002	11.210000000000013	11.459999999999988	10.820000000000004
3	10.829999999999998	13.070000000000006	11.000000000000005	11.610000000000003	9.930000000000003	10.409999999999999	10.119999999999996
4	10.090000000000001	10.470000000000004	9.789999999999994	9.160000000000007	10.699999999999998	8.450000000000006	9.399999999999995
5	8.690000000000001	11.579999999999993	10.19	10.039999999999994	9.27	8.899999999999983	9.480000000000011
6	11.169999999999984	9.459999999999997	9.149999999999986	11.190000000000002	9.279999999999994	8.069999999999991	10.989999999999972
7	9.389999999999995	8.540000000000001	8.150000000000015	8.1	10.130000000000013	7.96	9.920000000000021

Obrázek 75: Hodnoty  $Z^*$  získané pomocí racionálního kvadratického modelu variogramu

	1	2	3	4	5	6	7
1	11.110000000000014	11.040000000000026	11.750000000000003	10.170000000000025	9.370000000000022	10.110000000000024	9.630000000000019
2	10.960000000000012	10.280000000000008	9.780000000000002	10.550000000000011	11.210000000000004	11.460000000000017	10.820000000000013
3	10.830000000000004	13.070000000000025	11.000000000000021	11.610000000000003	9.930000000000025	10.410000000000018	10.120000000000003
4	10.090000000000021	10.470000000000018	9.790000000000004	9.160000000000018	10.700000000000019	8.450000000000017	9.399999999999995
5	8.690000000000005	11.580000000000009	10.190000000000024	10.040000000000017	9.270000000000017	8.900000000000006	9.480000000000018
6	11.170000000000016	9.460000000000017	9.150000000000003	11.190000000000019	9.280000000000002	8.070000000000022	10.990000000000014
7	9.390000000000017	8.540000000000008	8.150000000000034	8.100000000000016	10.130000000000001	7.960000000000025	9.920000000000004

Obrázek 76: Hodnoty  $Z^*$  získané pomocí exponenciálního modelu variogramu

	1	2	3	4	5	6	7
1	11.109999999999995	11.039999999999958	11.749999999999948	10.169999999999997	9.369999999999987	10.109999999999989	9.629999999999988
2	10.959999999999939	10.279999999999996	9.779999999999948	10.549999999999963	11.209999999999981	11.459999999999983	10.819999999999988
3	10.829999999999941	13.069999999999947	10.999999999999963	11.609999999999976	9.929999999999971	10.409999999999973	10.119999999999989
4	10.089999999999994	10.469999999999951	9.789999999999958	9.159999999999963	10.699999999999969	8.449999999999982	9.399999999999979
5	8.689999999999965	11.579999999999943	10.189999999999964	10.039999999999978	9.269999999999976	8.899999999999983	9.479999999999974
6	11.169999999999964	9.459999999999956	9.149999999999965	11.189999999999996	9.279999999999976	8.069999999999991	10.989999999999966
7	9.389999999999942	8.539999999999996	8.149999999999995	8.099999999999994	10.129999999999972	7.959999999999967	9.919999999999973

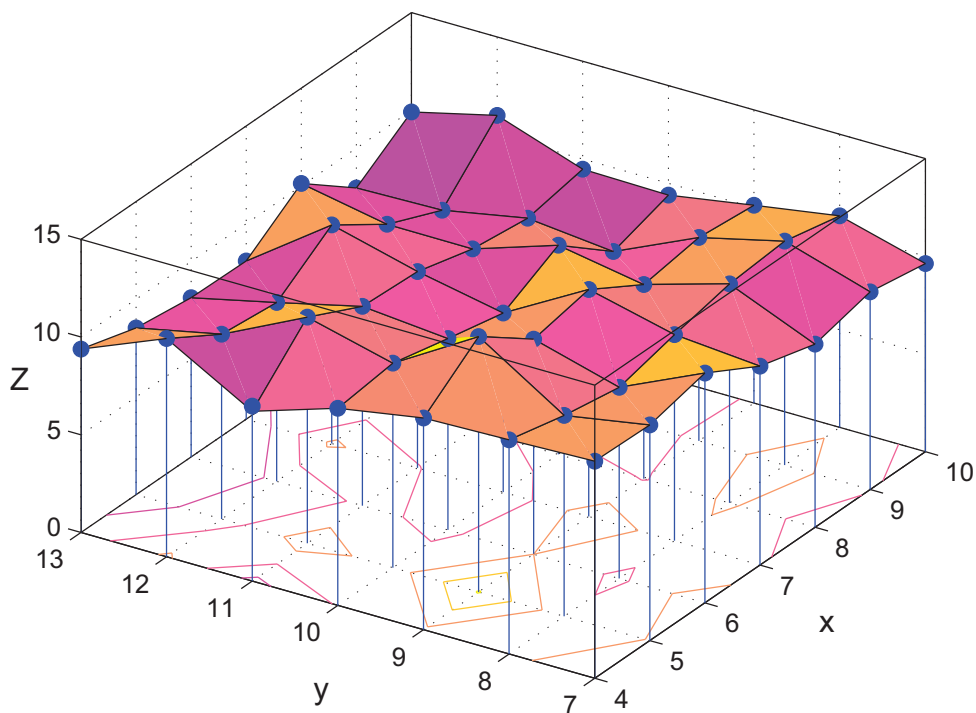
Obrázek 77: Hodnoty  $Z^*$  získané pomocí mocninného modelu variogramu

	1	2	3	4	5	6	7
1	11.110000000000056	11.040000000000067	11.750000000000048	10.170000000000051	9.370000000000058	10.110000000000047	9.630000000000043
2	10.960000000000052	10.280000000000053	9.780000000000005	10.550000000000034	11.210000000000036	11.460000000000002	10.820000000000013
3	10.830000000000004	13.070000000000045	11.000000000000003	11.610000000000031	9.929999999999993	10.410000000000002	10.119999999999989
4	10.090000000000044	10.470000000000029	9.790000000000026	9.160000000000004	10.699999999999998	8.449999999999985	9.399999999999956
5	8.690000000000042	11.580000000000028	10.190000000000007	10.040000000000006	9.269999999999966	8.899999999999998	9.479999999999956
6	11.170000000000046	9.460000000000019	9.150000000000025	11.189999999999994	9.279999999999996	8.069999999999945	10.989999999999966
7	9.390000000000036	8.540000000000045	8.150000000000013	8.099999999999971	10.129999999999962	7.959999999999959	9.919999999999943

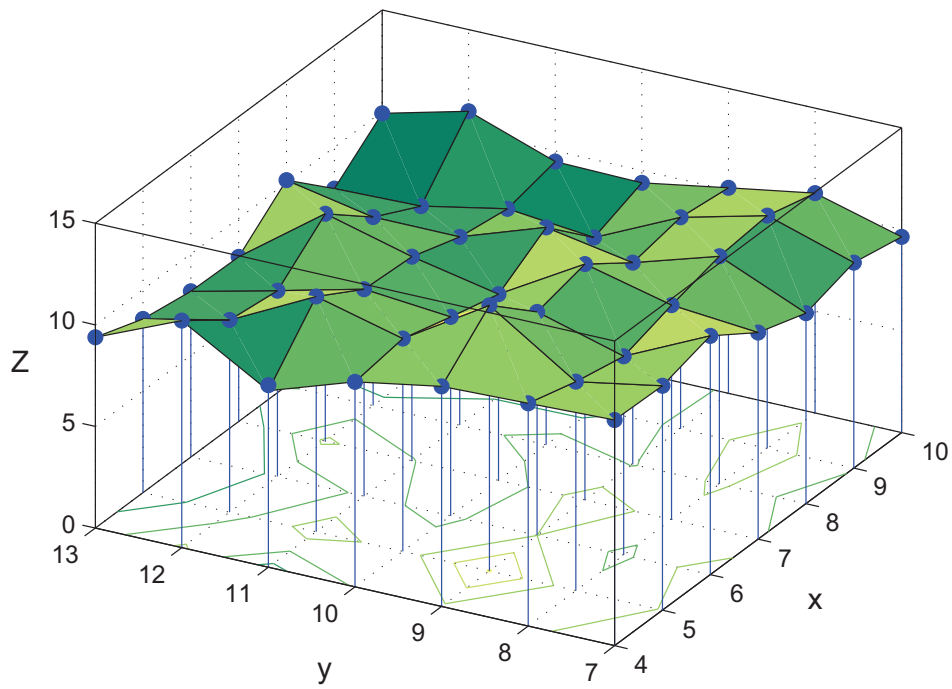
Obrázek 78: Hodnoty  $Z^*$  získané pomocí vlnového modelu variogramu

### 10.1.2 Grafické znázornění výsledků

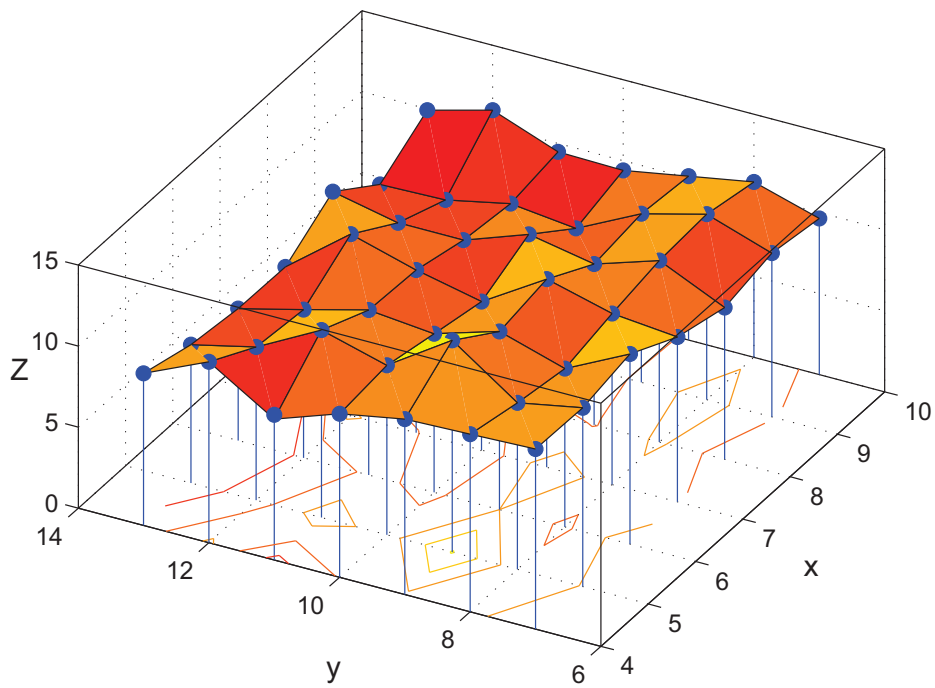
Pro lepší průhlednost jsem vykreslila pouze hodnoty  $Z$ , ke kterým jsem počítala  $Z^*$ . Hodnoty  $Z^*$  jsem navíc proložila plochou a zvolila vhodnější úhel pohledu.



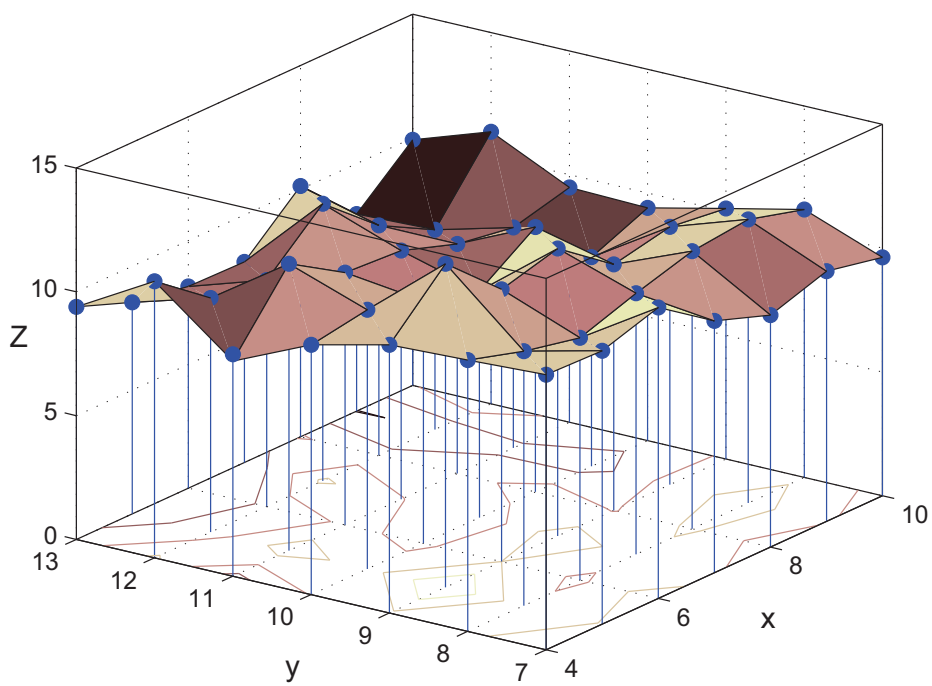
Obrázek 79: Vykreslení hodnot  $Z^*$  pro lineární model variogramu



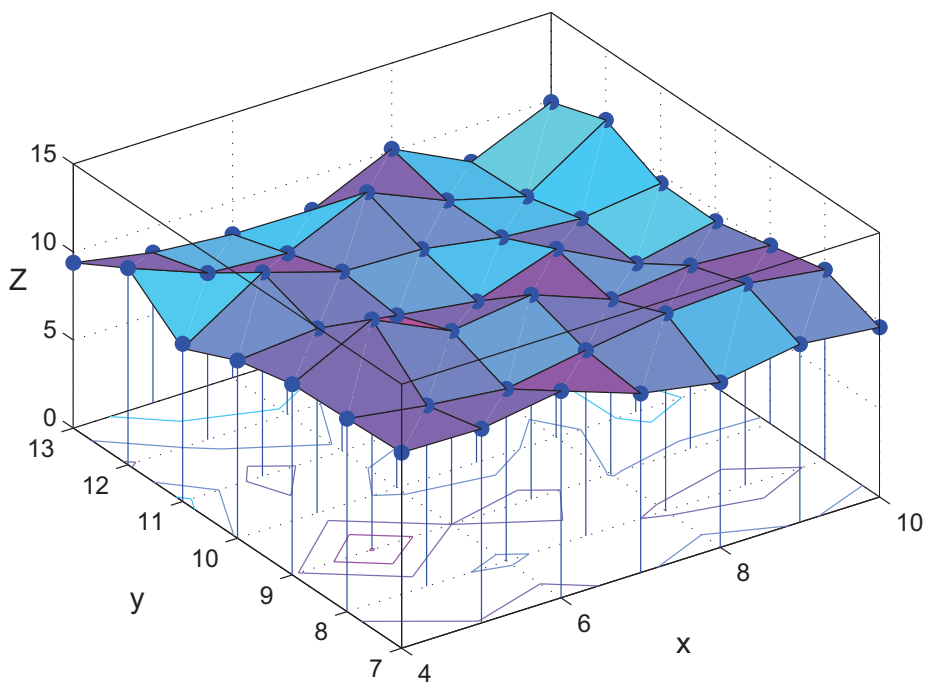
Obrázek 80: Vykreslení hodnot  $Z^*$  pro sférický model variogramu



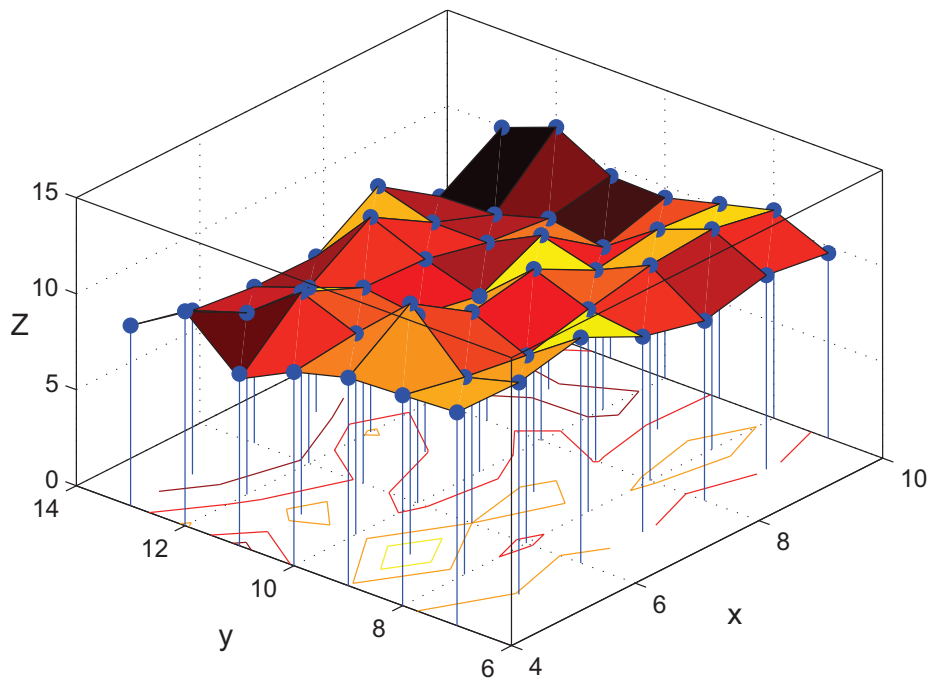
Obrázek 81: Vykreslení hodnot  $Z^*$  pro racionální kvadratický model variogramu



Obrázek 82: Vykreslení hodnot  $Z^*$  pro exponenciální model variogramu



Obrázek 83: Vykreslení hodnot  $Z^*$  pro mocninný model variogramu



Obrázek 84: Vykreslení hodnot  $Z^*$  pro vlnový model variogramu

### 10.1.3 Reziduální analýza

Jak je vidět na Obr.85 - Obr.90, hodnoty reziduí jsou si opět velmi blízké a opět se nedaří vykreslení jednotlivých reziduí pro všechny základní typy modelových variogramů.

	1	2	3	4	5	6	7
1	7.1054e-015	1.2434e-014	7.1054e-015	7.1054e-015	1.5987e-014	8.8818e-015	1.0658e-014
2	1.0658e-014	3.5527e-015	1.2434e-014	8.8818e-015	1.2434e-014	7.1054e-015	5.3291e-015
3	5.3291e-015	7.1054e-015	3.5527e-015	1.5987e-014	2.3093e-014	1.0658e-014	1.0658e-014
4	1.2434e-014	1.2434e-014	8.8818e-015	1.4211e-014	8.8818e-015	1.2434e-014	1.2434e-014
5	0.27	1.0658e-014	1.0658e-014	1.4211e-014	3.5527e-015	7.1054e-015	1.0658e-014
6	7.1054e-015	1.4211e-014	1.5987e-014	5.3291e-015	5.3291e-015	1.2434e-014	1.0658e-014
7	1.0658e-014	5.3291e-015	3.5527e-015	8.8818e-015	1.7764e-015	6.2172e-015	8.8818e-015

Obrázek 85: Rezidua odpovídajících si bodů pro lineární model

	1	2	3	4	5	6	7
1	1.0658e-014	1.7764e-015	1.4211e-014	-7.1054e-015	0	1.0658e-014	1.7764e-014
2	8.8818e-015	1.954e-014	7.1054e-015	1.2434e-014	1.4211e-014	7.1054e-015	0
3	7.1054e-015	7.1054e-015	1.7764e-015	8.8818e-015	1.2434e-014	7.1054e-015	7.1054e-015
4	1.0658e-014	1.0658e-014	1.0658e-014	2.1316e-014	1.7764e-014	8.8818e-015	5.3291e-015
5	0.27	1.0658e-014	3.5527e-015	8.8818e-015	5.3291e-015	1.7764e-014	1.5987e-014
6	1.2434e-014	1.5987e-014	1.0658e-014	5.3291e-015	1.5987e-014	1.7764e-014	2.8422e-014
7	1.954e-014	1.2434e-014	2.1316e-014	3.5527e-015	1.2434e-014	1.4211e-014	1.5987e-014

Obrázek 86: Rezidua odpovídajících si bodů pro sférický model

	1	2	3	4	5	6	7
1	0	3.5527e-015	2.1316e-014	7.1054e-015	1.4211e-014	-3.5527e-015	-8.8818e-015
2	1.0658e-014	5.3291e-015	1.0658e-014	-1.7764e-015	-1.2434e-014	1.2434e-014	-3.5527e-015
3	1.7764e-015	-5.3291e-015	-5.3291e-015	-3.5527e-015	-3.5527e-015	1.0658e-014	3.5527e-015
4	-1.0658e-014	-3.5527e-015	5.3291e-015	-7.1054e-015	1.7764e-015	-7.1054e-015	5.3291e-015
5	0.27	7.1054e-015	0	5.3291e-015	0	1.7764e-014	-1.0658e-014
6	1.5987e-014	3.5527e-015	1.4211e-014	-2.1316e-014	5.3291e-015	8.8818e-015	3.5527e-015
7	5.3291e-015	-1.7764e-015	-1.4211e-014	0	-1.2434e-014	0	-2.1316e-014

Obrázek 87: Rezidua odpovídajících si bodů pro racionální kvadratický model

	1	2	3	4	5	6	7
1	-1.4211e-014	-2.6645e-014	-3.0198e-014	-2.4869e-014	-2.3093e-014	-2.4869e-014	-1.7764e-014
2	-1.0658e-014	-8.8818e-015	-2.1316e-014	-1.0658e-014	-3.908e-014	-1.5987e-014	-1.2434e-014
3	-3.5527e-015	-2.4869e-014	-2.1316e-014	-3.0198e-014	-2.4869e-014	-1.7764e-014	-3.0198e-014
4	-2.1316e-014	-1.7764e-014	-5.3291e-015	-1.7764e-014	-1.954e-014	-1.7764e-014	5.3291e-015
5	0.27	-8.8818e-015	-2.4869e-014	-1.7764e-014	-1.7764e-014	-5.3291e-015	-1.7764e-014
6	-1.5987e-014	-1.5987e-014	-3.0198e-014	-1.954e-014	-2.1316e-014	-2.1316e-014	-1.4211e-014
7	-1.5987e-014	-8.8818e-015	-3.3751e-014	-1.5987e-014	-8.8818e-015	-2.4869e-014	-4.0856e-014

Obrázek 88: Rezidua odpovídajících si bodů pro exponenciální model

	1	2	3	4	5	6	7
1	4.9738e-014	4.0856e-014	5.1514e-014	3.0198e-014	1.2434e-014	1.0658e-014	1.2434e-014
2	6.2172e-014	3.908e-014	5.1514e-014	3.7303e-014	1.954e-014	1.7764e-014	1.2434e-014
3	5.862e-014	5.3291e-014	3.7303e-014	2.3093e-014	2.8422e-014	2.6645e-014	1.0658e-014
4	6.0396e-014	4.9738e-014	4.0856e-014	3.7303e-014	3.0198e-014	1.7764e-014	2.1316e-014
5	0.27	5.6843e-014	3.5527e-014	2.1316e-014	2.3093e-014	1.7764e-014	2.6645e-014
6	3.5527e-014	4.4409e-014	3.1974e-014	3.908e-014	2.3093e-014	8.8818e-015	3.3751e-014
7	5.862e-014	3.908e-014	4.9738e-014	5.3291e-015	2.8422e-014	3.2863e-014	2.6645e-014

Obrázek 89: Rezidua odpovídajících si bodů pro mocninný model

	1	2	3	4	5	6	7
1	-5.6843e-014	-6.7502e-014	-4.7962e-014	-5.1514e-014	-5.862e-014	-4.7962e-014	-4.2633e-014
2	-5.1514e-014	-5.3291e-014	-5.1514e-014	-3.3751e-014	-3.5527e-014	-1.954e-014	-1.2434e-014
3	-3.908e-014	-4.4409e-014	-3.0198e-014	-3.1974e-014	7.1054e-015	-1.7764e-015	1.0658e-014
4	-4.4409e-014	-2.8422e-014	-2.6645e-014	-3.5527e-015	1.7764e-015	1.4211e-014	4.4409e-014
5	0.27	-2.8422e-014	-7.1054e-015	-7.1054e-015	3.3751e-014	1.954e-014	4.4409e-014
6	-4.6185e-014	-1.7764e-014	-2.4869e-014	5.3291e-015	3.5527e-015	5.5067e-014	3.3751e-014
7	-3.5527e-014	-4.6185e-014	-1.2434e-014	2.8422e-014	3.908e-014	4.0856e-014	5.6843e-014

Obrázek 90: Rezidua odpovídajících si bodů pro vlnový model

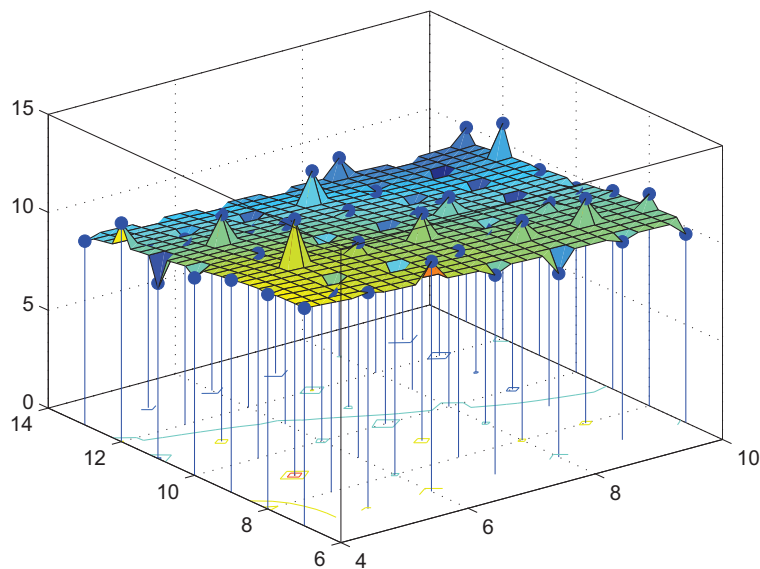
## 10.2 Srovnání výsledků

Podle hodnot reziduí můžeme říci, že se nepotvrdil vliv volby variogramu na predikci. U reálného příkladu ale dochází k tomu, že u některých odhadů vycházejí rezidua pouze kladná (Obr.85), záporná (Obr. 88) nebo se znaménka reziduí střídají (Obr. 87, Obr. 90). U modelů, pro které rezidua střídají znaménka, není odhad systematicky nadhodnocen nebo podhodnocen a tudíž by příslušnému modelu měla být dána přednost. U tohoto příkladu by tedy bylo vhodné zvolit racionální kvadratický nebo vlnový model.

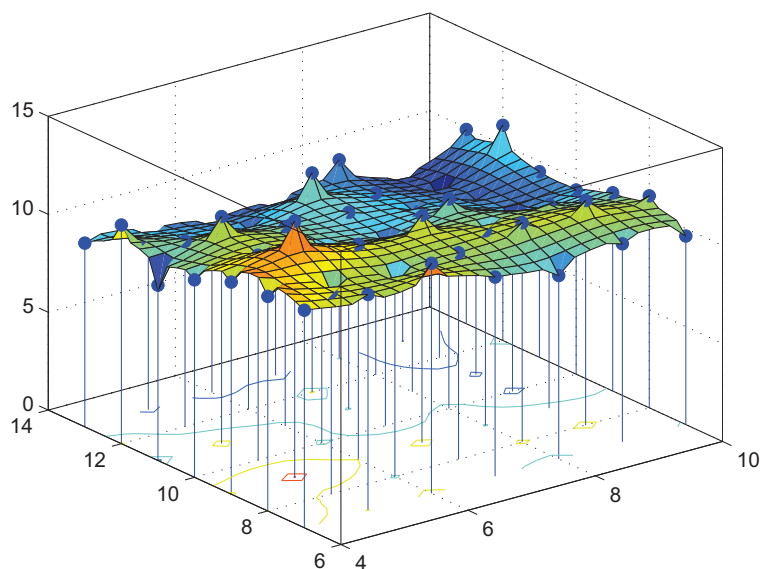


### 10.3 Zjemnění škály měřítka souřadnic bodů predikce

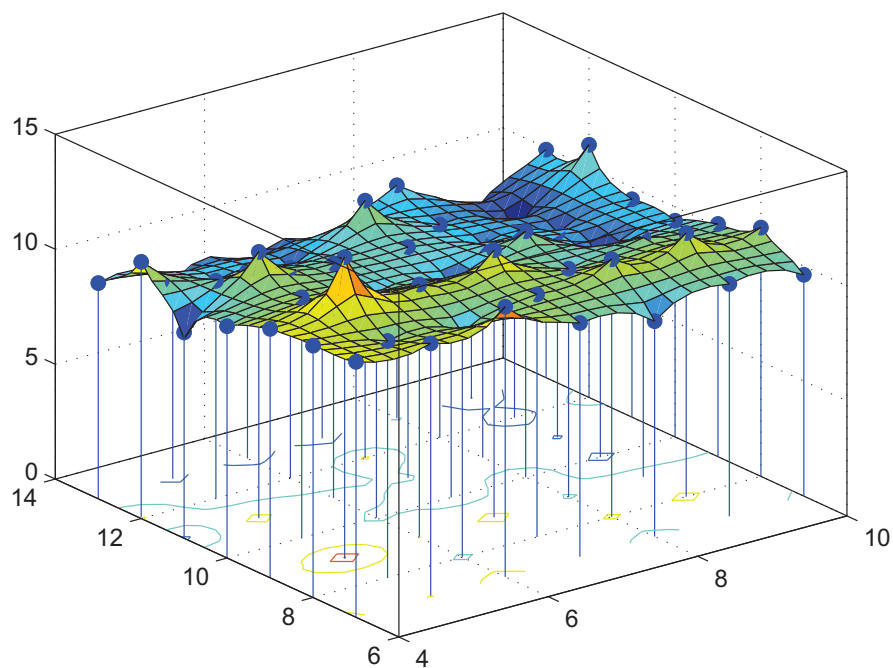
Pro úplnost uvedeme ještě obrázky pro zjemněné vektory  $x_0$  a  $y_0$  a zhodnotit vliv volby variogramu pomocí těchto obrázků. Vektory  $x_0$  a  $y_0$  zjemníme na  $x_0 = 4 : 0.25 : 10$  a  $y_0 = 7 : 0.25 : 14$  dostaneme následující obrázky.



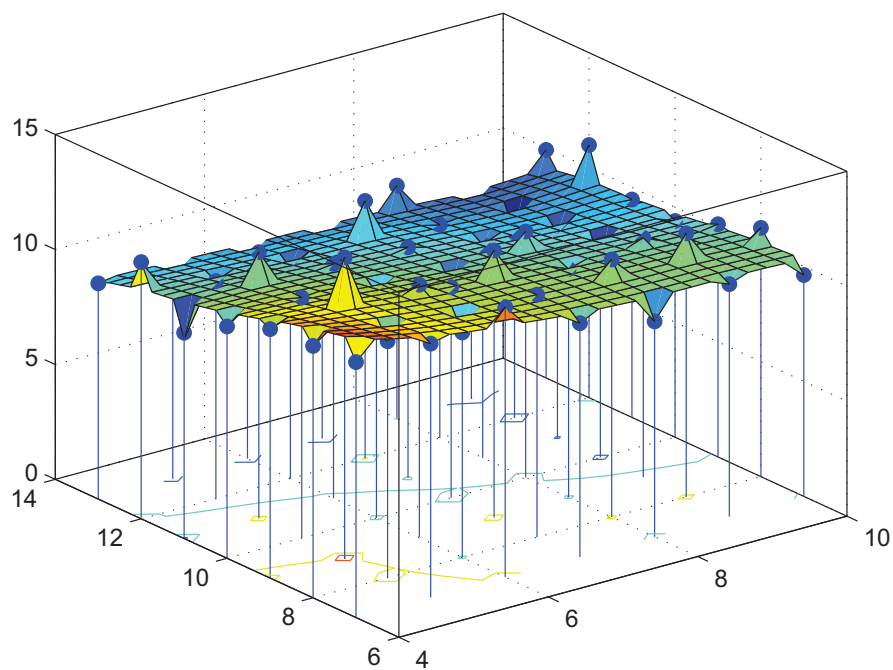
Obrázek 91: Vykreslená plocha odhadů získaných pomocí lineárního modelu pro zjemněné vektory  $x_0$  a  $y_0$



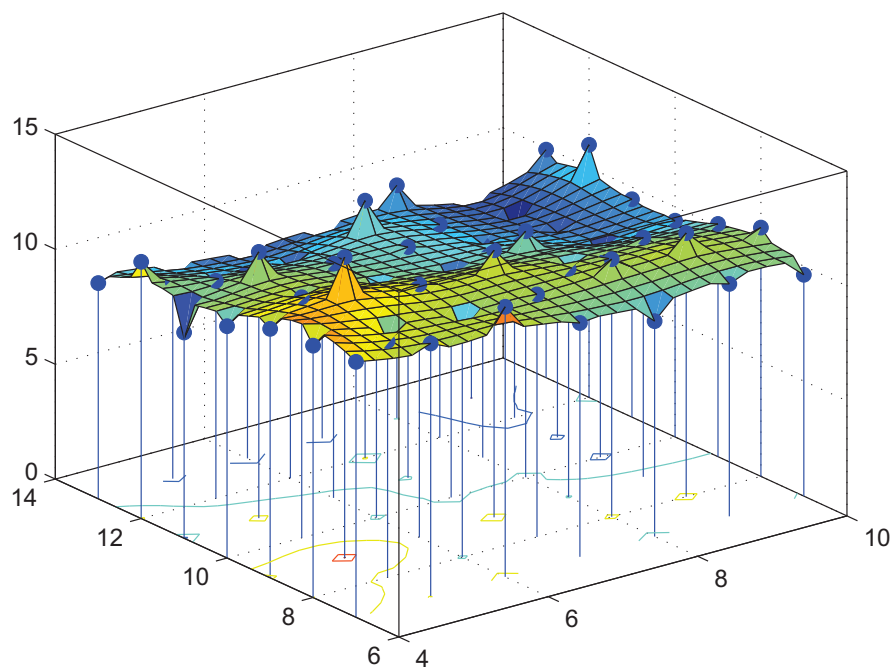
Obrázek 92: Vykreslená plocha odhadů získaných pomocí sférického modelu pro zjemněné vektory  $x_0$  a  $y_0$



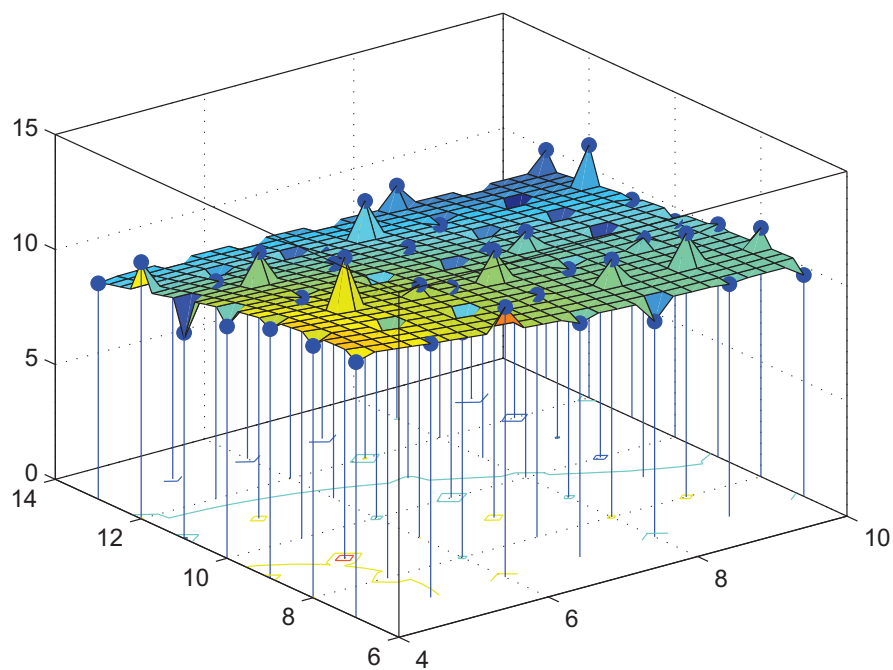
Obrázek 93: Vykreslená plocha odhadů získaných pomocí racionálního kvadratického modelu pro zjemněné vektory  $x_0$  a  $y_0$



Obrázek 94: Vykreslená plocha odhadů získaných pomocí exponenciálního modelu pro zjemněné vektory  $x_0$  a  $y_0$



Obrázek 95: Vykreslená plocha odhadů získaných pomocí mocninného modelu pro zjemněné vektory  $x_0$  a  $y_0$



Obrázek 96: Vykreslená plocha odhadů získaných pomocí vlnového modelu pro zjemněné vektory  $x_0$  a  $y_0$

Odhady jsou velmi podobné i pro zjemněné vektory, všechny modely jsou pro odhady variogramů vhodné a žádný z modelů nemá výrazně odlišný vliv na odhad od ostatních.

## 11 Závěr

V předložené diplomové práci byly popsány vybrané regresní metody pro predikci kovariančně stacionárního procesu, podle literatury byly vybrány metody pro odhad vario-gramu a byl studován vliv tohoto odhadu na predikci. Různé modely pro odhad vario-gramu byly porovnány a jejich vliv na predikci procesu byl posouzen pomocí programové implementace.

Program byl použit na 3 datových strukturách. Simulovaná data byla zastoupena re-gresní rovinou a obecnou lineární regresní plochou. Do simulací byla přidávána nekorelo-vaná náhodná chyba. Implementace reálných dat byla provedena na příkladě o uhelném popelu (Coal-Ash Data). Ukázalo se, že volba modelu nemá v těchto příkladech vliv na predikci. Autorem vyvinuté programy jsou nahrány na přiloženém CD.

## 12 Seznam použitých zdrojů

- [1] Cressie, N.A.C.: *Statistics for spatial data*. John-Wiley & Sons, Inc., New York, 1993.
- [2] Rehman, S.U. and Shapiro, A.: *An Integral transform approach to cross-variograms modeling*. Computational Statistics & Data Analysis 22 1996 p. 213-233.
- [3] Cressie, N.A.C. and Hawkins, D.M.: *Robust estimation of the variogram*. Journal of the International Association for Mathematical Geology, 12, 1980, p.115-125.
- [4] Anděl, J.: *Základy matematické statistiky*. MATFYZPRESS, Praha, 2005. 1. vydání, 358 s. ISBN 80-86732-40-1.
- [5] Karpíšek, Z.: *Matematika IV - Pravděpodobnost a statistika: Učební text*. CERM Brno, Brno, 2003.
- [6] Michálek, J.: *Dvě poznámky o metodě kriging*. Robust 2000, p. 189-197.
- [7] Pfefferová, A.: *Effectiveness of Design of Experiment*. [Diplomová práce.] Brno: VUT, FSI, 2008. 75 p.
- [8] Maroš, B.: *Empirické modely I: Analýza inženýrského experimentu*. CERM Brno, 2001. 112 s. ISBN 80-214-1984-9.