

VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS

Ingrida Borisenko

ERDVINIŲ DUOMENŲ MODELIAVIMAS
NAUDOJANT STACIONARIŲ ERDVINIŲ
PROCESŲ SEMIVARIOGRAMAS

DAKTARO DISERTACIJA

FIZINIAI MOKSLAI,
MATEMATIKA (01P)



LEIDYKLA
Vilnius TECHNIKA 2009

Disertacija rengta 2001–2009 metais Klaipėdos universitete ir Vilniaus Gedimino technikos universitete.

Disertacija ginama eksternu.

Mokslinis konsultantas

prof. dr. Kęstutis DUČINSKAS (Klaipėdos universitetas, fiziniai mokslai, matematika – 01P).

VGTM leidyklos TECHNIKA 1661-M mokslo literatūros knyga
<http://leidykla.vgtu.lt>

ISBN 978-9955-28-480-2

© Vilniaus Gedimino technikos universitetas, 2009

© Borisenko, I., 2009

VILNIUS GEDIMINAS TECHNICAL UNIVERSITY

Ingrida BORISENKO

MODELLING OF SPATIAL DATA USING SEMIVARIOGRAMS OF STATIONARY SPATIAL PROCESSES

DOCTORAL DISSERTATION

PHYSICAL SCIENCES,
MATHEMATICS (01P)



LEIDYKLA
Vilnius TECHNIKA 2009

Doctoral dissertation was prepared at Klaipėda University and Vilnius Gediminas Technical University in 2001–2009.

The dissertation is defended as an external work.

Scientific Consultant

Prof Dr Kęstutis DUČINSKAS (Klaipėda University, Physical Sciences, Mathematic – 01P).

Santrauka

Disertacijoje nagrinėjama erdvinių duomenų su stacionariomis klaidomis modeliavimo per semivariogramas ir tiesinio prognozavimo metodika.

Erdvinių duomenų skiriamasis bruožas – jų išsidėstymas erdvėje, kuris dažniausiai aprašomas geografinėmis koordinatėmis. Tokių duomenų modeliavimas semivariogramomis, ir prognozavimas krigingu yra vienas iš svarbių geostatistikos mokslo uždavinių. Krigingas yra stochastinis prognozavimo metodas, kuris prie tam tikrų sąlygų pateikia geriausią tiesinę nepaslinktą prognozę. Krigingo rezultatų paklaidos priklauso nuo to kaip tiksliai erdvinių duomenų sklaida aprašoma kovariacine funkcija arba semivariograma. Darbe dėmesys skiriamas semivariogramoms, nes jos aprašo platesnę erdvinių procesų klasę.

Pagrindinis disertacijos tikslas yra apibendrinti ir realizuoti vieningą erdvinių duomenų su stacionariomis klaidomis modeliavimo metodiką, pagrįstą semivariogramomis. Darbo objektai yra semivariogramos, jų modeliai, įvairūs erdvinių duomenų prognozavimo metodai bei erdvinių duomenų modeliavimo, prognozavimo etapai. Šių objektų analizė bei interpretacija prie tam tikrų sąlygų leidžia gauti geriausių erdvinių duomenų modeliavimo bei prognozavimo rezultatus.

Taip pat disertaciniame darbe empiriniam Matern'o semivariogramų įvertinimui MoM pateikta dispersijų-kovariacijų matricos išraiška per teorines semivariogramas stacionaraus Gauso duomenų modelio atveju.

Tiriami erdvinių duomenų vidurkio modelio parametrų bei semivariogramų vertinimo metodai. Išvedama vidurkio parametrų apibendrinto mažiausių kvadratų įvertinio formulė per semivariogramas.

Vėliau vidurkio modelio parametrams D ir A kriterijų funkcijų išraiškos pateikiamos per semivariogramas. Šie kriterijai panaudoti nustatant optimalų taškų išsidėstymą taisyklingoje 2×2 gardelėje.

Taip pat Lahiri *et al.* (2002) apibrėžtam didėjančios erdvės asimptotikos erdviniam imčių planui LIDA įrodytas MoM įvertinio asimptotinis normališkumas atveju, kai duomenų modelio klaidos yra stacionarios ir Gausinės. Išvesta asimptotinės dispersijų-kovariacijų matricos išraiška izotropinės eksponentinės semivariogramos atveju.

Praktiniams skaičiavimams atlikti naudoti įvairūs atviro kodo R sistemos paketai.

Disertaciją sudaro įvadas, keturi skyriai, rezultatų apibendrinimas, naudotos literatūros ir autoriaus publikacijų disertacijos tema sąrašai bei keturi priedai.

Abstract

The main objective of this thesis is to propose the methodology for modeling and prediction of spatial data through semivariograms, when the spatial data are generated by spatial processes with stationary errors. The main distinguishing of spatial data is their distribution in space, which usually expressed in geographic coordinates. The modeling of such data through semivariograms and prediction by kriging is one of main objective of geostatistic science. Kriging is a stochastic prediction method, which under certain conditions provides the best linear unbiased prediction. The efficiency of kriging depends on how accurately the covariance function or semivariogram describes the variation of spatial data. This work focuses on semivariograms, because it covers the wider class of spatial processes.

The objects of thesis are semivariograms, their models, the various prediction methods of spatial data and the stages of modeling of spatial data. The analysis and interpretation of these objects under some conditions provide the best results of spatial data modeling and prediction.

Also the expression of asymptotic variance-covariance matrix for empirical semivariogram estimator of Matheron (MoM) is derived through theoretical semivariogram in case of stationary Gaussian model of data.

The methods, which are used for evaluation of parameters of mean models or semivariograms, are analysed. The formula of mean parameters of generalized least squares estimator is derived through semivariogram.

Further, the expressions of D and A criterion functions for parameters of mean models are given through semivariogram. These criteria are often used for determination of optimal spatial sampling design. The received expressions were applied to identify the optimal design of points in regular 2x2 grid, with the center (0,0).

Also, for increasing domain asymptotic spatial sampling design the asymptotical normality of MoM estimator, when errors of data model are stationary and Gaussian is proofed. The expression of asymptotic variance-covariance matrix in the case of exponential isotropic semivariogram is derived.

To perform practical calculations the various functions of open source *R* packages were used.

The scientific work consists of the general characteristic of the dissertation, 4 chapters, conclusions, list of literature, list of publications and addenda.

Žymėjimai

Simboliai

$X_n \xrightarrow{d} X$	– konvergavimas pagal pasiskirstymą ($n \rightarrow \infty$)
$\mathbf{I}_A(x) = \begin{cases} 1, & x \in A, \\ 0, & \text{priešingu atveju} \end{cases}$	– aibės A indikatorius
$ x = \left(\sum_{i=1}^p x_i^2 \right)^{1/2}$	– jei $x = (x_1, \dots, x_p)^T \in R^p$
$ x _1 = \sum_{i=1}^p x_i $	
$ A = K$	– jei $A = \{a_1, \dots, a_K\}$, t. y. aibę A sudaro baigtinis elementų skaičius
$\mathbf{1}_n$	– n-matis vektorius-stulpelis, kurio visi elementai lygus 1
$\mathbf{I}_n$	– $n \times n$ vienetinė matrica, kurios įstrižainės elementai lygūs 1, o visi lygus 0
$\mathbf{J}_n = \mathbf{1}_n \mathbf{1}_n^T$	– $n \times n$ matrica, kurios visi elementai lygus 1
A^T	– transponuota matrica, t. y. matricos A eilutės tampa A^T stulpeliais
$\det(A)$	– kvadratinės matricos A determinantas
$tr(A) = \sum_i a_{ii}$	– kvadratinės matricos pėdsakas
$V(\mathbf{X}) = \text{cov}(\mathbf{X}, \mathbf{X})$	– atsitiktinio dydžio $\mathbf{X}$ kovariacijų matrica
$\text{cov}(\mathbf{X}, \mathbf{Y})$	– atsitiktinių vektorių $\mathbf{X}$ ir $\mathbf{Y}$ kovariacijų matrica

Santrumpos

VKPK	– vidutinė kvadratinė prognozės klaida (<i>angl. MSPE – Mean Squared Prediction Error</i>)
SMKM	– svertinis mažiausių kvadratų metodas (<i>angl. Weighted Least Squares</i>)
PMKM	– paprastas mažiausių kvadratų metodas (<i>angl. Ordinary Least Squares</i>)

AMKM	– apibendrintas mažiausių kvadratų metodas (<i>angl. General Least Squares</i>)
MTLM	– maksimalaus tikėtinumo liekanoms metodas (<i>angl. Restricted Error Maximum Likelihood</i>)
MTM	– maksimalaus tikėtinumo metodas (<i>angl. Maximum Likelihood Method</i>)
PKS	– paklaidų kvadratų suma (<i>angl. SSE , Sum Square of Error</i>)

Turinys

IVADAS	1
Tiriamoji problema ir jos aktualumas	1
Tyrimų objektas	2
Darbo tikslas	3
Darbo uždaviniai	3
Tyrimų metodika	3
Darbo mokslinis naujumas ir jo reikšmė	3
Darbo rezultatų praktinė reikšmė	4
Ginamieji teiginiai	4
Darbo rezultatų aprobavimas	5
Disertacijos struktūra	5
 1.ERDVINIŲ DUOMENŲ MATEMATINIAI MODELIAI	6
1.1. Erdviniai duomenys	6
1.2. Skaliariniai erdviniai procesai ir jų geostatistinės charakteristikos	9
1.3. Vidurkio modeliai ir jų parametrų įvertiniai	11
1.4. Semivariogramų parametrai ir jų modeliai	13
1.5. Pirmojo skyriaus apibendrinimas ir disertacijos uždavinių formulavimas	26

2.ERDVINIŲ DUOMENŲ STATISTINIO MODELIAVIMO IR PROGNOZAVIMO METODIKA	28
2.1. Prognozavimo metodų klasifikavimas. Bendriniai modeliavimo ir prognozavimo etapai	29
2.2. Statistinė erdvinių duomenų analizė	31
2.3. Semivariogramų modeliavimas	34
2.4. Krigingo tipai	46
2.5. Modelių patikros metodai	59
2.6. Semivariogramos parametrų įtaka prognozei. Krigingo trūkumai ir privalumai	62
2.7. Antrojo skyriaus apibendrinimas. Papildomų disertacijos uždavinių formulavimas.	64
3.VIDURKIO PARAMETRŲ IR SEMIVARIOGRAMŲ ĮVERTINIAI	66
3.1. Vidurkio parametrų AMKM įvertinio išraiška per semivariogramas	67
3.2. Taškų išsidėstymo 2x2 gardelėje optimizavimas.....	70
3.3. Empirinio semivariogramos įvertinio (MoM) asimptotinė analizė.....	74
3.4. Trečiojo skyriaus apibendrinimas ir išvados.....	79
4.SIŪLAMOS ERDVINIŲ DUOMENŲ MODELIAVIMO IR PROGNOZAVIMO METODIKOS TAIKYMAS	80
4.1. 1990 metų gylio ir dugno nuosėdų duomenų modeliavimas ir prognozė.....	82
4.2. 2005 metų gylio duomenų modeliavimas ir prognozė.....	103
4.3. Ketvirtojo skyriaus apibendrinimas ir išvados.....	114
BENDROSIOS IŠVADOS.....	115
LITERATŪROS SĄRAŠAS	117
AUTORIAUS MOKSLINIŲ PUBLIKACIJŲ DISERTACIJOS TEMA SĄRAŠAS	124
PRIEDAI.....	126
A priedas. Pradinės statistinės duomenų analizės komandos	126
B priedas. 1990 gylio duomenų prognozė ordinariu krigingu	128
C priedas. 1990 gylio duomenų prognozė universaliu krigingu	129
D priedas. 2005 gylio duomenų, atvaizduotų tinklu, prognozė medianinio išlyginimo krigingu	130

Contents

INTRODUCTION	1
Topicality of the problem	1
Research object.....	2
Aim of work.....	3
Tasks of the work.....	3
Methodology of research	3
Scientific novelty of the work results	3
Practical value of the work results.....	4
Defended propositions	4
Approval of the work.....	5
Structure of the dissertation	5
 1. THE MATHEMATICAL MODELS OF SPATIAL DATA	6
1.1. Spatial data.....	6
1.2. Scalar spatial process and their geostatistical characteristics	9
1.3. Models of mean and their parametric estimators.....	11
1.4. Models of semivariogram and their parameters.....	13
1.5. Generalisation of the first chapter and formulation of tasks of the dissertation.....	26
 2.THE METHODOLOGY OF STATISTICAL MODELLING AND PREDICTION OF SPATIAL DATA	28

2.1. The clasification of prediction methods. Generic stages of modelling and prediction	29
2.2. The statistical analysis of spatial data	31
2.3. The modelling of semivariograms	34
2.4. Types of krigings	46
2.5. The validation of models	59
2.6. Influence of semivariogram parameters to the prediction. The disadvantages and advantages of kriging	62
2.7. Generalisation of the second chapter. Formulation of additional tasks of the dissertation	64
3.ESTIMATORS OF PARAMETERS OF MEAN AND SEMIVARIOGRAMS	66
3.1. The expresion of GLS estimator of mean parameters evaluated in terms of semivariograms	67
3.2. Optimisation of spatial design 2x2 grid	70
3.3. The asymptotic analysis of MoM estimator	74
3.4. Generalisation of the third chapter	79
4.THE APPLICATION OF PROPOSED METHODOLOGY OF MODELLING AND PREDICTION	80
4.1. The modelling and prediction of the depth and sediments data of 1990	82
4.2. The modelling and prediction of the depth data of 2005	103
4.3. Generalisation of the fourth chapter	114
GENERAL CONCLUSIONS.....	115
REFERENCES.....	117
LIST OF AUTHOR SCIENTIFIC PUBLICATIONS IN SUBJECT OF DISSERTATION	124
ADDENDAS	126
Addenda A. The function of preliminary spatial analysis	126
Addenda B. The prediction of the depth data of 1990 by ordinary kriging.....	128
Addenda C. The prediction of the depth data of 1990 by universal kriging.....	129
Addenda D. The prediction of the gridded depth data of 2005 by median polishing kriging.....	130

Ivadas

Tiriamoji problema ir jos aktualumas

Erdvinė statistika – viena jauniausių statistikos mokslo krypčių, praėjusio šimtmečio penktame dešimtmetyje pradėta naudoti kalnakasyboje. Per pusšimtį metų nuo šios statistikos atšakos atsiradimo, išaugo ją nagrinėjančių mokslininkų ratas bei taikymo sričių spektras, buvo išleista nemažai teorinio (Cressie 1993) ir praktinio (Isaaks, Srivastava 1989) pobūdžio metodinių priemonių. Šiuo metu erdvinės statistikos metodai taikomi ekologijoje, kiekybinėje geologijoje, vaizdų apdorojime ir analizėje, epidemiologiniuose, klimato pokyčių ir net Visatos tyrimuose. Lietuvoje gi erdvinių duomenų tyrimo metodologija pradėta nagrinėti tik šio šimtmečio pradžioje, šio mokslo taikymai vykdyti tik kelių mokslininkų (Dumbrasukas, A., Kumetaitis, A.; Kumetaitienė, A. ir kt.). Tad labai svarbu plėtoti šią sritį, tobulinti esamus metodus, išnagrinėti erdvinių duomenų modeliavimo metodus ir pateikti bendrinę erdvinių duomenų modeliavimo metodologiją.

Norint taikyti erdvinės statistikos metodus, būtina žinoti duomenų išsidėstymą erdvėje, kuris dažniausiai aprašomas geografinėmis koordinatėmis. Tad, vienas esminių bruožų, išskiriančių erdvinę statistiką nuo klasikinės, yra tas, kad ji modeliuoja ir erdvinį trendą, ir erdvinę koreliaciją.

Viena pagrindinių erdvinės statistikos užduočių yra erdvinių duomenų matematinio modelio sudarymas, kuris gali būti naudojamas interpoliacijai (ekstrapoliacijai) ar kitiems tikslams įgyvendinti. Duomenų prognozei atlikti gali būti taikomas krigingo metodas, kuris prie tam tikrų sąlygų minimizuoja vidutinę kvadratinę prognozės klaidą – pateikia geriausią tiesinę nepaslinktą prognozę. Priklausomai nuo erdvinių duomenų pobūdžio, erdvinio proceso vidurkio modelio ir kitų duomenų savybių krigingai skirstomi į kelis tipus.

Erdvinių duomenų sklaidą aprašoma kovariacine funkcija arba semivariograma. Pastaroji dažniau naudojama praktiniuose skaičiavimuose, ypač juos atliekant programiniais paketais, kadangi apima platesnę erdvinių procesų klasę. Turimų duomenų pagrindu sudaroma empirinė semivariograma (dažniausiai naudojant momentų metodu gaunamą Matheron'o įvertinį (MoM)), kuriai parenkamas ir priglodinamas vienas iš semivariogramos parametrinių modelių. Dažniausiai naudojami semivariogramų modeliai aprašomi keliais semivariogramos parametrais, kurių pagrindiniai yra grynuolis, slenkstis ir slenkščio atstumas.

Semivariogramų parametrų įvertinimui dažnai taikomi mažiausių kvadratų metodai (svertinis, paprastas bei apibendrintas mažiausių kvadratų metodai). Kaip nurodyta Lahiri ir kitų autorių straipsnyje "On asymptotic distribution and asymptotic efficiency of least squares estimators of spatial variogram parameters" (2002), prie bendrų sąlygų, lyginant su kitais mažiausių kvadratų vertinimo metodais, apibendrintas mažiausių kvadratų metodas (AMKM) vertinimo metodas yra asimptotiškai efektyvus.

Mažiausių kvadratų metodai gali būti taikomi ir vidurkio modelio parametrų įvertinimui. Kadangi AMKM metodas literatūroje pateikiamas per kovariacines funkcijas, svarbu užrašyti šio įvertinio išraišką per semivariogramas.

Taip pat reiktų atkreipti dėmesį, kad semivariogramų parametrų įvertiniai analitiniu būdu nėra gaunami. Todėl aktuali šių įvertinių asimptotinė analizė.

Kitas dažnai erdvinėje statistikoje pasitaikantis uždavinys, yra optimalaus imčių plano nustatymas. Plačiau taikomi du eksperimentų planų optimalumo kriterijai: kriterijus D, kriterijus A. Vėlgi, kadangi praktiniuose skaičiavimuose dažniau naudojamos semivariogramos, aktualu šiuos planus suformuluoti semivariogramų pavidalu.

Tyrimų objektas

Disertacijoje autorė tiria erdvinių duomenų generuotų erdviniais procesais su stacionariomis klaidomis modeliavimo, prognozavimo ir gautų rezultatų tyrimo metodiką.

Darbo tikslas

Pagrindinis darbo tikslas yra apibendrinti ir realizuoti vieningą erdvinių duomenų generuojamų erdviniais procesais su stacionariomis klaidomis modeliavimo metodiką, pagrįstą semivariogramomis.

Darbo uždaviniai

1. Apžvelgti adityvius erdvinių procesų modelius su stacionariomis klaidomis, aprašyti pagrindines erdvinių procesų charakteristikas.
2. Pasiūlyti apibendrintą erdvinių duomenų modeliavimo metodiką, naudojant įvairius parametrinius vidurkių ir stacionarių semivariogramų modelius.
3. Aprašytą bei pasiūlytą metodiką pritaikyti Kuršių marių gylio ir dugno nuosėdų duomenims modeliuoti ir prognozuoti.
4. Atlikti erdvinių imčių planų palyginimus, taikant įvairius optimalumo kriterijus.
5. Atlikti MoM įvertinio statistinių savybių analizę.

Tyrimų metodika

Analitinėje darbo dalyje aprašant erdvinių duomenų matematinius modelius bei erdvinių duomenų modeliavimo ir prognozavimo metodiką naudojami literatūros ir lyginamosios analizės metodai. Erdvinių duomenų modeliavimo etapams apibendrinti naudojami sąvokų apibendrinimo ir koncepcinio modeliavimo metodai. Darbe apibendrintų modeliavimo ir prognozavimo metodų praktinei realizacijai naudojami eksperimentinės analizės metodai.

Darbo mokslinis naujumas ir jo reikšmė

Šioje disertacijoje pateikiama vieninga erdvinių duomenų su stacionariomis klaidomis modeliavimo metodika, pagrįsta semivariogramomis. Analizuojant erdvinių duomenų su stacionariomis klaidomis modeliavimo metodiką, pastebėtos galimybės tobulinti kai kuriuos erdvinės statistikos metodus.

Šiame darbe pateikti nauji rezultatai:

1. MoM įvertinys pateiktas per erdvinio plano (Laplaso) matricą.
2. Užrašyta β parametrų apibendrinto mažiausių kvadratų metodo įvertinio išraiška per semivariogramas.

3. β parametru įvertiniui išvestos D ir A kriterijų funkcijų išraiškos per semivariogramas.
4. Stacionaraus Gauso erdvinio proceso atvejui įrodytas MoM semivariogramos įvertinio asimptotinis normališkumas, taikant didėjančios srities asimptotikos schemą.
5. Izotropinės eksponentinės semivariogramos atvejui išvesta minėto įvertinio asimptotinės dispersijų-kovariacijų matricos išraiška.
6. Pateikta apibendrinta erdvinų duomenų generuojamų erdviniais procesais su stacionariomis klaidomis modeliavimo ir prognozavimo metodika, leidžianti:
 - a) duomenų sklaidą aprašant semivariogramomis atsižvelgti į erdvinį trendą bei anizotropiją;
 - b) duomenis prognozuoti ordinariu, universaliu, medianinio išlyginimo krigingais ar kokrigingu;
 - c) taikyti kelis geriausio modelio ar prognozės rezultatų patikros metodus (klaidų kvadratų sumos, kryžminės patikros metodus, Akaike kriterijų).

Darbo rezultatų praktinė reikšmė

Darbe pateikta vieninga erdvinų duomenų su stacionariomis klaidomis modeliavimo metodika gali būti taikoma įvairaus pobūdžio erdvinų duomenų analizėje. Teoriniai tyrimų rezultatai gali būti naudojami ieškant optimalių imčių planų. Praktiniai tyrimų rezultatai gali būti pavyzdžiu kaip realizuoti apibendrintą erdvinų duomenų metodiką atviro kodo sistemos R įrankiais.

Ginamieji teiginiai

1. Pasiūlyta vieninga erdvinų duomenų modeliavimo per semivariogramas ir prognozavimo ordinariu, universaliu, medianinio išlyginimo krigingais ir kokrigingu metodika.
2. Stacionaraus Gauso duomenų modelio atvejui pateikta MoM semivariogramų įvertinio dispersijų-kovariacijų matricos išraiška per teorines semivariogramas.
3. Apibendrinto mažiausių kvadratų įvertinio išraiška per semivariogramas β parametrus užrašoma dispersijų-kovariacijų matricas atitinkamai pakeičiant semivariogramų matricomis.
4. β parametru apibendrinto mažiausių kvadratų įvertinio D ir A kriterijų funkcijų išraiškos išvestos per semivariogramas.
5. MoM semivariogramos įvertinio normališkumas atvejui, kai duomenų modelio klaidos yra stacionarios ir Gausinės.

6. Išvesta MoM semivariogramos įvertinio asimptotinės dispersijų-kovariacijų matricos išraiška Lahiri ir kt. (2002) apibrėžtam didėjančios erdvės asimptotikos erdviniam imčių planui izotropinės eksponentinės semivariogramos atveju.

Darbo rezultatų aprobavimas

Disertacijos tema yra atspausdintos 8 mokslinėse publikacijos: viena – mokslo žurnale, cituojamame Inspec duomenų bazėje (Krūminienė (Borisenko) 2006), penkios – recenzuojamuose mokslo žurnaluose (Krūminienė (Borisenko), Dučinskas, Garška 2003; Krūminienė (Borisenko), Garška 2004; Krūminienė (Borisenko), Dučinskas 2004; Krūminienė (Borisenko), Dučinskas 2004; Krūminienė (Borisenko) 2004), viena – recenzuojamoje tarptautinės konferencijos medžiagoje (Dučinskas, Borisenko 2009), viena – kitose tarptautinių ir respublikinių konferencijų medžiagose (Krūminienė (Borisenko) 2004).

Disertacijoje atliktų tyrimų rezultatai buvo paskelbti devyniose mokslinėse konferencijose Lietuvoje ir užsienyje:

- 9-oje magistrantų ir doktorantų konferencijoje „Informacinė visuomenė ir universitetinės studijos“, 2004 m., Kaune;
- Tarptautinėje konferencijoje „Mathematical Modelling and Analysis“ 2003 m., Trakuose, 2004 m., Jurmaloje (Latvija);
- Tarptautinėje konferencijoje „Technologijos mokslo darbai Vakarų Lietuvoje“, 2004 m., Klaipėdoje;
- Lietuvos matematikų draugijos konferencijoje, 2003 m., Vilniuje, 2004 m., Kaune, 2007 m., Vilniuje;
- Tarptautinėje šiaurės šalių matematinės statistikos konferencijoje (NORDSTAT), 2008 m., Vilniuje;
- Tarptautinėje taikomų stochastinių modelių ir duomenų analizės konferencijoje (ASMDA), 2009 m., Vilniuje.

Disertacijos struktūra

Disertacija sudaryta iš žymėjimų ir santrumpų sąrašo, įvado, keturių skyrių, išvadų, 4 priedų, naudotos literatūros sąrašo iš 106 pozicijų. Darbo pabaigoje pateikiamas straipsnių disertacijos tematika sąrašas.

Darbe atlikti skaičiavimai buvo dalinai remiami Lietuvos valstybinio mokslo ir studijų fondo (2004 m. tema „Statistinė erdvinių duomenų analizė“, reg. Nr. T-05258).

Erdvinių duomenų matematiniai modeliai

Skyriuje analizuojami erdvinių duomenų tipai, skaliariniai erdviniai procesai, pagrindinės jų charakteristikos. Nurodomi erdvinių duomenų vidurkio modeliai. Taip pat analizuojamas vienas pagrindinių ir dažnai praktikoje naudojamų sklaidos rodiklių – semivariograma – bei jos parametrai, nurodomi keli semivariogramos įvertiniai, pateikiami populiariausi semivariogramų parametriniai modeliai bei jų naudojimo rekomendacijos.

1.1. Erdviniai duomenys

Geostatistika nagrinėja stebinius erdvėje. Kiekvienas stebinys susijęs su tašku erdvėje. „Tašku“ vadiname atskirą tašką arba mažą sritį erdvėje. „Taškiniu stebiniu“ tame taške vadiname temperatūros, vietovės aukščio, vandens telkinio gylio matavimo rezultatą.

Terminas „Geografiškai orientuoti duomenys“ (*angl. Georeferenced Data*) naudojamas tam, kad apibrėžti kiekybinius duomenis, susietus su geografinėmis vietovės koordinatėmis. Jie gali būti išsidėstę dvimatėje ar trimatėje erdvėje, vaizduoti tam tikrus geologinius ar atmosferinius derinius. Visi erdviniai procesai gali būti aprašyti vieninga forma.

Tegu $\mathbf{s} \in \mathbb{R}^d$ yra d -matės Euklido erdvės taškas, kuriame atliekamas stebiny. Gali būti, kad $\mathbf{s} \equiv (x, y)$ yra taško padėtis dvimatėje erdvėje, arba $\mathbf{s} \equiv (x, y, z)$ – taško padėtis trimatėje erdvėje, arba $\mathbf{s} \equiv (\text{ilguma}, \text{platuma})$ ($\mathbf{s} \equiv (\text{ilguma}, \text{platuma}, \text{aukštis})$). Tiksliau tariant, stebinio koordinatės gali būti nurodytos Dekarto koordinatėse arba geografinėmis koordinatėmis (Dučinskas, Šaltytė-Benth 2003). Tačiau atstumą matuojant pagal ilgumos ir platumos koordinates atsiranda atstumo iškraipymų. Transponuojant geografines koordinates į Dekarto koordinates (pvz., naudojant UTM projekciją) absčių ašies kryptys sutampa su Rytų kryptimi ir vadinama „easting“, o ordinačių ašies kryptis sutampa su šiaurės kryptimi ir vadinama „northing“.

Laikoma, kad skaliarinio požymio $\mathbf{Z}$ stebiny erdvės taške $\mathbf{s} \in \mathbb{R}^d$ yra atsitiktinis dydis, pasiskirstęs pagal normalųjį (Gauso) skirstinį. Jį žymėsime $\mathbf{Z}(\mathbf{s})$. Matematinis erdvinių duomenų (erdvinės populiacijos) modelis yra erdvinis procesas su erdviu indeksu:

$$\{\mathbf{Z}(\mathbf{s}) : \mathbf{s} \in D\}, \text{ čia } D \subset \mathbb{R}^d. \quad (1.1)$$

Erdvinio proceso realizacija žymima $\{\mathbf{z}(\mathbf{s}), \text{ čia } \mathbf{s} \in D \subset \mathbb{R}^d\}$. Kai duomenų prognozė atliekama erdvėje, erdvinio proceso realizacija dažnai vadinama regionalizuotu kintamuoju (*angl. Regionalized Variable*).

Pagal Cressie (1993) yra išskiriami trys erdvinių duomenų tipai:

- geostatistiniai duomenys,
- gardelės duomenys (*angl. Lattice*) ir
- taškiniai vaizdai (*angl. Point Pattern*) (1.1 lentelė).

Geostatistiniai duomenys formuoja erdvinį paviršių. Modeliuojant gardelės duomenis rezultatas gaunamas suskaičiuojamam erdvės taškų kiekiui. Tarp šio tipo duomenų nėra prasmės atlikti interpoliacijos, nes duomenys nesudaro tolydaus paviršiaus.

Taškinio proceso duomenys skiriasi nuo pirmų dviejų erdvinių duomenų tipų tuo, kad aibė D , kurioje išsidėsčiusios įvykių matavimų vietos, yra atsitiktinė. Vienas pagrindinių geostatistinių ir gardelės duomenų analizės tikslų yra išanalizuoti $\mathbf{Z}(\mathbf{s})$ arba $E[\mathbf{Z}(\mathbf{s})]$ savybes, tuo tarpu nagrinėjant taškinius procesus vienas pagrindinių tikslų yra nustatyti tikimybinės aibės D savybes (Doraj-Raj 2001).

Nepriklausomai nuo to, kuris erdvinių duomenų tipas nagrinėjamas, naudojamas baigtinis duomenų rinkinys. Jo pagrindu sudaromas modelis ir pateikiamos išvados (ver Hoef 2002). Tokią erdvinių duomenų klasifikaciją pirmasis pateikė Cressie (1993). Šiame disertaciniame darbe bus modeliuojami arba teoriškai nagrinėjami geostatistiniai ir gardeliniai duomenys. Juos

generuojančius atsitiktinius procesus vadinsime **erdviniais procesais** (angl. *Spatial Process*).

1.1 lentelė. Erdvinių duomenų tipai

Table 1.1. Spatial data types

Erdvinių duomenų tipai	Duomenis generuojančių procesų apibūdinimas	Tikslas	Procesų pavyzdžiai
Geostatistiniai duomenys	• D – tolydus fiksuotas $\mathcal{R}^d$ erdvės poaibis; • s – gali būti tolydus poaibyje D; • $Z(s) \in \mathcal{R}^d$.	Prognuoti Z reikšmes erdvės taškuose.	• Temperatūrų reikšmės; • aukščio reikšmės tam tikrose srityse taškuose; • tam tikro ežero užterštumo koncentracija.
Gardelės duomenys	• D – fiksuotas, dažniausiai baigtinis taisyklingai arba skaitlus $\mathcal{R}^d$ erdvėje išsidėsčiusių taškų rinkinys; • $Z(s)$ atsitiktinis vektorius (gali būti daugiamatis) erdvės taške s.	Modeliuoti pasiskirstymų junginį, ryšius tarp kintamųjų arba stebinių skirtumus taškuose.	• Gyvūnų kiekis gretimose srityse; • vaisių kiekis medyje; • nelaimingų įvykių kiekis tam tikrame kelio ruože.
Taškiniai duomenys	• D – $\mathcal{R}^d$ atsitiktinis erdvės taškų, kuriuose įvyko tam tikro tipo įvykis, rinkinys; • $Z(s)$ atsitiktinis arba neatsitiktinis vektorius (gali būti daugiamatis) erdvės taške $s \in D$.	Klasterizavimas (Sain 2004)	• Medžių išsidėstymas miške; • žvaigždžių išsidėstymas danguje; • žmonių, kuriems vystosi plaučių vėžys, vietos; • gyvūnų gimstamumo vietos.

1.2. Skaliariniai erdviniai procesai ir jų geostatistinės charakteristikos

Šiame darbe nagrinėsime adityvų erdvinio $\{Z(s) : s \in D, \text{čia } D \subset \mathbb{R}^d\}$ stebinio taške s modelį:

$$Z(s) = \mu(s) + \varepsilon(s), \quad (1.2)$$

čia $\mu(s) \equiv E\{Z(s)\}$ – vidurkio funkcija (erdvinis trendas, poslinkis (*angl. Drift*)), $\varepsilon(s)$ – atsitiktinė klaida. Kitos pagrindinės modelio (1.2) charakteristikos yra dispersija $\sigma^2(s) = \text{var}(Z(s)) = E[Z(s) - \mu(s)]^2$ taške s bei kovariacinė funkcija $C(s_1, s_2) = \text{cov}(Z(s_1), Z(s_2)) = E[(Z(s_1) - \mu(s_1))(Z(s_2) - \mu(s_2))]$ tarp stebinių taškuose s_1 ir s_2 .

Atsitiktinės klaidos matematinis modelis aibėje D yra erdvinis procesas $\{\varepsilon(s) : s \in D\}$ su nuliniu vidurkiu ir kovariacine funkcija $\text{cov}(\varepsilon(s_1), \varepsilon(s_2))$.

1.1 apibrėžimas. Erdvinis procesas $\{Z(s) : s \in D, \text{čia } D \subset \mathbb{R}^d\}$ vadinamas **griežtai stacionariu** (pirmos eilės stacionariu) (*angl. Strong Stationarity*), jei bendras stebinių $\{Z(s_1), Z(s_2), \dots, Z(s_n)\}$ skirstinys toks pat kaip ir $\{Z(s_1 + h), Z(s_2 + h), \dots, Z(s_n + h)\}$, $\forall s_1, s_n$ ir $\forall h$. T. y. visas stebinio matavimų vietas pastūmus ta pačia kryptimi ir tuo pačiu atstumu proceso reikšmių skirstinys ir tikimybinės charakteristikos nesikeičia. Tai klasikinės statistikos termino „vienodai pasiskirstę ir nepriklausomi“ analogas.

Jei $\{Z(s) : s \in D, \text{čia } D \subset \mathbb{R}^d\}$ yra griežtai stacionarus erdvinis procesas, tai $E\{Z(s)\} = \mu$ visiems s .

Griežtas stacionarumas dažniai sunkiai įgyvendinamas, kadangi vidurkis skirtingose srityse dažnai skiriasi arba duomenyse yra akivaizdus trendas (Blyth, Cake *et al.* 2008).

1.2 apibrėžimas. Erdvinis procesas $\{Z(s) : s \in D, \text{čia } D \subset \mathbb{R}^d\}$ vadinamas **stacionariu** (antros eilės stacionariu, silpnai stacionariu) (*angl. Second-Order Stationarity*), jei:

- $\mu(s) \equiv \mu$ visiems $s \in D$,
- $C(s_i, s_j) = C(s_i - s_j) = \text{cov}(Z(s_i), Z(s_j))$ visiems $s_i, s_j \in D$.

Pagal apibrėžimą, $C(0)$ yra dispersija σ^2 . Čia 0 – nulinis vektorius.

Iš griežto stacionarumo išplaukia antros eilės stacionarumas (bet ne atvirkščiai), tačiau Gauso erdvinio proceso atveju šie stacionarumai ekvivalentūs.

Dažnai šio tipo stacionarumas taip pat nesunkiai įgyvendinamas, nes taikymuose pastebėta, kad didėjant sričiai, kovariacinė funkcija dažniausiai taip pat didėja.

1.3 apibrėžimas. Erdvinis procesas vadinamas **vidiniai stacionariu** (*angl. Intrinsic Stationarity*), jei:

- $\mu(\mathbf{s}) \equiv \mu$ visiems $\mathbf{s} \in D$,
- variograma $2\gamma(\mathbf{s}_i, \mathbf{s}_j) = 2\gamma(\mathbf{s}_i - \mathbf{s}_j) = \text{var}(Z(\mathbf{s}_i) - Z(\mathbf{s}_j))$ visiems

$\mathbf{s}_i, \mathbf{s}_j \in D$.

Semivariograma (kai kur literatūroje lietuvių kalba dar vadinama pusvariograma (Blyth, Cake *et al.* 2008)) tai pusė variogramos ir ji žymima:

$$\gamma(\mathbf{h}) = \gamma(\mathbf{s}_i - \mathbf{s}_j) = \frac{\text{var}(Z(\mathbf{s}_i) - Z(\mathbf{s}_j))}{2} = E\{[Z(\mathbf{s}_i) - Z(\mathbf{s}_i + \mathbf{h})]^2\}. \quad (1.3)$$

Iš griežto stacionarumo ir antros eilės stacionarumo išplaukia vidinis stacionarumas, bet ne atvirkščiai.

1.4 apibrėžimas. Jeigu $D = R^n$ tai stacionaraus erdvinio proceso spektrinio tankio funkcija $g(w)$ visiems $w \in R^n$ apibrėžiama tokiu būdu:

$$g(w) = (2\pi)^{-n} \int_{R^n} \cos(\mathbf{h}w) c(\mathbf{h}) d\mathbf{h}. \quad (1.4)$$

Jei erdvinis procesas yra antros eilės stacionarus, tai semivariograma ir kovariacinė funkcija išsiskaičiuoja viena per kitą:

$$\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h}) \quad (1.5)$$

arba

$$\gamma(\mathbf{h}) = \sigma^2 - C(\mathbf{h}). \quad (1.6)$$

Prieš tai pateikta lygtis rodo, kad jei kovariacinė funkcija yra žinoma, tai ir semivariograma yra žinoma. Praktikoje semivariogramos yra naudojamos dažniau nei kovariacinės funkcijos, nes tam, kad naudoti semivariogramą, nereikia žinoti vidurkio reikšmės.

Erdvinė autokoreliacija tai matas, kuris nusako kaip kintamasis koreliuoja pats su savimi atsižvelgiant į jo padėtį erdvėje. Erdvinės autokoreliacijos priežastys detaliau aptartos 2.2. skyrelyje.

Stacionariems procesams ryšys tarp autokoreliacijos ir kovariacinės funkcijos aprašomas tokia formule $\rho(\mathbf{h}) = \frac{C(\mathbf{h})}{C(\mathbf{0})}$.

Vidiniai stacionarūs procesai skirstomi į izotropinius ir anizotropinius.

1.5 apibrėžimas. Vidinis stacionarus procesas yra izotropinis (*angl. Isotropy*), jei $\gamma(\mathbf{s}_i, \mathbf{s}_j) = \gamma_0(\|\mathbf{s}_i - \mathbf{s}_j\|)$, t. y. jo pokytis priklauso tik nuo atstumo tarp taškų ir nepriklauso nuo krypties.

Anizotropija (*angl. Anisotropy*), kai dėl erdvinio proceso autokoreliacija viena kryptimi didesnė negu kita kryptimi (Blyth, Cake *et al.* 2008; Cressie 1993).

Yra žinoma, kad stacionarių Gauso erdvinių procesų įvertinimui be stacionarumo sąlygos būtina ir pakankama ergodiškumo sąlyga. Stacionarus erdvinis procesas vadinamas ergodišku (*angl. Ergodicy*), jei momentas visos srities atžvilgiu gali būti gautas erdviniais empiriniais momentais.

Ergodiškų erdvinių procesų atveju $C(\mathbf{h}) \rightarrow 0$, kai $|\mathbf{h}| \rightarrow \infty$, todėl $C(\mathbf{0})$ gaunamas kaip $\gamma(\mathbf{h})$ riba, kai $|\mathbf{h}| \rightarrow \infty$. Tačiau bendru atveju tokia riba nebūtinai turi egzistuoti. Jei $\lim_{|\mathbf{h}| \rightarrow \infty} \gamma(\mathbf{h})$ egzistuoja, erdvinis procesas yra antros eilės stacionarus su kovariacine funkcija $C(\mathbf{h}) = \lim_{|\mathbf{h}| \rightarrow \infty} \gamma(\mathbf{h}) - \gamma(\mathbf{h})$, tačiau, jei ši riba neegzistuoja, tai erdvinis procesas nėra antros eilės stacionarus (Dučinskas, Šaltytė-Benth 2003).

1.3. Vidurkio modeliai ir jų parametrų įvertiniai

Nagrinėjamas skaliarinis erdvinis procesas. Požymio $\mathbf{Z}$ matavimų erdvės taškuose $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n \in D$ (čia $D \subset \mathbb{R}^d$) modelis:

$$\mathbf{Z}_n = \boldsymbol{\mu}_n + \boldsymbol{\varepsilon}_n, \quad (1.7)$$

čia:

- požymio vektorius $\mathbf{Z}_n = (Z(\mathbf{s}_1), Z(\mathbf{s}_2), \dots, Z(\mathbf{s}_n))^T$,
- vidurkio vektorius $\boldsymbol{\mu}_n = (\mu(\mathbf{s}_1), \mu(\mathbf{s}_2), \dots, \mu(\mathbf{s}_n))^T$,
- klaidų vektorius $\boldsymbol{\varepsilon}_n = (\varepsilon(\mathbf{s}_1), \varepsilon(\mathbf{s}_2), \dots, \varepsilon(\mathbf{s}_n))^T$.

Dažniausiai išskiriami du vidurkio modeliai:

- pastovaus vidurkio;
- nepastovaus vidurkio (tendo paviršiaus, regresinis).

Pastovaus vidurkio atveju vidurkio funkcija yra pastovi, t. y. $E\{\mathbf{Z}(\mathbf{s})\} = \boldsymbol{\mu} = \text{const}$. Tuomet $\mathbf{Z}_n$ vidurkių vektorius:

$$\boldsymbol{\mu}_n = \boldsymbol{\mu} \mathbf{1}_n, \quad (1.8)$$

čia $\mathbf{1}_n = (1_1, 1_2, \dots, 1_n)^T$ – vienetinis n - matis vektorius.

Nepastovaus vidurkio atveju, vidurkis nėra pastovus srityje D ir tiesiškai priklauso nuo žinomų funkcijų $\{f_q(\mathbf{s}), q = 1, \dots, Q\}$. Tada vidurkio funkcija $\mu(\mathbf{s})$ turi pavidalą:

$$\mu(\mathbf{s}) = \mathbf{f}^T(\mathbf{s})\boldsymbol{\beta}, \quad (1.9)$$

čia $\mathbf{f}(\mathbf{s}) = (f_0(\mathbf{s}), \dots, f_Q(\mathbf{s}))^T$, o $\boldsymbol{\beta} = (\beta_0, \dots, \beta_Q)^T$ – modelio parametrų vektorius. Jeigu $f(s)$ yra $\mathbf{s}$ koordinačių polinamai, tai turime trendo paviršiaus modelį:

a) tiesinio trendo atveju:

$$f_1(\mathbf{s}) = 1, f_2(\mathbf{s}) = x, f_3(\mathbf{s}) = y, \text{ kai } Q = 3,$$

b) kvadratinio trendo atveju:

$$f_1(\mathbf{s}) = 1, f_2(\mathbf{s}) = x, f_3(\mathbf{s}) = y, f_4(\mathbf{s}) = x^2, f_5(\mathbf{s}) = y^2, f_6(\mathbf{s}) = xy, \text{ kai } Q = 6.$$

Regresinio modelio atveju $\{f_q(\mathbf{s}), q = 1, \dots, Q\}$ yra regresoriai (nepriklausomi kintamieji). Disertaciniame darbe nagrinėjami regresiniai modeliai, kai $f_0(\mathbf{s}) = 1$. Tuomet abiem nepastovaus vidurkio atvejais stebinių n taškuose $(\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n)$ vidurkio modelis vektoriniu pavidalu bus

$$\boldsymbol{\mu}_n = \mathbf{X}\boldsymbol{\beta}, \quad (1.10)$$

čia

$$\mathbf{X}_{[n \times Q]} = \begin{pmatrix} 1 & f_2(\mathbf{s}_1) & \cdots & f_Q(\mathbf{s}_1) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & f_2(\mathbf{s}_n) & \cdots & f_Q(\mathbf{s}_n) \end{pmatrix} = (\mathbf{1}_n, \mathbf{F}). \quad (1.11)$$

Nepastovaus vidurkio atveju dažnai $\boldsymbol{\beta}$ nėra žinomas. Paprastu mažiausių kvadratų (PMKM) metodu (jis bus aprašytas 2.3.2 skyrelyje) gaunamas $\boldsymbol{\beta}$ įvertinys ir jo dispersijų-kovariacijų matrica yra tokia:

$$\hat{\boldsymbol{\beta}}_{\text{PMKM}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Z}_n, \\ \mathbf{V}(\hat{\boldsymbol{\beta}}_{\text{PMKM}}) = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1},$$

o apibendrintu mažiausių kvadratų (AMKM) metodu gaunamas $\boldsymbol{\beta}$ įvertinys ir dispersijų-kovariacijų matrica yra:

$$\hat{\boldsymbol{\beta}}_{\text{AMKM}} = (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{Z}_n,$$

$$\mathbf{V}(\hat{\boldsymbol{\beta}}_{\text{AMKM}}) = (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1},$$

čia $\boldsymbol{\Sigma}$ žymi $\mathbf{Z}_n$ dispersijų-kovariacijų $n \times n$ eilės matricą, kurios pavidalas stacionaraus erdvinio proceso atveju yra:

$$\mathbf{V}(\mathbf{Z}_n) = \boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 & C(\mathbf{s}_1 - \mathbf{s}_2) & C(\mathbf{s}_1 - \mathbf{s}_3) & \dots & C(\mathbf{s}_1 - \mathbf{s}_n) \\ C(\mathbf{s}_2 - \mathbf{s}_2) & \sigma^2 & C(\mathbf{s}_2 - \mathbf{s}_3) & \dots & C(\mathbf{s}_2 - \mathbf{s}_n) \\ C(\mathbf{s}_3 - \mathbf{s}_1) & C(\mathbf{s}_3 - \mathbf{s}_2) & \sigma^2 & \dots & C(\mathbf{s}_3 - \mathbf{s}_n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C(\mathbf{s}_n - \mathbf{s}_1) & C(\mathbf{s}_n - \mathbf{s}_2) & C(\mathbf{s}_n - \mathbf{s}_3) & \dots & \sigma^2 \end{pmatrix}. \quad (1.12)$$

Anksčiau minėtų PMKM ir AMKM metodų atvejais daroma prielaida, kad kovariacinė funkcija $C(\cdot)$ yra žinoma.

Dabar aprašysime vidurkio modelius vektorinio erdvinio proceso atveju. Šioje situacijoje stebima $p \geq 2$ priklausomų kintamųjų, tuomet modelio (1.7) komponentai visiems $\mathbf{s}_i \in D$, $i = 1, \dots, n$ perrašomi tokiu principu:

- požymio vektorius $\mathbf{Z}(\mathbf{s}_i) = (Z_1(\mathbf{s}_i), Z_2(\mathbf{s}_i), \dots, Z_p(\mathbf{s}_i))^T$,
- vidurkio vektorius $\boldsymbol{\mu}(\mathbf{s}_i) = (\mu_1(\mathbf{s}_i), \mu_2(\mathbf{s}_i), \dots, \mu_p(\mathbf{s}_i))^T$,
- klaidų vektorius $\boldsymbol{\varepsilon}(\mathbf{s}_i) = (\varepsilon_1(\mathbf{s}_i), \varepsilon_2(\mathbf{s}_i), \dots, \varepsilon_p(\mathbf{s}_i))^T$.

Jeigu naudojamas *pastovaus vidurkio modelis*, tai visiems $\mathbf{s} \in D$:

$$\boldsymbol{\mu}(\mathbf{s}) = \mu \mathbf{1}_p, \quad (1.13)$$

o jei naudojamas *nepastovaus vidurkio modelis*, tai:

$$\boldsymbol{\mu}(\mathbf{s}) = \left(\sum_{q=0}^Q \beta_{q1} f_q(\mathbf{s}), \sum_{q=0}^Q \beta_{q2} f_q(\mathbf{s}), \dots, \sum_{q=0}^Q \beta_{qp} f_q(\mathbf{s}) \right)^T. \quad (1.14)$$

Vektorinio erdvinio proceso modelis bus naudojamas tik aprašant kokrigingo prognozės metodą disertacijos skyrelyje 2.4.4.

1.4. Semivariogramų parametrai ir jų modeliai

Tarkime, kad nagrinėjamas stacionarus erdvinis procesas $\mathbf{Z}(\mathbf{s})$ su žinomu vidurkiu $\boldsymbol{\mu}(\mathbf{s}) = \boldsymbol{\mu}$ ir dispersija $C(\mathbf{0}) = \sigma^2$. Stacionaraus erdvinio proceso atveju semivariograma yra dviejų duomenų reikšmių tikėtinas kvadratinis skirtumas,

aprašomas atskyrimo vektoriumi arba vėlinimo vektoriumi $\mathbf{h}$ (formulė (1.3)). Empirines semivariogramas galima skirstyti į:

- visakryptes (kai semivariogramai skaičiuoti svarbus tik atstumas tarp stebinių) (izotropinio semivariogramos modelio atvejis);
- kryptines (kai semivariogramai skaičiuoti svarbus atstumas ir kryptis tarp stebinių) (anizotropinio semivariogramos modelio atvejis).

Smulkiau panagrinėsime izotropinio semivariogramos modelio atvejį.

Visakryptę semivariogramą γ (1.1 pav.) apibūdina keli parametrai:

- *Grynuolis* – yra semivariogramos reikšmė, kai atstumas tarp taškų artėja prie nulio. Jis atitinka reikšmių pokytį taške, kurio nepaaiškina erdvinė struktūra. Grynuolį galima išskaidyti į matavimų paklaidą ir paklaidą, atsiradusią dėl pokyčio mikromasteliu. Šiame darbe jis žymimas θ_0 .
- *Slenkstis* – semivariogramos reikšmė, ties kuria semivariograma įgauna maksimumą ir tampa pastovia (jei taip įvyksta), t. y. išnyksta statistinė priklausomybė tarp matavimo taškų. Slenkstį dažnai sudaro dvi dalys: trūkis pradžioje, vadinamas *grynuolio* reiškiniu, ir *dalinis slenkstis* (angl. *Partial Sill*). Abi dalis sudėjus gaunamas *slenkstis* (žymėsime $\theta_0 + \theta_1 = \sigma^2$).
- *Slenksčio atstumas* – tai atstumas, ties kuriuo semivariograma įgauna maksimumą ir tampa pastovia (jei taip įvyksta). Čia žymima θ_2 raide.

Tam, kad apskaičiuoti visakryptės semivariogramos taškus atliekama tokia veiksmų seka:

- paskaičiuojamas atstumas tarp stebinių porų (žinoma, kad tam tikros stebinių poros kartosis);
- gautos stebinių poros grupuojamos į k klases pagal atstumą tarp jų (pvz., pirma klasė – taškai iki 100 metrų, antra klasė – 100–200 metrų ir t. t.);

Stacionariam erdviniam procesui semivariogramos įvertinys grindžiamas momentų metodu (MoM) (pagal Matheron 1963) skaičiuojamas kiekvienai taškų klasei atskirai tokia formule:

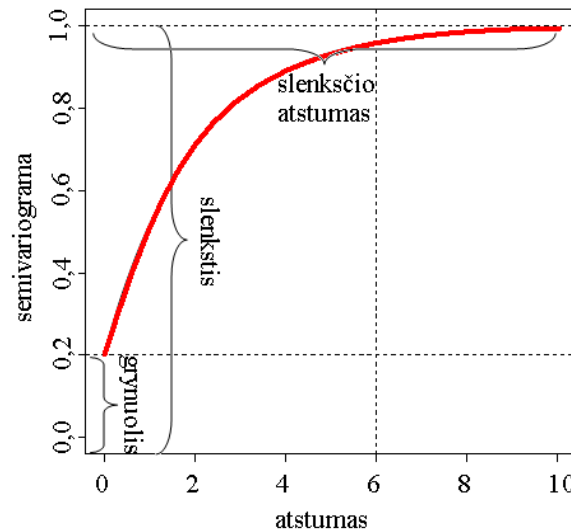
$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2|N(\mathbf{h})|} \sum_{(\mathbf{s}_i, \mathbf{s}_j) \in N(\mathbf{h})} [Z(\mathbf{s}_i) - Z(\mathbf{s}_j)]^2, \quad (1.15)$$

čia $|N(\mathbf{h})|$ nurodo aibės $N(\mathbf{h})$ porų skaičių, $\mathbf{s}_i - \mathbf{s}_j = \mathbf{h}$.

Izotropiniam stacionariam erdviniam procesui semivariogramos įvertinys skaičiuojamas tokia formule:

$$\hat{\gamma}_0(\mathbf{h}) = \frac{1}{2|N_0(\mathbf{h})|} \sum_{(\mathbf{s}_i, \mathbf{s}_j) \in N_0(\mathbf{h})} [Z(\mathbf{s}_i) - Z(\mathbf{s}_j)]^2. \quad (1.16)$$

Praktikoje dažniau sutinkamas atvejis, kai matavimo taškai nesudaro gardelės. Tuomet taikant (1.15) formulę $N(\mathbf{h}) = \{(\mathbf{s}_i, \mathbf{s}_j) : \mathbf{s}_i - \mathbf{s}_j = \mathbf{h}; \text{čia } \mathbf{h} \in T(\mathbf{h}_k); i \neq j; i, j = 1, \dots, n\}$, $T(\mathbf{h}_k)$ – atskyrimo žingsnis (nedidelė sritis apie $\mathbf{h}$), o $k = 1, \dots, K$. „Paprastai atskyrimo žingsnis pasirenkamas toks, kad jo ir žingsnių skaičiaus sandauga būtų apytiksliai lygi pusei didžiausio atstumo tarp matuojamų taškų“ (Blyth, Cake *et al.* 2008)¹. Taip siekiama išvengti „krašto problemos“, kai nagrinėjamos srities kraštuose randama mažiau stebinių nei srities viduje arba nagrinėjama tyrimų sritis yra tik dalis geografinio regiono, kuriame vyksta nagrinėjamas erdvinis procesas. Pastaruoju atveju didesnio geografinio regiono viduje vykstantys įvykiai gali turėti įtakos įvykiams tiriamos srities viduje. Kraštų įtaką galima įvertinti įvairiais metodais, kurie detaliau aprašyti Diggle (2003).



1.1 pav. Visakryptės semivariogramos forma ir jos parametrai
Fig. 1.1. The form and parameters of omnidirectional semivariogram

¹ *Gstat* pakete maksimalus lagas lygus 1/3 maksimaliams pločiui ir pagal nutylėjimą sudaroma 15 vienodo atstumo lagų.

Taip pat buvo pasiūlyti ir kiti erdvinės sklaidos charakteristikų empiriniai įvertiniai: madograma, radiograma, standartizuota semivariograma ir kt. (Deutsch ir Journel 1998).

Empirinės semivariogramos įverčio paklaida priklauso nuo daugelio faktorių:

- vidinės hipotezės stacionarumo įgyvendinimo;
- klasių parinkimo ir porų kiekio klasėse ir t. t.

Dažnai sudėtinga vizualiai apibrėžti klases. Tačiau:

- skaičiavimuose reikia naudoti pakankamai klasių, tam, kad pakankamai gerai vizualizuoti apskaičiuotos semivariogramos elgesį;
- kuo daugiau skaičiavimuose naudojama stebinių porų, tuo rezultatai yra tikslesni.

Journel ir Huijbregts (1978) siūlo tam tikros klasės semivariogramos reikšmės apskaičiavimui naudoti 30–50 stebinių porų, tačiau dauguma patikslina, kad jei naudojama mažiau nei 50 taškų porų, gauti rezultatai yra nepatikimi, o esant didesniai nei 250 stebinių kiekiui – galima beveik neabejoti rezultatų patikimumu. Webster ir Oliver (1992) pateikia puikius pavyzdžius, kokie nepatikimi gali būti semivariogramos rezultatai, kai jos reikšmių apskaičiavimui naudojama nepakankamai stebinių porų. Didesnis stebinių kiekis ypač reikalingas bandant nustatyti duomenų anizotropiją. Taip pat rekomenduojama sudaryti ne mažiau kaip 10 klasių. Sudarant klases, patartina pirmos klasės riba nustatyti nulį.

Klasikinį MoM semivariogramos įvertinį, pateiktą formulėje (1.15), galima užrašyti taip:

$$\hat{\gamma}(\mathbf{h}) = \frac{\mathbf{Z}^T \mathbf{L}(\mathbf{h}) \mathbf{Z}}{2|\mathbf{N}(\mathbf{h})|}, \quad (1.17)$$

čia $\mathbf{L}(\mathbf{h})$ – grafų teorijoje gerai žinoma *Laplaso* matrica, kuri erdvinėje statistikoje vadinama erdvinio dizaino matrica (*angl. Spatial Design Matrix*) (Hillier, Martellosio 2006).

Nagrinėjamoje situacijoje *Laplaso* matricos $\mathbf{L}(\mathbf{h})$ elementai $L_{ij}(\mathbf{h})$ apibūdinami tokiu būdu:

- $L_{ij}(\mathbf{h}) = 0$, jei pora $(\mathbf{s}_i, \mathbf{s}_j) \notin N(\mathbf{h})$, $i \neq j$;
- $L_{ii}(\mathbf{h}) = 0$, jei $(\mathbf{s}_i, \mathbf{s}_j) \notin N(\mathbf{h})$, visiems $j = 1, \dots, n$ ir $i \neq j$;
- $L_{ij}(\mathbf{h}) = -1$, kai $i \neq j$ ir $(\mathbf{s}_i, \mathbf{s}_j) \in N(\mathbf{h})$;

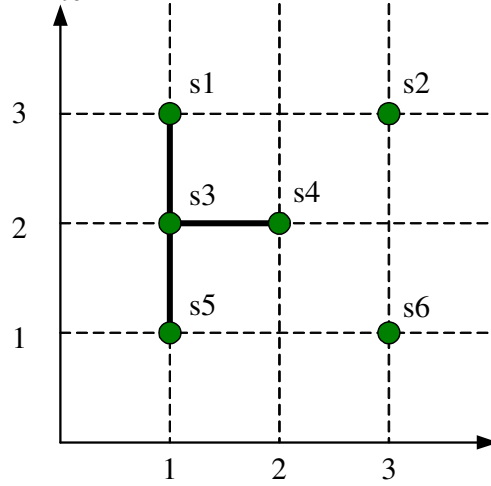
d) $L_{ii}(\mathbf{h}) = m_i(\mathbf{h})$, jei egzistuoja bent vienas $\mathbf{s}_j (j \neq i)$ toks, kad pora $(\mathbf{s}_i, \mathbf{s}_j) \in N(\mathbf{h})$, čia $m_i(\mathbf{h})$ – nurodo tokių porų skaičių.

Akivaizdu, kad šiai matricai galioja tokios savybės:

a) $\sum_{j=1}^n L_{ik} = \sum_{j=1}^n L_{il} = 0$, visiems $k, l = 1, \dots, n$;

b) $\sum_{i=1}^n m_i(\mathbf{h}) = \sum_{i=1}^n L_{ii}(\mathbf{h}) = N(\mathbf{h})$.

Pvz., tarkime, kad 6 taškai išsidėstę Euklido plokštumoje taip kaip pavaizduota 1.2 paveikslėlyje.



1.2 pav. Taškų išsidėstymas Euklido plokštumoje
Fig. 1.2. Locations of points in Euclidean plane

Imdami $\mathbf{h} = (0, 1)$ pavaizduosime *Laplaso* matricą anizotropiniu atveju:

$$L(\mathbf{h}) = \begin{pmatrix} 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 2 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Panaudojus *Laplaso* matricą nesunku parodyti, kad įvertinys $\hat{\gamma}(\mathbf{h})$ yra nepaslinktas, nes:

$$E\left(\hat{\gamma}(\mathbf{h})\right) = \text{tr}(L(\mathbf{h})\Sigma) / 2N(\mathbf{h}) = 2N(\mathbf{h}) \frac{(C(\mathbf{0}) - C(\mathbf{h}))}{2N(\mathbf{h})} = \gamma(\mathbf{h}). \quad (1.18)$$

1 lema. Tarkime, kad $\mathbf{h}_1, \mathbf{h}_2 \in D$ ir $\mathbf{Z} \sim N_n(0, \Sigma)$, tada:

$$\begin{aligned} \text{cov}\left\{\hat{\gamma}(\mathbf{h}_1), \hat{\gamma}(\mathbf{h}_2)\right\} = \\ \frac{1}{2|N(\mathbf{h}_1)||N(\mathbf{h}_2)|} \cdot \sum_{(\mathbf{s}_{i1}, \mathbf{s}_{j1}) \in N(\mathbf{h}_1)} \sum_{(\mathbf{s}_{i2}, \mathbf{s}_{j2}) \in N(\mathbf{h}_2)} [\gamma(\mathbf{s}_{i1} - \mathbf{s}_{j2}) + \gamma(\mathbf{s}_{i2} - \mathbf{s}_{j1}) \\ - \gamma(\mathbf{s}_{i1} - \mathbf{s}_{i2}) - \gamma(\mathbf{s}_{j1} - \mathbf{s}_{j2})]^2. \end{aligned} \quad (1.19)$$

Įrodymas. MoM įvertinius taškuose $\mathbf{h} = \mathbf{h}_k$, $k = 1, 2$ pažymėkime tokiu būdu:

$$\hat{\gamma}(\mathbf{h}_k) = \frac{1}{2|N(\mathbf{h}_k)|} \sum_{(\mathbf{s}_{ik}, \mathbf{s}_{jk}) \in N(\mathbf{h}_k)} (Z(\mathbf{s}_{jk}) - Z(\mathbf{s}_{ik}))^2. \quad (1.20)$$

Naudodamiesi (1.17) ir daugiamačių Gauso skirstinių savybėmis gauname:

$$\text{cov}\left(\hat{\gamma}(\mathbf{h}_j), \hat{\gamma}(\mathbf{h}_k)\right) = \frac{\text{cov}(\mathbf{Z}^T L(\mathbf{h}_j) \mathbf{Z}, \mathbf{Z}^T L(\mathbf{h}_k) \mathbf{Z})}{4N_j(\mathbf{h})N_k(\mathbf{h})} = \frac{\text{tr}(L(\mathbf{h}_j)\Sigma L(\mathbf{h}_k)\Sigma)}{2N_j(\mathbf{h})N_k(\mathbf{h})}.$$

Pasinaudojus aukščiau apibrėžtomis *Laplaso* matricos savybėmis gauname:

$$\begin{aligned} \text{tr}(L(\mathbf{h}_1)\Sigma L(\mathbf{h}_2)\Sigma) = \\ (C(\mathbf{s}_{i1} - \mathbf{s}_{j2}) + C(\mathbf{s}_{i2} - \mathbf{s}_{j1}) - C(\mathbf{s}_{i1} - \mathbf{s}_{i2}) - C(\mathbf{s}_{j1} - \mathbf{s}_{j2}))^2. \end{aligned} \quad (1.21)$$

Pasinaudojus ryšio tarp semivariogramos ir kovariacijos lygtimi (1.5) užbaigiame lemos įrodymą.

1 lemoje, pasinaudojus *Laplaso* matricos savybėmis, Gauso atvejui dispersijų-kovariacijų matricos tarp MoM įvertinio reikšmių skirtingiems vėlinimo vektoriams formulė išvedama per semivariogramas atvejui, kai imties dydis yra baigtinis.

Geostatistikoje buvo pasiūlyti ir kiti semivariogramos įvertiniai (Cressie, Hawkins 1980; Genton 1998). Cressie ir Hawkins semivariogramos įvertinys:

$$\hat{\gamma}_{CH}(\mathbf{h}) = \frac{\left\{ \frac{1}{2|N(\mathbf{h})|} \sum_{(s_i, s_j) \in N(\mathbf{h})} [Z(s_i) - Z(s_j)]^{1/2} \right\}^4}{0,457 + \frac{0,494}{|N(\mathbf{h})|}}. \quad (1.22)$$

Cressie ir Howkins įdėja buvo gauti simetrišką skirstinį transformuojant duomenis su prielaida, kad $[Z(s_i + \mathbf{h}) - Z(s_i)]^2$ skirstinys būtų artimas Gauso skirstiniui.

Šie įvertiniai nėra robastiški išskirtims, nes įvertinys $\hat{\gamma}(\mathbf{h}) \rightarrow \infty$, kai nors vienas $Z(\mathbf{s}) \rightarrow \infty$. Tad Cressie-Howkins taip pat pasiūlė medianinę semivariogramą (Cressie 1993):

$$\hat{\gamma}_{Me}(\mathbf{h}) = \frac{\left\{ \text{Median} \sum_{(s_i, s_j) \in N(\mathbf{h})} [Z(s_i) - Z(s_j)]^{1/2} \right\}^4}{0,457}. \quad (1.23)$$

Matheron (MoM), Cressie ir Howkins įvertinių savybės palygintos Roustant, Dupuy, Helbert (2007).

Empiriniai semivariogramos taškai tiesiogiai nėra naudojami tolesniuose geostatistikos etapuose: parametrinis semivariogramos modelis yra priderinamas prie empirinių semivariogramos taškų. Empirinė semivariograma tolesnėje geostatistinėje analizėje negali būti naudojama dėl daugelio priežasčių:

1. Semivariogramos funkcija reikalinga visam atstumui ir visomis vektoriaus $\mathbf{h}$ kryptimis aplinkos ribose, tačiau empirinei semivariogramai apskaičiuojamos reikšmės tam tikruose atstumo intervaluose ir tam tikromis kryptimis. Tad būtina interpoliuoti semivariogramos funkciją $\mathbf{h}$ reikšmėms, kur nėra daug duomenų.
2. Taip pat yra būtina atsižvelgti į geologinę informaciją – anizotropiją, trendą, imties sudarymo klaidas ir t. t. ir tą informaciją įtraukti į erdvinę koreliaciją, kad semivariograma geriausiai atvaizduotų tikrąją geologinį pokytį.
3. Naudojamas izotropinis semivariogramos modelis turi tenkinti neteigiamo apibrėžtumo sąlygą. Ši sąlyga užtikrina, kad krigingo lygtys bus išspręstos, o gauta krigingo dispersija bus teigiama (Gringarten, Deutsch 2001).

Tarkim $\gamma(\cdot)$ yra antros eilės stacionaraus proceso semivariograma. Ji turi būti sąlyginai neigiamai apibrėžta (*angl. Conditionally Negative Definite*) funkcija (žr. Cressie 1993, psl. 86), t. y. jei $a_1, a_2, \dots, a_n$ yra konstantos ir $\sum a_i = 0$, tai

$$\sum_i \sum_j a_i a_j \gamma(\mathbf{s}_i - \mathbf{s}_j) \leq 0. \quad (1.24)$$

Dėl šių priežasčių geostatistikoje naudojami parametriniai semivariogramos modeliai, aprašomi neigiamai apibrėžtomis funkcijomis, tokiomis kaip sferinė, eksponentinė, *Gauso* ir t. t. (Christakos 1984, Cressie 1993, Journel, Huijbregts 1978). Bet kuri neigiamai apibrėžta semivariogramos funkcija gali būti naudojama, taip pat kaip ir semivariogramos ar kovariacijos reikšmių lentelė, tačiau pasirinktos funkcijos naudojimas reikalauja pirminio tos funkcijos teigiamo apibrėžtumo patikrinimo (Myers 1991).

Kai kurie autoriai (Rossiter 2004) parametrinius semivariogramų modelius skirsto į tokias grupes:

1. Semivariogramos su „švarių“ grynoliu (nėra jokios erdvinės struktūros): $\gamma_0(t) = \theta_0$, čia $t = |\mathbf{h}|$. Šiuo atveju interpoliacija negalima, imties vidurkis aprašo kiekvieno duomenų taško reikšmę.

2. *Neribojami* modeliai, kur dispersija didėja kartu su sritimi. Šio tipo semivariogramose gali nebūti slenksčio arba slenksčio atstumo, spėjama, kad sritis yra mažesnė nei modeliuojamo erdvinio proceso plotis. Prie tokių modelių priskiriamas *laipsninis modelis* $\gamma_0(t) = \theta_0 + \theta_1 t^\xi$, $0 < \xi < 2$:

- $\xi = 1$, tiesinis modelis;
- $\xi < 1$, išgaubtas modelis;
- $\xi > 1$, išlenktas modelis.

3. *Ribojami* modeliai, kai tam tikrame plote. Prie tam tikro slenksčio atstumo dispersija pasiekia slenkstį. Keli jų:

- tiesinis modelis su slenksčiu (1.3 pav.):

$$\gamma_0(t, \boldsymbol{\theta}) = \begin{cases} 0 & \text{jei } t = 0 \\ \theta_0 + \theta_1 t, & \text{jei } 0 < t < \theta_2, \\ \theta_0 + \theta_1 & \text{jei } t \geq \theta_2 \end{cases} \quad (1.25)$$

čia θ_1 ir θ_2 – teigiamos konstantos, o semivariograma artėja prie begalybės, kai $t \rightarrow \infty$;

- sferinis (1.4 pav.):

$$\gamma_0(t, \boldsymbol{\theta}) = \begin{cases} 0 & \text{jei } t = 0 \\ \theta_0 + \theta_1 \left[\frac{3}{2} \frac{t}{\theta_2} - \frac{1}{2} \left(\frac{t}{\theta_2} \right)^3 \right] & \text{jei } 0 < t \leq \theta_2. \\ \theta_0 + \theta_1 & \text{jei } t > \theta_2 \end{cases} \quad (1.26)$$

4. *Ribojami* modeliai, kai dispersija pasiekia slenkstį prie kelių efektyvių slenksčio atstumų (kai $\theta_0 + \theta_1 * 0,95$):

- eksponentinis (efektyvus slenksčio atstumas $3\theta_2$) (1.5 pav.):

$$\gamma_0(t, \boldsymbol{\theta}) = \begin{cases} 0 & \text{jei } t = 0 \\ \theta_0 + \theta_1 \left(1 - e^{-\frac{t}{\theta_2}} \right) & \text{jei } t > \theta_2. \end{cases} \quad (1.27)$$

- *Gauso* (efektyvus slenksčio atstumas yra $\sqrt{3\theta_2}$) (1.6 pav.):

$$\gamma_0(t, \boldsymbol{\theta}) = \begin{cases} 0 & \text{jei } t = 0 \\ \theta_0 + \theta_1 \left(1 - e^{-\frac{t^2}{\theta_2^2}} \right) & \text{jei } t > \theta_2. \end{cases} \quad (1.28)$$

5. *Mattern* klasė. Šią funkcijų klasę sukūrė Mattern (1960), tačiau kitos semivariogramų ir kovariacinės funkcijos naudojamos dažniau dėl jų išraiškų paprastumo. Visgi Handcock ir Stein (1993) bei Handsock ir Wallis (1994) demonstruoja *Mattern* funkcijų klasės lankstumą dirbant su įvairiais duomenų rinkiniais. Parametrinės (izotropinės) semivariogramos funkcijos, sudarytos *Mattern* klasės pagrindu, išraiška :

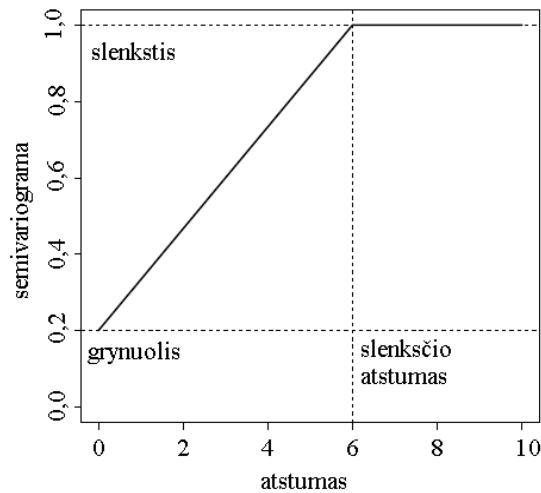
$$\gamma_0(t, \boldsymbol{\theta}) = \theta_0 + \theta_1 \left[1 - \frac{1}{2^{\theta_3-1} \Gamma(\theta_3)} \left(\frac{t}{\theta_2} \right)^{\theta_3} K_{\theta_3} \left(\frac{t}{\theta_2} \right) \right], \quad (1.29)$$

čia $\theta_3 > 0$ – formos parametras, kuris reguliuoja proceso diferencijuotumą, $\Gamma(\cdot)$ – įprasta *gamma* funkcija, o K_{θ_3} yra θ_3 trečios eilės modifikuota *Bessel*'io funkcija. Atskiru atveju, kai $\theta_2 = \frac{1}{2}$ semivariograma įgauna eksponentinę formą, o kai $\theta_2 \rightarrow \infty$ – *Gauso* semivariogramos formą (Smith 2001).

Kiekvienas parametrinis modelis kuriamas toks, kad kuo tiksliau atitiktų konkretaus tipo reiškinių. Tikslų taisyklių, kaip pasirinkti „geriausią“ semivariogramos modelį nėra. Visgi „geriausias“ parametrinis semivariogramos modelis gali būti pasirinktas taip:

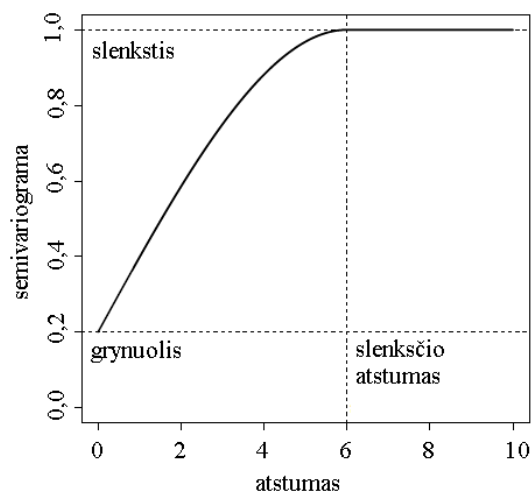
- semivariogramos modelis parenkamas priklausomai nuo analizuojamos problemos (tarkim Brauno judesiui tinkamiausias yra tiesinis semivariogramos modelis);
- palyginus kelis parametrinius semivariogramų modelius, svertiniu mažiausių kvadratų sumų metodu skaičiuojama paklaidų kvadratų suma (toliau PKS) (*angl. SSE – Sum Square of Error*) ir išrenkamas modelis, turintis minimaliausią reikšmę šios statistikos reikšmę;
- pritaikius kelis parametrinius semivariogramų modelius, galima pritaikyti kryžminės patikros metodą (2.5 skyrelis).

Gauso modelis be grynuolio efekto turi vienintelį sprendinį, kai kriginio tiesinių lygčių sistema yra išsprendžiama. Todėl, taikant *Gauso* semivariogramos modelį, grynuolio efekto reikšmė turi būti prilyginama nuliui. Erdvinio lauko, kuriam tinkamiausias *Gauso* modelis, elgsena panaši kaip ir eksponentinio modelio atveju, tik *Gauso* semivariogramos modelio atveju egzistuoja didelė priklausomybė nuo kaimyninių reikšmių.

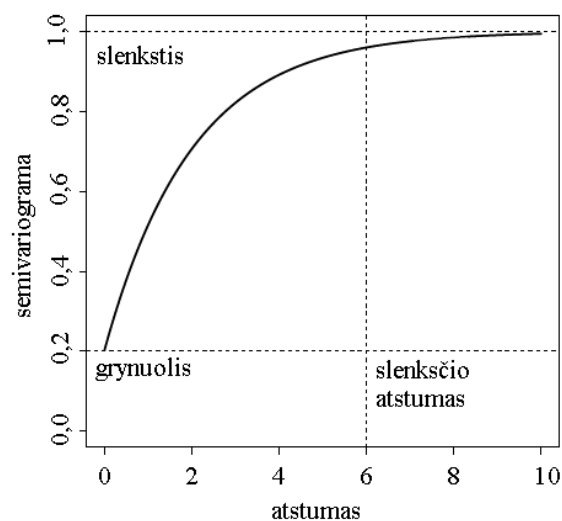


1.3 pav. Tiesinis su slenksčiu semivariogramos modelis (Rossiter 2004)

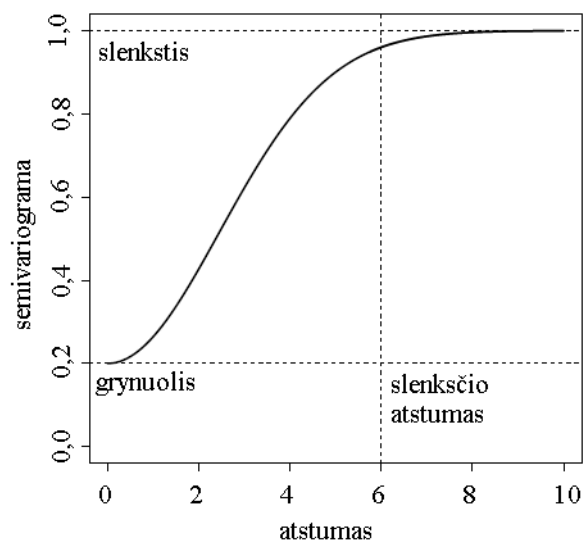
Fig. 1.3 Linear semivariogram model with sill (Rossiter 2004)



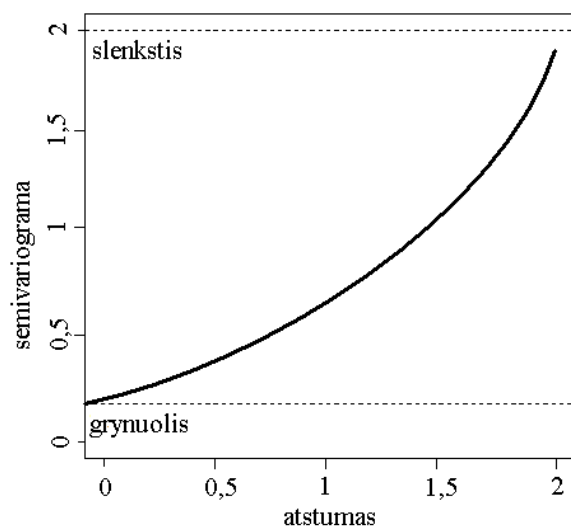
1.4 pav. Sferinis semivariogramos modelis (Rossiter 2004)
Fig. 1.4. Spherical semivariogram model (Rossiter 2004)



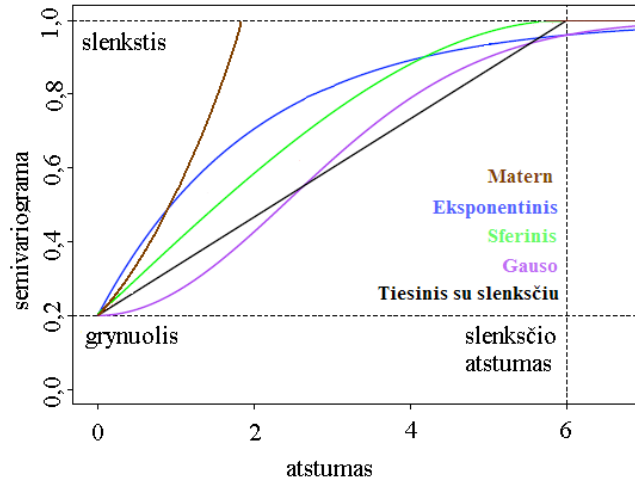
1.5 pav. Eksponentinis semivariogramos modelis (Rossiter 2004)
Fig. 1.5. Exponential semivariogram model (Rossiter 2004)



1.6 pav. *Gauso* semivariogramos modelis (Rossiter 2004)
Fig. 1.6. *Gaussian* semivariogram model (Rossiter 2004)



1.7 pav. *Matheron* semivariogramos modelis (Rossiter 2004)
Fig. 1.7. *Matheron* semivariogram model (Rossiter 2004)



1.8 pav. Semivariogramų modelių palyginimas (Rossiter 2004)

Fig. 1.8. The comparison of semivariogram models (Rossiter 2004)

Jei įvertinta semivariograma pradžioje yra tiesinė, turėtų būti taikomi eksponentinis arba sferinis semivariogramų modeliai. Sferinis modelis teisingas tik kai $d = 1, 2$ arba 3 , tačiau aukštesnėje dimensijoje šis modelis netenkina neigiamo apibrėžtumo sąlygos. Eksponentinis modelis yra paprastesnė sferinio modelio forma, tačiau neturi slenksčio. Tiesinio modelio su slenksčiu atveju didėjant atstumui proceso autokoreliacija mažėja link nulio.

Kartais siekiant parinkti geresnį modelį empirinės semivariogramos taškams, naudojama tiesinė kelių parametrinių semivariogramų modelių kombinacija:

$$\gamma(t, \boldsymbol{\theta}) = \sum_{i=1}^n |\lambda_i| \gamma_i(t, \boldsymbol{\theta}) \quad (1.30)$$

Pvz., semivariogramai, kuri nepasiekia pastovaus slenksčio, tačiau turi parabolinę elgseną jos pradžioje, galima apjungti *Gauso* ir tiesinį modelį.

Tiesinę kelių parametrinių semivariogramų modelių kombinaciją patartina naudoti, kai to reikalauja nagrinėjamo proceso struktūra. Modelių svoriai dažniausiai parenkami bandymų metodu, o bendra svorių suma turi būti lygi slenksčiui (Isaaks, Srivastava 1989).

Erdvinis pokytis tarp dviejų koreliuotų (priklausomų) erdvinį kintamųjų yra aprašomas kryžmine semivariograma. Kai vidinės hipotezės prielaidos teisingos, kryžminė semivariograma aprašoma formule:

$$\gamma_{kl}(\mathbf{h}) = \gamma_{lk}(\mathbf{h}) = \frac{1}{2} E[(Z_k(\mathbf{s} + \mathbf{h}) - Z_k(\mathbf{s}))(Z_l(\mathbf{s} + \mathbf{h}) - Z_l(\mathbf{s}))]. \quad (1.31)$$

Šiuo atveju k ir l žymi skirtingus kintamuosius. Kryžminė semivariograma yra simetrinė vektoriau $\mathbf{h}$ atžvilgiu, t. y. $\gamma_{ij}(\mathbf{h}) = \gamma_{ji}(-\mathbf{h})$. Kryžminės semivariogramos įvertinys skaičiuojamas taip:

$$\hat{\gamma}_{kl}(\mathbf{h}) = \frac{1}{2[N(\mathbf{h})]} \sum_{(\mathbf{s}_i, \mathbf{s}_j) \in N(\mathbf{h})} [(Z_k(\mathbf{s}_i) - Z_k(\mathbf{s}_j))(Z_l(\mathbf{s}_i) - Z_l(\mathbf{s}_j))]. \quad (1.32)$$

Įvertinys gali būti apskaičiuotas taip pat nauju kintamuoju $Z_k(\mathbf{s}) + Z_l(\mathbf{s})$. Tuomet semivariogramos $\hat{\gamma}_{k+l}(\mathbf{h})$ įvertinys aprašomas formule:

$$\hat{\gamma}_{k+l}(\mathbf{h}) = \frac{1}{2} \left[\hat{\gamma}_{kl}(\mathbf{h}) - \hat{\gamma}_k(\mathbf{h}) - \hat{\gamma}_l(\mathbf{h}) \right]. \quad (1.33)$$

1.5. Pirmojo skyriaus apibendrinimas ir disertacijos uždavinių formulavimas

- Erdvinių duomenų analizėje yra kelios esminės sąlygos, tokios kaip griežtas stacionarumas, silpnas stacionarumas, vidinis stacionarumas bei izotropiškumas. Šios sąlygos vienais ar kitais atvejais turi būti įgyvendintos prieš atliekant erdvinių duomenų modeliavimą ir prognozavimą.
- Erdvinių duomenų požymio $\mathbf{Z}$ modelį sudaro erdvinio proceso vidurkio ir klaidų dedamosios. Dažniausiai išskiriami pastovaus ir nepastovaus vidurkio modeliai.
- Duomenų variacijai aprašyti dažnai naudojamos semivariogramos, kurios, priklausomai nuo erdvinio proceso tipo, skirstomos į izotropines (anizotropines), visakryptes (kryptinės) bei kryžmines.
- Praktikoje dažnai taikomas klasikinis Matheron'o semivariogramos įvertinys MoM (1.15), kuris gali būti išreikštas per erdvinę dizaino (Laplaso) matricą (1.17). Ši semivariogramos įvertinio išraiška patogi optimalaus eksperimento (plano) sudaryme.
- Kadangi MoM įvertinys nėra robusiškasis išskirtims, literatūroje siūlomi kiti įvertiniai, neįtakojami išskirčių. Kaip pavyzdys darbe minima Cressie-Howkins medianinė semivariograma.
- Tam, kad būtų atsižvelgta į anizotropiją, tendą bei kitas erdvinis procesus visoje nagrinėjamoje teritorijoje apibūdinančias

charakteristikas, empirinės semivariogramos taškams priderinamas vienas parametrinių – sferinis, *Gauso*, eksponentinis ar kt. – semivariogramų modelių.

- Esminis disertacijos uždavinys – išnagrinėti erdvinių duomenų modeliavimo ir prognozavimo etapus bei pasiūlyti bendrinę erdvinių duomenų modeliavimo semivariogramomis ir tiesinio prognozavimo metodiką.

Erdvinių duomenų statistinio modeliavimo ir prognozavimo metodika

Vienas dažniausių erdvinių duomenų modeliavimo tikslų yra duomenų prognozė (interpoliavimas ar ekstrapoliavimas) erdvėje. Prognozės rezultatų efektyvumas priklauso nuo duomenų erdvinės sklaidos aprašymo tikslumo. Čia svarbu atsižvelgti į erdvinį trendą, anizotropiškumą ir atitinkamai išreikšti šias duomenų savybes semivariograma.

Šis skyrius skirtas erdvinių duomenų modeliavimo ir prognozavimo etapų aprašymui, siekiant pasiūlyti vieningą erdvinių duomenų modeliavimo ir prognozavimo metodiką. Skyrius padalintas į šešias dalis. Pirmoje dalyje pateikiama erdvinių duomenų prognozavimo metodų klasifikacija bei bendriniai geostatistinių duomenų modeliavimo ir prognozavimo etapai. Antroje dalyje nagrinėjamas pirmasis – statistinės erdvinių duomenų analizės – etapas. Trečioje dalyje detalai analizuojamas semivariogramų modeliavimas; duomenų trendo ir anizotropijos identifikavimas, galimi anizotropijos tipai, jų pašalinimo būdai. Taip pat atskirame poskyryje nurodomi semivariogramų modelių parametrų vertinimo metodai, detaliau nagrinėjami mažiausių kvadratų metodai. Ketvirtas skyrius skirtas erdvinių duomenų prognozavimo metodui – krigingui. Čia pateikiamos įvairių krigingo tipų klasifikavimo priežastys, detaliau nagrinėjami

ordinarus, (ordinarus) blokinis, paprastas, universalus, medianinio išlyginimo krigingai bei kokrigingas. Penktoje dalyje nagrinėjami geriausio semivariogramos modelio ar krigingo patikros metodai, tokie kaip kryžminės patikros metodas, Akaike kriterijus. Paskutinėje, šeštoje, skyriaus dalyje apžvelgiama parametrinių semivariogramų modelių bei jų parametrų įtaką prognozei, analizuojami krigingo privalumai ir trūkumai lyginant su kitais prognozės metodais.

Skyriaus tematika paskelbti septyni autorės straipsniai (Krūminienė (Borisenko), Dučinskas, Garška 2003; Krūminienė (Borisenko), Garška 2004; Krūminienė (Borisenko), Dučinskas 2004; Krūminienė (Borisenko), Dučinskas 2004; Krūminienė (Borisenko) 2004); Krūminienė (Borisenko) 2004); Krūminienė (Borisenko) 2006).

2.1. Prognozavimo metodų klasifikavimas. Bendriniai modeliavimo ir prognozavimo etapai

Paviršiaus atvaizdavimo nagrinėjamiems duomenims tikslumas priklauso nuo daugelio faktorių: pradinių duomenų kiekio; pakankamai gero nagrinėjamo proceso suvokimo; adekvačių matematinių metodų, kurie aprašo proceso skirstinio dėsningumus, pritaikymo; galimybės naudoti patogius instrumentus duomenų analizei, paviršių sudarymui ir gautų rezultatų patikrai.

Duomenų prognozavimas galimas keliais būdais:

- **Interpoliacija**, kai apskaičiuojamos funkcijos reikšmės tarpiniuose taškuose;
- **Ekstrapoliacija**, kai nustatoma funkcijos elgsena už matavimo duomenų ribų;
- **Parinka** (glodinimas, *angl. Fitting*) – nagrinėjamiems duomenims vartotojo nuožiūra parenkamas tinkamiausias kreivės pavidalas, o kreivės koeficientai paskaičiuojami taikant vieną ar kelis aproksimacijos gerumo kriterijus (pvz., mažiausių vidutinių kvadratinių nuokrypių metodas). Priklausomai nuo parinktos funkcijos ir apskaičiuojamų jos koeficientų skiriama tiesinė ir netiesinė parinka.

Prognozės procedūros skiriasi priklausomai nuo:

- naudojamo tinklo narvelių dydžio, orientacijos;
- naudojamo duomenų modelio (pvz., paviršiaus duomenų modelis (*angl. TIN – Triangulated Irregular Network*), rastras, vektorinis poligonas);
- imties duomenų pobūdžio.

Dažniausiai prognozavime naudojami keliuose taškuose išmatuoti duomenys, vadinami formaliais duomenimis (*angl. Hard Data*). Informacija apie fizinius procesus arba reiškinius (intuityvi informacija), kuri sudaro tam tikrą raštą, vadinama neformaliais duomenimis (*angl. Soft Data*).

Reiktų paminėti, kad dauguma geostatistikų tą patį procesą vadina vienu vardu: arba *prognoze*, arba *įvertinimu*. Tuo tarpu (Schabenberger, Pierce 2001) teigia, kad prognozė – tai atsitiktinio kintamojo reikšmės nustatymas, o įvertinimas – nežinomos konstantos reikšmės nustatymas. Jei nagrinėjamas dydis randasi erdvinio proceso s matavimuose, reikia prognozuoti $\mathbf{Z}(s)$ arba signalą $\mathbf{S}(s)$. Jei visų erdvinio proceso, kuris sudaro specifinę atsitiktinę funkciją, realizacijų vidurkis yra matavimo vietoje s , tuomet reikia įvertinti $E[\mathbf{Z}(s)]$.

Prognozavimo (įvertinimo) procedūros gali būti klasifikuojamos į:

- taško ir srities prognozavimo procedūras;
- vietines (lokalias) ir visuotinas (globalias);
- tikslas ir apytiksles;
- stochastines bei deterministines;
- pažingsnines ir staigias prognozavimo procedūras (Waters 1997).

Deterministiniai algoritmai (procedūros) taikant įvairias matematines funkcijas bando taškams su x , y ir z koordinatėmis priderinti tam tikrą paviršiaus tipą. Deterministinių prognozavimo algoritmų grupei priklauso trendo paviršiaus įvertinimas (pagrįstas globaliais arba lokaliais polinomis), interpoliacija splineais ir kt. Deterministiniams prognozavimo algoritmams būdingi šie trūkumai:

- nėra atsižvelgiama į duomenų erdvines koreliacijos ypatybes, kai tuo tarpu erdvinis duomenis charakterizuoja priklausomybė arba autokoreliacija erdvėje;
- deterministiniai metodai neleidžia įvertinti prognozavimo patikimumo.

Stochastiniuose algoritmuose keliamas klausimas, ar erdvinio duomenų rinkinys (tam tikra nagrinėjamo reiškinio abstrakcija) turi ar neturi atsitiktinę reikšmių sklaidą. Plačiau stochastiniai algoritmai nagrinėjami aptariant krigingo metodą (2.4 skyrelis). Kai kurios interpoliavimo algoritmų savybės pateiktos Blyth, Cake *et al.* (2008).

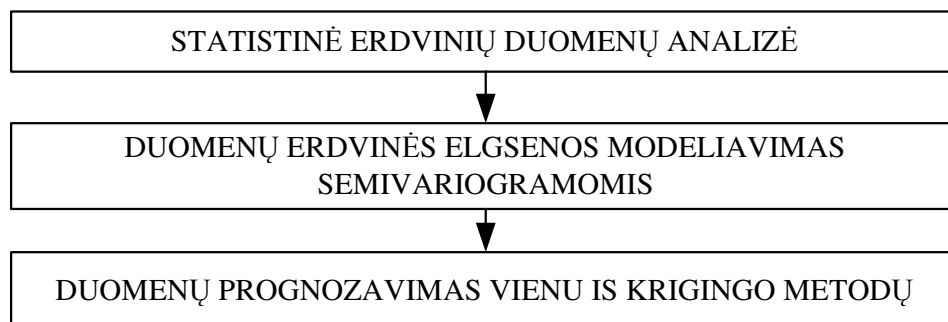
Daugelis aplinkos tyrimų, tokių kaip aplinkos užterštumo analizė ir pan., susiję su erdvinėmis struktūromis ar stebinių raštu. Tad, dažnai siekiant sumažinti tyrimo išlaidas, pravartu iš anksto numatyti stebinių išsidėstymo raštą, kuris leistų gauti tikslesnius tak tikro požymio prognozės rezultatus.

Optimalaus erdvinio dizaino (plano) sudarymas turi ilgą statistinę istoriją. Didžioji dalis darbų koncentruojasi ties nepriklausomų duomenų rinkinio sudarymu. Fedorov (1996) ir Muller (2001) nagrinėja atsitiktinių laukų sudarymą. Apskritai optimalūs planai skirstomi į tikimybę ar modeliu

grindžiamus planus. Tikimybe grindžiamo plano atveju dažniausiai atliekama paprasta atsitiktinė atranka be grąžinimo. Modeliu grindžiamo plano atveju taikomi arba regresinis, arba atsitiktinio proceso modeliai. Regresinio modelio su nepriklausomais duomenimis atveju optimalumas pasiekiamas efektyvių regresijos koeficientų įverčių sąskaita. Optimalumo kriterijus yra apibrėžta plano matricos funkcija ir „geriausias“ taškų rinkinys (planas) optimizuoja šį kriterijų per visas plano matricas (žr. Pukelsheim (1993) arba Muller (2001)). Aproximacija anizotropijos kontekste yra nagrinėjama Arbia and Lafratta (2002) darbe.

Nagrinėjant stacionarius Gauso procesus su regresine struktūra, yra svarbūs du klausimai, susiję su optimaliu planu: (1) Koks yra optimalus stebinių planas siekiant atlikti prognozę neišmatuotame srities taške? (2) Koks yra optimalus ėminių planas, siekiant įvertinti kovariacinės funkcijos parametrus? Kadangi, kaip jau buvo minėta darbe, prognozavimas yra vienas dažniausių tyrimo tikslų, pirmas klausimas literatūroje nagrinėjamas plačiau (žr. McBratney *et al.* 1981, Su ir Cambanis (1993), Ritter (1996), ir Zhu (2002)). Nagrinėjant antrą klausimą, pastovaus vidurkio atveju, klasikinė kovariacinės struktūros įvertinimo procedūra remiasi semivariogramomis (žr. Warrick ir Myers (1987), Bogaert ir Russo (1999), bei Muller ir Zimmerman (1999)). Šiuo atveju patogu taikyti semivariogramos įvertinį, išreikštą per erdvinio dizaino (Laplaso) matricą.

Literatūroje dažniausiai pateikiami tokie bendriniai geostatistinių duomenų, kuriems šiame darbe skiriamas išskirtinis dėmesys, etapai (2.1 pav.):



2.1 pav. Bendriniai geostatistinio modeliavimo ir prognozavimo etapai

Fig. 2.1. Generic stages of geostatistical modeling and prediction

2.2. Statistinė erdvinių duomenų analizė

Pradedant 1950-taisiais, kai kurie erdvinių statistinių duomenų analizės (ESDA) metodai, naudoti erdviniams raštams identifikuoti bei charakterizuoti,

buvo pakeisti ir papildyti naujais metodais. Jie priklauso kelioms tyrimų sritims (augalų ekologija, gyvūnų ekologija, geografija), naudojami skirtingiems tyrimo tikslams realizuoti (pradedant duomenų tiriamąja analize, baigiant rezultatų pateikimu), matematiniam taikymui (nuo variacijos-kovariacijos iki grafų metodų) ir tenkina tam tikras prielaidas (stacionarumo arba pseudo-stacionarumo). Tad metodų, taikomų erdviniams duomenims įvertinti, pasirinkimas priklauso nuo tyrimo tikslo, stebinių arba imties sudarymo tipo.

Bet kuriuo atveju prieš analizuojant geostatistinius duomenis, pirma tikrinama jų atributinė informacija. Matavimus, kurie rodo aiškias klaidas (pvz., neigiamos PH reikšmės) būtina pašalinti. Tuomet vykdoma duomenų topologinės informacijos loginė patikra, kuri gali būti atlikta *Geografinių Informacinių Sistemų (GIS)* priemonėmis. Jos tikslas nustatyti, ar matavimai atitinka logines nagrinėjamos srities ribas (pvz., dirvožemio matavimai negali būti ežere ir panašiai) (Fortin, Dale, ver Hoef 2001).

Taip pat erdvinės statistikos turi tenkinti stacionarumo ir izotropiškumo sąlygas, aprašytas 1.2 skyrelyje. Be to yra žinoma, kad nuo nagrinėjamos srities dydžio ir formos priklauso kaip tiksliai erdvinė statistika įvertina erdvinio rašto intensyvumą, plotį ir tipą. Čia dažnai susiduriama su krašto problema (Diggle 2003).

Bendru atveju erdvinėje statistinėje duomenų analizėje (ESDA) patikrinamas duomenų skirstinio atitikimas Gauso skirstiniui, išskirčių egzistavimas, globalus trendas, stacionarumas, autokoreliacija erdvėje ir t. t. Tam gali būti panaudoti ir pradinės duomenų analizės (EDA) metodai, tokie kaip histograma, penkiaskaitė suvestinė, Q-Q normališkumo diagrama, ANOVA ir kt.

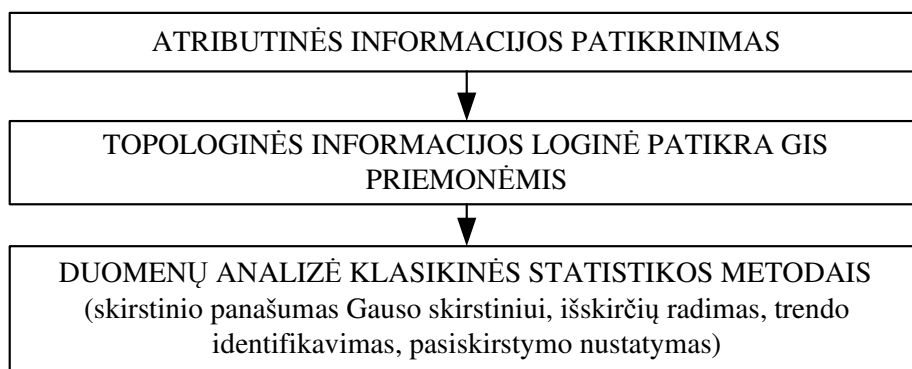
Erdvinės autokoreliacijos instrumentai tikrina, ar nagrinėjamas požymio reikšmės viename erdvės taške nėra priklausomos nuo aplinkinių kintamojo reikšmių. Teigiama erdvinė autokoreliacija rodo, kad netoli esančios reikšmės yra panašios, tuo tarpu kai neigiama autokoreliacija nurodo, kad artimoje asociacijoje esančios reikšmės yra nepanašios. Kai statistiškai reikšminga erdvinė autokoreliacija neegzistuoja, teigiama, kad erdvinių duomenų skirstinys yra atsitiktinis (Chou 1997).

Erdvinė autokoreliacija gali atsirasti dėl kelių priežasčių. Yra skiriamos netikra, interpoliuota, tikra ir iššaukta autokoreliacijos. Net kai stebiniai yra statistiškai nepriklausomi, netikra autokoreliacija gali būti stebima dėl erdvinių duomenų sutapimų. Interpoliuota autokoreliacija atsiranda, kai erdvinis paviršius priglodinamas, interpoliuojamas arba ekstrapoliuojamas. Tikroji erdvinė autokoreliacija yra priežastinės sąveikos tarp arti esančių stebinių rezultatas. Pagaliau erdvinė autokoreliacija gali būti iššaukta priklausomo kintamojo sąryšio su kitais erdviškai autokoreliuotais kintamaisiais. Tikroji ir iššauktoji autokoreliacija yra erdvės – laiko procesų rezultatas (pvz., migracija). Yra daug

erdvinės autokoreliacijos indikatorių, kurie detaliau nagrinėjami Cliff, Ord (1973, 1981); Goodchild (1987); Haining (1990); Chou (1997).

Taip pat prieš vykdant semivariogramų modeliavimą parenkamas duomenų skirstinys arba atliekama parametrizacija (Deutsch, Journel 1998). Kai kuriuose algoritmuose (pvz., *p-laukų*) reikalaujama empirinę semivariogramą modeliuoti duomenis transformavus į tolydųjį skirstinį. Tačiau tolydaus ir Gauso skirstinių semivariogramos labai panašios, tad dažniausiai naudojamas Gauso skirstinys. Dauguma ekologinių duomenų yra dichotominiai arba teigiami su dešinės pusės asimetrija (tokie kaip vandens užterštumas). Indikatorinis metodas reikalauja nagrinėjamą kintamąjį perkoduoti į binarinį. Kai duomenys yra iškreipti arba turi išskirtis, paskaičiuotos semivariogramos reikšmės būna chaotiškos. Tuomet siūloma skaičiuoti robustines semivariogramas (1.4 skyrelis).

Kai duomenų skirstinyje stebima dešinės pusės asimetrija dažnai taikomas lognormališkumas, bet gali būti taikomos ir kitos normališkumo transformacijos (Ecker, Gelfand 1999). Duomenis normalizavus panaikinami skirtumai tarp ekstremalių reikšmių, o teorinio slenksčio reikšmė tampa artima 1.



2.1 pav. Statistinės erdvinį duomenų analizės etapai
Fig. 2.2. The stages of statistical analysis of spatial data

Visgi erdvinį duomenų transformavimas turi tam tikrus trūkumus. Paprastai kiekvienas masyvo komponentas transformuojamas atskirai ir bendras duomenų masyvo skirstinio priartėjimas prie Gauso skirstiniui sunkiai pasiekiamas. Be to transformuotus duomenis sunku interpretuoti, ypač jei atskiriems duomenų masyvo kintamiesiems naudotos skirtingos transformacijos. Taip pat, net jei prognozuotų duomenų atgalinė transformacija galima, tam tikrais atvejais baigtinis rezultatas būna paslinktas (Miller 1984, Cressie 1993). Dėl išvardintų trūkumų, kaip alternatyva, duomenų rinkinio atvaizdavimui naudojami parametriniai daugialypiai skirstiniai.

Tad bendru atveju statistinė erdvinė duomenų analizė schematiškai gali būti atvaizduota schema, pateikta 2.1 paveiksle.

2.3. Semivariogramų modeliavimas

Semivariogramų modeliavimas atliekamas keliais etapais:

1. Apskaičiuojama ir grafiškai atvaizduojama empirinė semivariograma. Paskaičiuotos empirinės semivariogramos elgsena gali atskleisti nagrinėjamo proceso ypatumus: vaizduoti chaotiškumą, cikliškumą, anizotropiškumą ar trendą. Dauguma semivariogramų, sudaromų gamtinių procesų pagrindu, vaizduoja kelių minėtų požymių kombinaciją.

Kai nagrinėjami procesai kinta chaotiškai, yra netiesiniai bei neturi jokios erdvinės autokoreliacijos struktūros, empirinės semivariogramos taškai išsidėsto chaotiškai. Tokios elgsenos semivariogramos vadinamos grynuolio efektu (*angl. Nugget Effect*). Rudos pramonėje grynuolio efektu vadinami atsitiktiniai pokyčiai, kurie susidaro, jei gręžinyje randamas aukso grynuolis.

Geologiniai reiškiniai dažnai po tam tikro laiko kartojasi, sudarydami pakartotinus arba cikliškus pokyčius fizikinėse nagrinėjamos proceso savybėse. Tokiu atveju empirinės semivariogramos taškai vaizduoja ciklišką (sinusoidinį) autokoreliacijos kitimą. Tokia semivariogramos elgsena vadinama plyšio efektu (*angl. Hole Effect*). Plyšio efektas gali susidaryti ir dėl duomenų trūkumo. Pvz., rudos pramonėje semivariograma vaizduoja geologinį cikliškumą, jei gręžinyje atsirado plyšys. Plyšio efektas gali būti semivariogramos modeliuose su arba be slenksčio.

2. Nustatoma, ar duomenyse egzistuoja trendas. Nubrėžiama horizontali linija, vaizduojanti teorinį slenkstį. Jei duomenys buvo normalizuoti, eksperimentinės (pastovios) dispersijos reikšmė lygi 1. Diskrečių binarinių kintamųjų dispersija lygi $p(1-p)$, čia p – nagrinėjamos kategorijos proporcija. Reiktų pastebėti, kad dauguma programinių priemonių, skirtų geostatistinių duomenų analizei, neturi instrumentų, leidžiančių atvaizduoti teorinį slenkstį.

Jei empirinės semivariogramos taškai ženkliai kyla virš teorinio slenksčio, tai tikėtina, kad duomenyse egzistuoja trendas. Yra trys šios problemos sprendimo būdai:

- a) priderinti tam tikro laipsnio trendo paviršių ir tolesnius skaičiavimus atlikti su trendo liekanomis;
- b) bandyti rasti kryptį, kurioje trendas nėra akivaizdus ir naudoti semivariogramą, sudarytą šia kryptimi;
- c) ignoruoti problemą ir naudoti tiesinę arba laipsninę semivariogramą (Blyth, Cake *et al.* 2008).

Tam tikro laipsnio trendo paviršius nustatomas mažiausiu kvadratų (MK) metodu. T. y. trendo paviršius aprašomas matavimų geografinių koordinatų

tiesine funkcija, kurios kvadratinis nuokrypis nuo trendo yra minimizuojamas. Tam, kad nustatyti, kurio laipsnio trendo polinomas geriausiai aprašo nagrinėjamą paviršių, gali būti sudarytas ANOVA modelis bei patikrinta hipotezė apie trendo paviršiaus priglodinimo reikšmingumą.

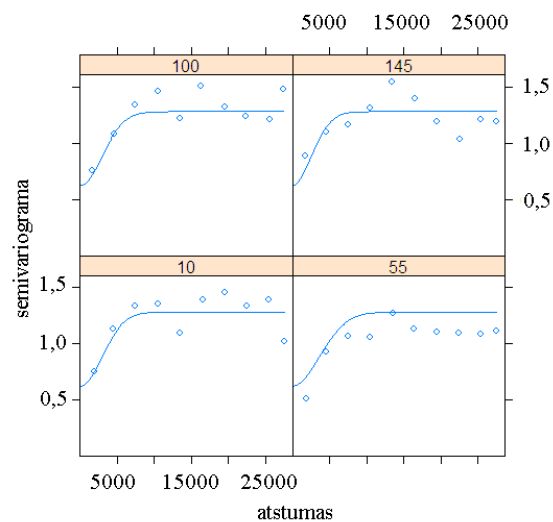
Įvertinant trendo tipą ir pašalinant determinuotą trendo dalį susiduriama su tam tikromis problemomis. Dėl svaraus (žytaus) trendo, kintamasis yra nepastovus, t. y. negalima tikėtis, kad modelio vidurkis bus nepriklausomas nuo vietovės koordinatų. Trendo pašalinimas sudaro paslinktumą, tačiau tas paslinktumas yra mažiau kenksmingas nei klaidos, susidarančios, jei trendas yra nepašalinamas. Trendą pašalinus semivariogramų analizė ir visi tolesni įvertinimai atliekami liekanų pagrindu. Trendas pridedamas prie įvertintų reikšmių tyrimo pabaigoje.

Tiriant trendo paviršių svarbu prisiminti, kad:

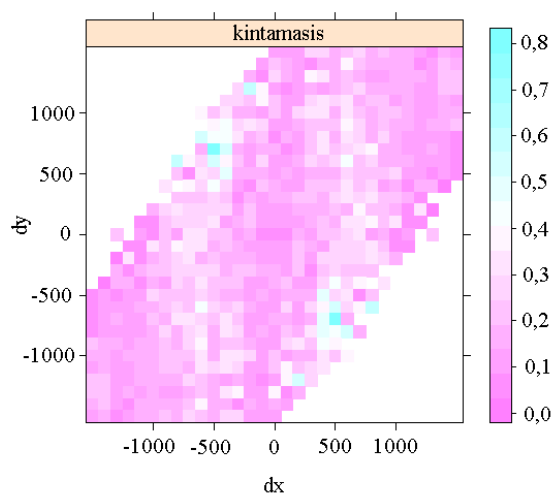
- reikalingas adekvatus duomenų kiekis ($n > Q+1$, čia n – duomenų kiekis, o Q – trendo modelio koeficientų kiekis);
- duomenys neturi grupuotis (ypač kai n yra mažas);
- reikia būti atsargiems atliekant ekstrapoliaciją už nagrinėjamos srities ribų;
- atsargiai naudoti aukštesnio laipsnio polinomines funkcijas.

Anizotropijai nustatyti Kaluzny, Vega, Cardosa ir Shelly (1996) siūlo kryptines semivariogramas (2.2 pav.); Isaaks ir Srivastava (1989) siūlo rožės diagramą (*angl. Rose Diagram*) ir empirinės semivariogramos (kovariacijos) paviršiaus kontūro grafiką R^2 erdvėje (kitai dar vadinamas semivariogramos (kovariacijos) žemėlapiu (*angl. Semivariogram Map*) (2.3 pav.)), kur kiekviena paviršiaus tinklo celė vaizduoja semivariogramos (kovariacijos) reikšmę duotu vėlinimo vektoriumi ir kryptimi. Visgi dažniausiai praktikuojamos visakryptės semivariogramos.

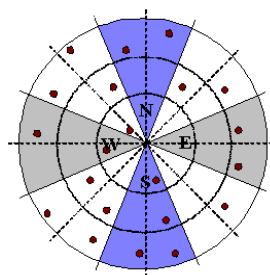
Variogramos, kurios kuriamos neatsižvelgiant į stebinių porų kryptį vadinamos *visakryptėmis* semivariogramomis. *Kryptinės* semivariogramos skaičiuojamos naudojant tam tikra kryptimi orientuotas stebinių poras. Dažniausiai kryptinių semivariogramų skaičiavimui pasirenkamos 4 pagrindinės kryptys (2.4 pav.): šiaurės – pietų, vakarų – rytų, šiaurės vakarų – pietryčių, šiaurės rytų – pietvakarių kryptys (0° , 45° , 90° ir 135°) kiekvienam atstumo intervalui su $22,5^\circ$ atskyrimo kampu (2.5 pav.). Tačiau šių kryptų pasirinkimas nėra privalomas. Galima sudaryti ir daugiau kryptinių semivariogramų, parenkant mažesnę vėlinimo vektorių (1.4 skyrelis). Visgi vėlinimo žingsnis $T(\mathbf{h}_k)$ turi būti pakankamai didelis, kad jis apimtų pakankamą stebinių kiekį, ir pakankamai mažas, kad atskleistų duomenų anizotropiškumą.



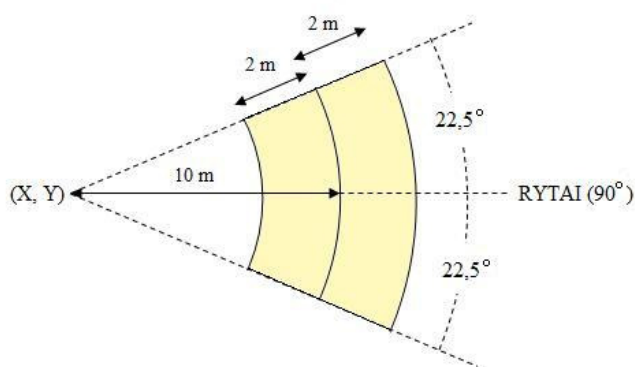
2.2 pav. Kryptinės semivariogramos, sudarytos *R* sistemos priemonėmis
Fig. 2.3. Directional semivariograms created by *R* system



2.3 pav. Semivariogramos žemėlapis
Fig. 2.4. The semivariogram map

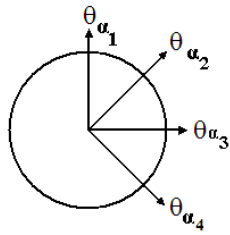


2.4 pav. Pagrindinės kryptinių semivariogramų kryptys ir toleracijos kampai
Fig. 2.5. Main tolerance angles and directions of directional semivariograms

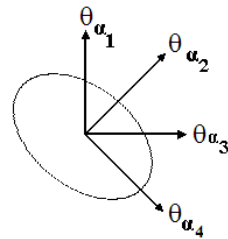


2.5 pav. Tolerancijos sritis, parenkant stebinius, semivariogramai sudaryti
Fig. 2.6. The tolerance area of choice of observations
 for creation of semivariogram

2.7 paveiksle parodytas pavyzdys, kai sudaromos kryptinės semivariogramos 4 pagrindinėmis kryptimis (0° , 45° , 90° ir 135°), ir visų kryptinių semivariogramų slenksčio atstumas yra vienodas. Tokiu atveju semivariogramos žemėlapis vaizduoja apskritimą (izotropinis atvejis, 2.6 pav). Tiksliau taškų, dalyvaujančių kringo prognozėje, paieškos sritis (semivariogramos žemėlapis) sudaro apskritimą. Anizotropijos atveju paieškos sritis yra elipsės formos (2.7 pav). Detaliau ši situacija ir jos sprendimo būdai nagrinėjami 2.3.1 skyrelyje.



2.6 pav. Izotropinis atvejis
Fig. 2.7. The isotropic case



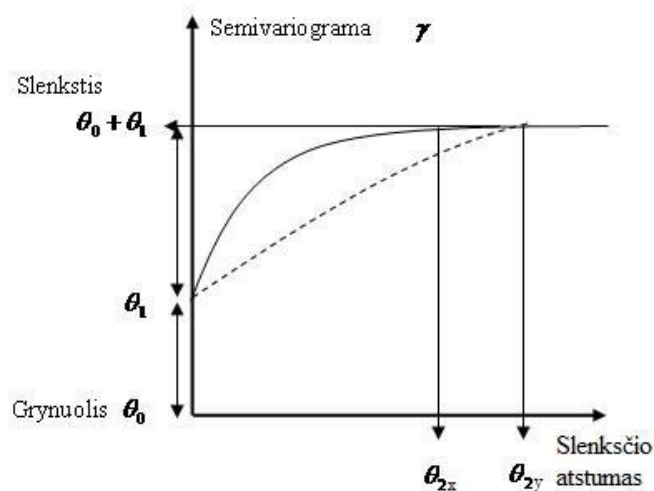
2.7 pav. Geometrinės anizotropijos atvejis
Fig. 2.8. The case of geometric anisotropy

3. Empirinės semivariogramos taškams, atsižvelgus į tendą ir eliminavus anizotropiją (jei šios duomenų savybės egzistavo), semivariogramai galima parinkti vieną iš parametrinių semivariogramos modelių (sferinį, eksponentinį, *Gauso* ir t. t.). Dažnai semivariogramoms pritaikomi keli parametriniai modeliai ir skaičiuojama skirtumo tarp empirinių semivariogramos taškų bei parinkto parametrinio modelio kreivės paklaidų kvadratų sumą (PKS).

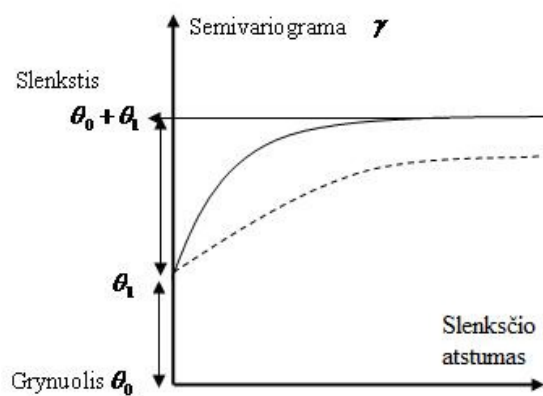
Praktikoje dažnai nutolstama nuo griežtos semivariogramų interpretavimo ir modeliavimo metodikos dėl duomenų trūkumo ir neadekvačių programinių priemonių, kurios nesuteikia galimybių interpretuoti tendą bei anizotropiją (Gringarten, Deutsch 2001).

2.3.1. Geometrinės anizotropijos pašalinimas

Prieš atliekant duomenų prognozę, kryptinės semivariogramos turi būti transformuojamos taip, kad įgytų identišką reikšmę bet kuria kryptimi ir bet kuriuo vėlinimo vektoriumi. Daugelis autorių (Cressie 1993, Deutsch 1998, Goovaerts 1997, Journel, Huijbregts 1978, Wackernagel 2003) išskiria du anizotropijos tipus: geometrinę ir zoninę. Teigiama, kad duomenims būdinga geometrinė anizotropija, jei esant vienodam kryptinių semivariogramų slenksčiui, jų slenksčių atstumai nėra lygūs (2.8 pav.). Zoninės anizotropijos atveju, skiriasi kryptinių semivariogramų ir slenksčio atstumas, ir slenkstis (2.9 pav.).



2.8 pav. Kryptinės semivariogramos, vaizduojančios geometrinę anizotropiją
Fig. 2.9. Directional semivariograms with geometric anisotropy



2.9 pav. Kryptinės semivariogramos, vaizduojančios zoninę anizotropiją
Fig. 2.10. Directional semivariograms, with zonal anisotropy

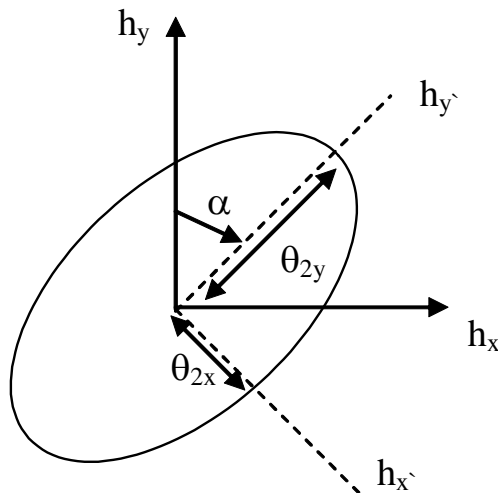
Tarkim sudarius kryptines semivariogramas buvo pastebėtas geometrinis anizotropiškumas. Tuomet būtina identifikuoti semivariogramą su didžiausiu semivariogramos slenksčiu bei mažiausiu semivariogramos slenksčiu.

Semivariogramos, kurios slenkstis maksimaliausias, kryptis retai sutampa su viena iš įprastų (pagrindinių) koordinačių ašių (tokiais atvejais pasukti ašis nereikėtų).

Jeigu semivariogramos su maksimaliu slenkščio atstumu ir semivariogramos su minimaliu slenkščio atstumu ašys ortogonalios ir šios semivariogramos yra to paties tipo, ašis galima pasukti α kampu (afininė transformacija²).

Praktikoje retai pasitaiko atvejų, kai semivariogramos, kurios slenkstis yra maksimaliausias, kryptis sutaptų su pagrindine koordinačių ašių kryptimi.

Tad tarkim, kad anizotropijos ašys randasi kryptimi (x', y') , kuri skiriasi nuo pagrindinės koordinačių sistemos krypties (x, y) tam tikru laipsniu α (2.10 pav.).



2.10 pav. Originalios koordinačių sistemos (x, y) transformacija į koordinačių sistemą (x', y')

Fig. 2.11. The transformations of orthogonal coordination system (x, y) to the coordination system (x', y')

Tokiu atveju būtina atlikti tiesinę koordinačių transformaciją ir apskaičiuoti naujas koordinates h'_x ir h'_y taip, kad:

$$\gamma(\mathbf{h}) = \gamma(h_x, h_y) = \gamma(\mathbf{h}') . \quad (2.1)$$

² Afininis koordinačių transformavimas – tai toks transformavimas, kai mastelis abscisų ir ordinatų ašių kryptimis keičiamas nevienodai, tokiu būdu keičiant srities formą (Balevičius G., Kriaučiūnaitė – Neklejonovienė V. 2008. Geodeziniai matavimai. Kaunas, Ardiva.)

čia $\gamma(\mathbf{h}') = \gamma\sqrt{(\mathbf{h}'_x)^2 + (\mathbf{h}'_y)^2}$ – semivariogramos modelis, atsižvelgiantis į geometrinę anizotropiškumą, o

$$\mathbf{h}'_x = a_{11}\mathbf{h}_x + a_{12}\mathbf{h}_y$$

$$\mathbf{h}'_y = a_{21}\mathbf{h}_x + a_{22}\mathbf{h}_y$$

ir a_{11} , a_{12} , a_{21} bei a_{22} yra svorių koeficientai.

Taigi vektorius iš standartinės koordinačių sistemos gali būti pertvarkytas į naują koordinačių sistemą transformuojant $\mathbf{R}$ matricą:

$$\mathbf{h}' = \begin{bmatrix} \mathbf{h}'_x \\ \mathbf{h}'_y \end{bmatrix} = \begin{bmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{bmatrix} \begin{bmatrix} \mathbf{h}_x \\ \mathbf{h}_y \end{bmatrix} = \mathbf{R}\mathbf{h}. \quad (2.2)$$

Jei semivariogramų su minimaliu slenksčio atstumu θ_{2x} ir maksimaliu slenksčio atstumu θ_{2y} kryptis neortogonalios, tuomet transformuojamas atstumas tarp $\mathbf{s}_i$ ir $\mathbf{s}_j$.

Atstumo transformacija $\mathbf{h} \rightarrow \mathbf{h}'$ gali būti atvaizduota matricos pavidalu:

$$\mathbf{h}' = \mathbf{T}\mathbf{R}\mathbf{h} = \mathbf{A}\mathbf{h}, \quad (2.3)$$

čia

$$\mathbf{T} = \begin{pmatrix} \frac{1}{\theta_{2x}} & 0 \\ 0 & \frac{1}{\theta_{2y}} \end{pmatrix}, \quad (2.4)$$

o matrica $\mathbf{A}$:

$$\mathbf{A} = \begin{pmatrix} \frac{1}{\theta_{2x}} & 0 \\ 0 & \frac{1}{\theta_{2y}} \end{pmatrix} \begin{bmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{bmatrix}, \quad (2.5)$$

čia θ_{2x} žymi minimalų vienos iš kryptinių semivariogramų slenksčio atstumą, o θ_{2y} – žymi maksimalų vienos iš kryptinių semivariogramų slenksčio atstumą.

$b = \frac{\theta_{2x}}{\theta_{2y}}$ vadinamas anizotropijos santykiu, kuris gali įgyti reikšmes intervale $[0, 1]$.

Taigi pirmoji transformacijos matrica **T** atlieka ištempimą arba sutraukimą, o antroji matrica **R** atlieka ašių pasukimą. Tad matrica **A** apibrėžia transformaciją iš pradinės erdvės (iš apskritimo su spinduliu a) (2.6 pav.) į anizotropinę erdvę (į elipsę, t. y. į formą, kuri gaunama tiesiškai transformavus apskritimą) (Goovaerts 1997, Hohn 1999).

Tad pagrindiniai geometrinės anizotropijos šalinimo etapai:

- nustatomos kryptys, kuriomis vienos iš kryptinių semivariogramų slenksčio atstumas yra maksimaliausias ir vienos iš kryptinių semivariogramų slenksčio atstumas – minimaliausias. Dažnai semivariogramos su minimaliausiu slenksčio atstumu kryptis parenkama statmenai semivariogramos su didžiausiu slenksčio atstumu kryptčiai;
- pasukamos duomenų ašys taip, kad atitiktų anizotropijos kryptį;
- kryptinės semivariogramos pertvarkomos į visakryptę semivariogramą su standartizuotu pločiu lygiu 1;
- dauginant **T** ir **R** matricas sudaroma vienetinė matrica **A**.

Transformuotai koordinačių sistemai gali būti pritaikytas tam tikras parametrinis semivariogramos modelis, o semivariogramą galima perrašyti tokiu pavidalu (Lophaven 2001):

$$\gamma(\mathbf{h}) = \gamma_0([\mathbf{A}\mathbf{h}]). \quad (2.6)$$

Geometrinės anizotropijos atveju:

- EkspONENTINIS semivariogramos modelis:

$$\gamma(\mathbf{h}) = \begin{cases} 0 & \text{jei } \mathbf{h} = 0 \\ \theta_0 + \theta_1 \left(1 - e^{-\sqrt{\mathbf{h}^T \mathbf{A}^T \mathbf{A} \mathbf{h}}}\right) & \text{jei } \mathbf{h} > \theta_2 \end{cases}.$$

- GAUSO semivariogramos modelis:

$$\gamma(\mathbf{h}) = \begin{cases} 0 & \text{jei } \mathbf{h} = 0 \\ \theta_0 + \theta_1 \left(1 - e^{-\mathbf{h}^T \mathbf{A}^T \mathbf{A} \mathbf{h}}\right) & \text{jei } \mathbf{h} > \theta_2 \end{cases}.$$

- SFERINIS semivariogramos modelis:

$$\gamma(\mathbf{h}) = \begin{cases} 0 & \text{jei } \mathbf{h} = 0 \\ \theta_0 + \theta_1 \left[\frac{3}{2} \frac{\sqrt{\mathbf{h}^T \mathbf{A}^T \mathbf{A} \mathbf{h}}}{\theta_2} - \frac{1}{2} \left(\frac{\sqrt{\mathbf{h}^T \mathbf{A}^T \mathbf{A} \mathbf{h}}}{\theta_2} \right)^3 \right] & \text{jei } 0 < \mathbf{h} \leq \theta_2 \\ \theta_0 + \theta_1 & \text{jei } \mathbf{h} > \theta_2 \end{cases}.$$

Geriausio semivariogramos modelio parinkimui anizotropijos atveju gali būti taikomi svertinis mažiausių kvadratų (*angl. WLS – Weighted Least Squares*), maksimalaus tikėtinumo ir mažiausio tikėtinumo liekanoms metodai.

Jei kryptinės semivariogramos yra tiesinė kelių semivariogramos modelių kombinacija, joje naudojami kryptinių semivariogramų parametriniai modeliai turi būti to paties tipo, t. y. visi sferiniai, visi eksponentiniai ir panašiai. Taip pat visos kryptinės semivariogramos turi turėti tą patį slenkstį ir svorį (λ_i).

Jei kryptinių semivariogramų slenkstis skiriasi, afininė transformacija nesuvienodins semivariogramų. Duomenų dispersija įvairiose nagrinėjamos srities zonose yra skirtinga, tad šis anizotropijos tipas pavadintas zonine anizotropija. Tokiu atveju sudedami du afininės transformacijos anizotropijos modeliai, kiekvienam nurodant labai aukštą anizotropijos santykį (taip, kad priešingos krypties dispersija yra beveik lygi nuliui). Plačiau apie zoninės anizotropijos šalinimo metodus rašoma Goovaerts (1997, pls. 93).

2.3.2. Semivariogramų parametrų vertinimo metodai

Semivariogramos modeliavimo procese svarbu parinkti parametrinį semivariogramos modelį, geriausiai aprašantį nagrinėjamos srities duomenų struktūros ypatumus. Vizualiai nustatytos parametrinio semivariogramos modelio parametrų reikšmės naudojamos parametrų optimizavimo procese.

Vertinant semivariogramų modelių parametrus gali būti taikomi mažiausių kvadratų (MK) arba maksimalaus tikėtinumo (MT) (*angl. ML – Maximum Likelihood*,) metodai. Pirmieji semivariogramų parametrų vertinimo procedūras taikant mažiausių kvadratų metodus aprašė David (1977) bei Journel, Huijbregts (1978). Darbus šia tema tesė Cressie (1985), Zhang *et al.* (1995), Genton (1997), Barry *et al.* (1997). Visi aprašytieji įvertiniai būdavo gaunami minimizuojant „atstumą“ tarp neparametrinių semivariogramų įvertinių ir parametrinių semivariogramų modelių. Šios procedūros įtrauktos į *R* sistemos paketus ir taikomos disertacijos praktinėje dalyje.

Yra trys dažnai taikomi netiesiniai mažiausių kvadratų metodai: SMK, PMKM ir AMKM. Gauso skirstinio populiacijų atveju taikomi maksimalaus tikėtinumo (MT) bei maksimalaus tikėtinumo liekanoms (MTL) (*angl. REML – Restricted Error Maximum Likelihood*) metodai, kuriuos taikant nėra būtina empirinė semivariograma. Taip pat pastaruoju metu vertinant erdvinių populiacijų parametrus vis plačiau taikomi regresinės analizės metodai. Tačiau jie pastarojoje disertacijoje nenagrinėjami.

Tarkime, kad pasirinkome K skirtingų vėlinimo vektorių $\mathbf{h}_1, \dots, \mathbf{h}_K$, priklausančių sričiai D ir apibrėžkime vektorius:

$$\hat{\gamma} = \left(\hat{\gamma}(\mathbf{h}_1), \dots, \hat{\gamma}(\mathbf{h}_K) \right)^T \text{ ir} \\ \gamma(\boldsymbol{\theta}) = (\gamma(\mathbf{h}_1, \boldsymbol{\theta}), \dots, \gamma(\mathbf{h}_K, \boldsymbol{\theta}))^T,$$

čia $\hat{\gamma}(\mathbf{h}_j)$ yra įvertinta semivariogramos reikšmė vėlinimo vektoriaus $\mathbf{h}_j$ klasėje, $\gamma(\mathbf{h}_j, \boldsymbol{\theta})$ – parametrinis semivariogramos modelis, kai $\mathbf{h} = \mathbf{h}_j$.

Vienoje iš paprasto mažiausių kvadratų metodo (*angl. OLS – Ordinary Least Squares*) (toliau PMKM) – būtinų sąlygų reikalaujama, kad regresinio modelio liekanos būtų nekoreliuotos, nes esant jų koreliacijai gali būti netiksliai prognozuoti regresijos parametrai. Šiuo metodu $\boldsymbol{\theta}$ parenkami taip, kad minimizuoti:

$$\left\{ \hat{\gamma} - \gamma(\boldsymbol{\theta}) \right\}^T \left\{ \hat{\gamma} - \gamma(\boldsymbol{\theta}) \right\}. \quad (2.7)$$

Apibendrinto mažiausių kvadratų metodo (toliau AMKM) (*angl. GLS – General Least Squares*) – atveju – $\boldsymbol{\theta}$ parenkami taip, kad minimizuoti:

$$\left\{ \hat{\gamma} - \gamma(\boldsymbol{\theta}) \right\}^T V(\boldsymbol{\theta})^{-1} \left\{ \hat{\gamma} - \gamma(\boldsymbol{\theta}) \right\}, \quad (2.8)$$

čia $V(\boldsymbol{\theta})^{-1}$ žymi $\hat{\gamma}$ kovariacinę matricą, kurios parametrai $\boldsymbol{\theta}$ yra nežinomi, tad $V(\boldsymbol{\theta})^{-1}$ elementus dažnai sudėtinga paskaičiuoti.

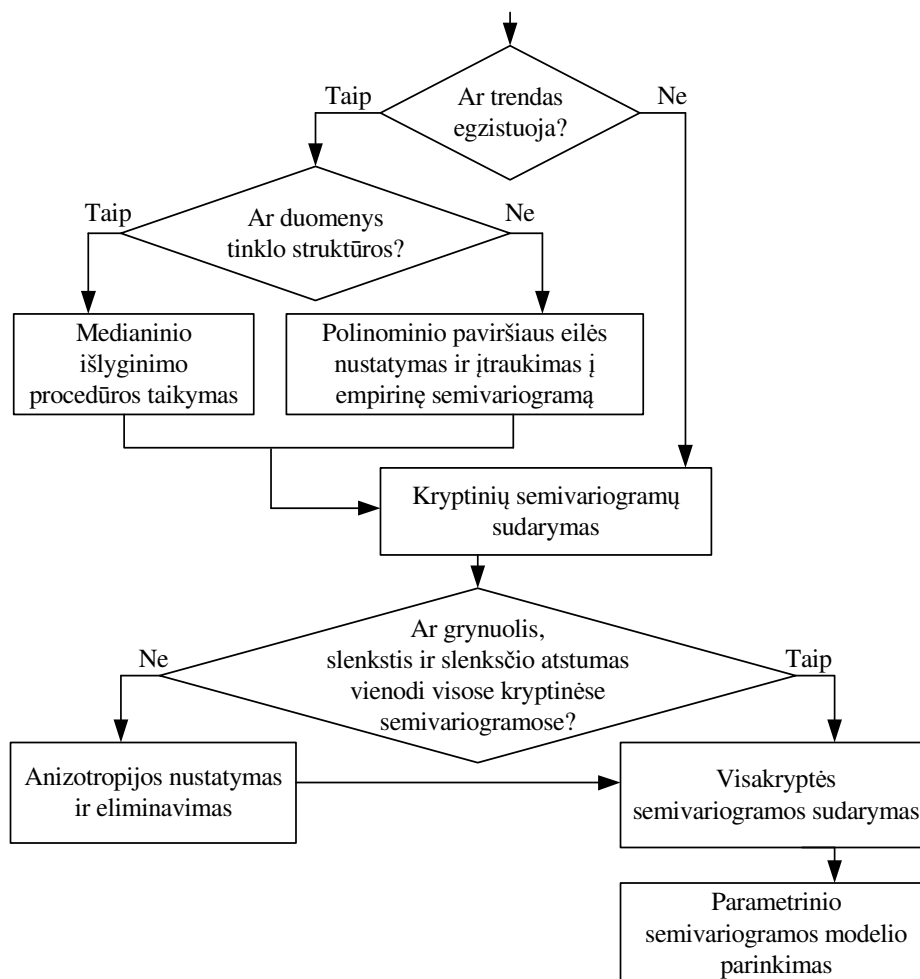
Aukščiau minėtų problemų išvengiama taikant svertinį mažiausių kvadratų metodą (toliau SMKM) (*angl. WLS – Weighted Least Squares*), kur $\boldsymbol{\theta}$ parenkami taip, kad minimizuoti:

$$\left\{ \hat{\gamma} - \gamma(\boldsymbol{\theta}) \right\}^T W(\boldsymbol{\theta})^{-1} \left\{ \hat{\gamma} - \gamma(\boldsymbol{\theta}) \right\}. \quad (2.9)$$

Cressie (1985) pasiūlė kitą – svertinį mažiausių kvadratų minimizavimą. Pagal pasiūlytą metodą, jei $\hat{\gamma}(\mathbf{h}_j)$ įvertinta baigtiniame $\{\mathbf{h}_j\}$ rinkinyje, $\boldsymbol{\theta}$ parenkami taip, kad minimizuoti

$$\sum_{j=1}^K N(\mathbf{h}_j) \left\{ \frac{\hat{\gamma}(\mathbf{h}_j)}{\gamma(\mathbf{h}_j; \boldsymbol{\theta})} - 1 \right\}^2. \quad (2.10)$$

Šis kriterijus suteikia didesnę svorį tų grupių semivariogramos reikšmėms, kuriose yra didesnis stebinių porų skaičius. Tačiau MK atveju įvertiniai yra koreliuoti ir turi skirtingas dispersijas.



2.11 pav. Semivariogramų modeliavimo etapai
Fig. 2.12. The stages of semivariogram modelling

SMKM metodas tinka tik semivariogramoms, bet netinka kovariacijoms. Tuo tarpu AMKM metodas gali būti taikomas abiem atvejais.

MT arba MTLM metodai (Warner *et al.* 1987, Lark 2000; Lark, Cullis 2004) taikomi tuomet, kai yra žinomi stebinių skirstiniai (dažniausiai Gauso atveju). Reiktų pastebėti, kad analogiški metodai taikomi vertinant kryžminės semivariogramos parametrus.

Maksimalaus tikėtinumo metodas (Kitanidis 1983, Mardia ir Marshall 1984) yra sudėtingesnis nei Cressie mažiausių kvadratų minimizavimo metodas (2.10), kadangi n duomenų kiekio atveju, būtinas atvirkštinės dispersijų-kovariacijų $n \times n$ matricos apskaičiavimas. Ribotą maksimalaus tikėtinumo metodą pasiūlė Cressie (1993). Tai alternatyva maksimalaus tikėtinumo metodui. MTLM metodas pasižymi ypač gerais rezultatais, kai analizuojama imtis yra maža, o nepastovaus vidurkio modelis (1.8) aprašomas aukšto laipsnio modelio parametru vektoriumi β (Smith 2000).

Cressie SMKMM atveju $W(\theta)$ yra įstrižainė (diagonalinė) matrica, kurios įstrižainėje yra $\hat{\gamma}$ komponentų dispersijos, t. y. $w_{jj} = \frac{2\gamma^2(\mathbf{h}_j; \theta)^2}{|N(\mathbf{h})|}$, čia $i = 1, \dots, K$.

Cressie (1993) pataria naudoti robastinį svorinį mažiausių kvadratų metodą. Diggle *et al.* (2003) pirmenybę teikia tikimybėmis paremtais metodais, kurie yra optimalūs modelio prielaidoms. Mažų imčių atveju ir/arba kai trendo paviršius turi daug parametru MTLM metodas pateikia mažiau nepaslinktus semivariogramų įverčius, nei MT metodas (Cressie 1993). Iš kitos pusės Diggle *et al.* (2003) teigia, kad MTLM labiau nei MT jautrus neteisingai parinktam trendo modeliui (Heikkinen 2003).

Apibendrinta semivariogramų modeliavimo etapų schema pateikta (2.11 pav.). Pastebėtina tai, kad tai vienas svarbiausių geostatistinių duomenų modeliavimo ir prognozavimo etapų, dažnai reikalaujantis pakankamai daug laiko resursų ir gerų nagrinėjamos srities žinių.

Semivariogramos modeliavimo kokybės įtaka prognozei aptariama 2.6 skyrelyje.

2.4. Krigingo tipai

Tokie duomenų prognozavimo metodai, kaip atvirkščiai proporcingo atstumo, atvirkščiai proporcingo atstumo kvadrato metodai ir kt., suteikia imties duomenims svorius, priklausomai nuo jų atstumo iki prognozuojamo taško. Tačiau šiuose metoduose neatsižvelgiama į duomenų panašumą lyginant su prognozuojamu tašku. Tuo tarpu kriginge teigiama, kad duomenys esantys arčiau pradinių imties duomenų yra labiau teigiamai koreliuoti.

Be to krigingas atlieka tikslią interpoliaciją, t. y. prognozuotos imties stebinių reikšmės, kurios buvo naudojamos nežinomų reikšmių prognozei, išlieka

nepakitusios. Krigingas taikomas projektuojant lėktuvus, automobilius, kompiuterinius vaizduoklius ir t. t. (Sacks *et al.* 1989; Meckesheimer *et al.* 2002; Simpson *et al.* (2001)) (van Beers 2004).

2.1 lentelė. Įvairūs krigingo metodai ir jų klasifikavimo priežastys
Table 2.1. Different kriging methods and clasification causes

		Krigingo metodo pavadinimas
Skirstymas pagal prognozuojamos srities dydį	Taškas	Taškinis
	Sritis	Blokinis
Kas yra žinoma	μ žinomas	Paprastas krigingas
	μ nežinomas, bet pastovus	Ordinarus krigingas
	$\mu = \mathbf{X}\beta$, β – nežinomas	Universalus krigingas
	$\mu(\mathbf{s}) = a + c(x) + r(y)$, $\mathbf{s}=(x, y)'$	Medianinio išlyginimo krigingas
Skirstinys	$\mathbf{Z}(\mathbf{s}) \sim \text{Gauso}$	Krigingas
	$\ln(\mathbf{Z}(\mathbf{s})) \sim \text{Gauso}$	Lognormalinis krigingas
	$\mathbf{Z}(\mathbf{s})$ – indikatorinis kintamasis	Indikatorinis krigingas
	Normalinis, pašalinus liekanas	Robastinis krigingas
Kintamųjų skaičius	Vienas kintamasis $\mathbf{Z}(\mathbf{s})$	Krigingas
	Daug kintamųjų $\mathbf{Z}_1(\mathbf{s}), \dots, \mathbf{Z}_n(\mathbf{s})$	Kokrigingas

Krigingo procedūros metu stebiniams suteikiami svoriai pagal tam tikrą svorinę funkciją. Svių suteikimo stebiniams ypatumai:

1. Artimi stebiniai įgauna didesnius svorius, nei labiau nutolę:
 - a) santykinė jų proporcija priklauso nuo stebinių vietos ir parametrinio semivariogramos modelio;
 - b) kuo didesnė grynuolio dispersija, tuo mažesni svoriai stebinių, esančių arčiau prognozuojamų reikšmių.

2. Grupotos reikšmės turi mažesnius svorius, nei izoliuotos tuo pačiu atstumu.

Daugeliu atveju 4–5 taškai, esantys arčiausiai prognozuojamos reikšmės, įgauna 80% visų likusių stebinių bendro svorio.

Taigi priklausomai nuo parinkto parametrinio semivariogramos modelio stebiniams nustatomi optimalūs svoriai taip, kad prognozuota nežinoma duomenų reikšmė būtų kuo tikslesnė. Kadangi semivariogramos reikšmės kinta priklausomai nuo atstumo, svoriai taip pat priklauso nuo stebinių skirstinio.

Priklausomai nuo duomenų skirstinio ir kitų faktorių skiriama keletas krigingo metodų (2.1 lentelė), dažnai naudojamos jų kombinacijos (universalus blokinis krigingas, lognormalinis kokrigingas ir t. t.) (Schabenberger, Pierce 2001). Pagal prognozuojamos srities apimtį krigingas skirstomas į blokinį bei taškinį. *Bloko krigingas* įvertina bloko reikšmę pagal artimiausius stebinius, tuo tarpu *taško krigingas*, remiantis, artimiausių stebinių vertėmis įvertina taško reikšmę. Krigingo prognozė taškui labai panaši į krigingo prognozė santykinai mažam blokui, bet tuo atveju apskaičiuotas krigingo standartinis nuokrypis yra labai didelis. Kai krigingu prognozuojamo taško koordinatės sutampa su tam tikro stebinio koordinatėmis, krigingo prognozė lygi imties duomenų reikšmei.

Pagal prognozuojamų duomenų savybes išskiriami: paprastas, ordinarus, universalus ir kt. krigingai:

2.4.1. Ordinalus, (ordinarus) blokinis, paprastas krigingai

Ordinarus krigingas – paprasčiausias ir dažniausiai naudojamas krigingas. Šio krigingo prielaidos: regioninis kintamasis yra stacionarus ir duomenyse nėra trendo (vidurkis – pastovus, bet nežinomas).

Remiantis šiomis prielaidomis bei naudojant stebinius $Z(s_i)$ ir tų stebinių svorinį vidurkį nežinoma duomenų reikšmė $Z(s_0)$ prognozuojama pagal formulę:

$$\hat{Z}_{OK}(s_0) = \sum_{i=1}^n \lambda_i Z(s_i), \quad (2.11)$$

čia $\hat{Z}_{OK}$ – ordinarus krigingo prognozė, λ_i – i -tojo kintamojo stebinio matavimo vietoje s_i svoris, o n – stebinių kiekis.

Teigiama, kad krigingo prognozė yra geriausia tiesinė nepaslinkta (*angl. BLUP – Best Linear Unbiased Prediction*), jei nagrinėjamas procesas yra stacionarus ir parinkta tinkamas semivariogramos modelis. Pasak Isaaks ir

Srivastava (1989), krigingu gaunama geriausia tiesinė nepaslinkta prognozė, jei tenkinamos 3 sąlygos:

- Kringas yra tiesinis įvertintojas, t. y.

$$\hat{Z}(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i) \quad (2.12)$$

yra svorinė tiesinė stebinių kombinacija, o λ_i – i-tojo stebinio svoris ir jis parinktas taip, kad minimizuoja vidutinę kvadratinę klaidą:

$$MSE = E\left\{[\hat{Z}(\mathbf{s}_0) - Z(\mathbf{s}_0)]^2\right\}. \quad (2.13)$$

- Kringu gauta prognozė yra nepaslinkta:

$$E[\hat{Z}(\mathbf{s}_0) - Z(\mathbf{s}_0)] = 0. \quad (2.14)$$

Vadinasi $E[\hat{Z}(\mathbf{s}_0)] = E[Z(\mathbf{s}_0)]$ prie visų μ . Tuomet $E[Z(\mathbf{s})] = \mu$ ir

$$E[\hat{Z}(\mathbf{s}_0)] = E\left[\sum_{i=1}^n \lambda_i Z(\mathbf{s}_i)\right] = \sum_{i=1}^n \lambda_i E[Z(\mathbf{s}_i)] = \mu \sum_{i=1}^n \lambda_i. \quad (2.15)$$

Taigi $\hat{Z}(\mathbf{s}_0)$ yra nepaslinktas $Z(\mathbf{s}_0)$ atžvilgiu, jei ir tik jei:

$$\sum_{i=1}^n \lambda_i = 1. \quad (2.16)$$

Taip pat:

$$E\left\{([\hat{Z}(\mathbf{s}_0) - Z(\mathbf{s}_0)]^2\right\} = \text{var}[\hat{Z}(\mathbf{s}_0) - Z(\mathbf{s}_0)],$$

o kadangi $\hat{Z}(\mathbf{s}_0)$ yra nepaslinktas $Z(\mathbf{s}_0)$ atžvilgiu, tuomet

$$\begin{aligned} E\left\{([\hat{Z}(\mathbf{s}_0) - Z(\mathbf{s}_0)]^2\right\} &= \text{var}[\hat{Z}(\mathbf{s}_0) - Z(\mathbf{s}_0)] = \\ &= \text{var}[\hat{Z}(\mathbf{s}_0)] + \text{var}[Z(\mathbf{s}_0)] - 2\text{cov}[\hat{Z}(\mathbf{s}_0), Z(\mathbf{s}_0)] = \end{aligned}$$

$$\begin{aligned}
&= \text{var} \left[\sum_{i=1}^n \lambda_i Z(\mathbf{s}_i) \right] + \text{var}[Z(\mathbf{s}_0)] - 2 \text{cov} \left[\sum_{i=1}^n \lambda_i Z(\mathbf{s}_i), Z(\mathbf{s}_0) \right] = \\
&= \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \text{cov}(\mathbf{s}_i - \mathbf{s}_j) + \sigma^2 - 2 \sum_{i=1}^n \lambda_i \text{cov}(\mathbf{s}_i - \mathbf{s}_0).
\end{aligned}$$

Iš semivariogramos ir kovariacinės funkcijos ryšio, pateikto (1.5) formulėje, išplaukia, kad vidutinė kvadratinė klaida yra:

$$\begin{aligned}
&E \left\{ ([Z(\mathbf{s}_0) - \hat{Z}(\mathbf{s}_0)])^2 \right\} = \\
&= \sigma^2 + \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j (\sigma^2 - \gamma(\mathbf{s}_i - \mathbf{s}_j)) - 2 \sum_{i=1}^n \lambda_i (\sigma^2 - \gamma(\mathbf{s}_i - \mathbf{s}_0)) = \\
&= \sigma^2 + \sigma^2 - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(\mathbf{s}_i - \mathbf{s}_j) - 2\sigma^2 + 2 \sum_{i=1}^n \lambda_i \gamma(\mathbf{s}_i - \mathbf{s}_0) = \\
&= - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(\mathbf{s}_i - \mathbf{s}_j) + 2 \sum_{i=1}^n \lambda_i \gamma(\mathbf{s}_i - \mathbf{s}_0).
\end{aligned}$$

- Krigingas turi minimizuoti įvertio klaidos dispersiją

$$\text{Var}[\hat{Z}(\mathbf{s}_0) - Z^*(\mathbf{s}_0)] \rightarrow \min.$$

Krigingo formulę perrašius matricų pavidalu gauname:

$$\begin{aligned}
&\begin{bmatrix} 0 & \gamma(\mathbf{s}_1 - \mathbf{s}_2) & \gamma(\mathbf{s}_1 - \mathbf{s}_3) & \cdots & \gamma(\mathbf{s}_1 - \mathbf{s}_n) & 1 \\ \gamma(\mathbf{s}_2 - \mathbf{s}_1) & 0 & \gamma(\mathbf{s}_2 - \mathbf{s}_3) & \cdots & \gamma(\mathbf{s}_2 - \mathbf{s}_n) & 1 \\ \gamma(\mathbf{s}_3 - \mathbf{s}_1) & \gamma(\mathbf{s}_3 - \mathbf{s}_2) & 0 & \cdots & \gamma(\mathbf{s}_3 - \mathbf{s}_n) & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma(\mathbf{s}_n - \mathbf{s}_1) & \gamma(\mathbf{s}_n - \mathbf{s}_2) & \gamma(\mathbf{s}_n - \mathbf{s}_3) & \cdots & 0 & 1 \\ 1 & 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \times \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \vdots \\ \lambda_n \\ \psi \end{bmatrix} = \\
&= \begin{bmatrix} \gamma(\mathbf{s}_1 - \mathbf{s}_0) \\ \gamma(\mathbf{s}_2 - \mathbf{s}_0) \\ \gamma(\mathbf{s}_3 - \mathbf{s}_0) \\ \vdots \\ \gamma(\mathbf{s}_n - \mathbf{s}_0) \\ 1 \end{bmatrix}, \tag{2.17}
\end{aligned}$$

vadinasi:

$$\mathbf{A}_{ok} = \lambda_{ok} \mathbf{b}_{ok},$$

čia $\mathbf{A}_{ok} = \begin{pmatrix} \mathbf{\Gamma} & \mathbf{1} \\ \mathbf{1}' & 0 \end{pmatrix}$, $\boldsymbol{\lambda}_{ok} = \begin{bmatrix} \boldsymbol{\Lambda} \\ \psi \end{bmatrix}$, o $\mathbf{b}_{ok} = \begin{bmatrix} \mathbf{\Gamma}_0 \\ 1 \end{bmatrix}$.

Iš šios formulės galima paskaičiuoti:

$$\boldsymbol{\lambda}_{ok} = \mathbf{A}_{ok}^T \mathbf{b}_{ok}.$$

Šiuo atveju vidutinė kvadratinė prognozės klaida lygi:

$$\begin{aligned} MSPE &= -\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(\mathbf{s}_i - \mathbf{s}_j) + 2 \sum_{i=1}^n \lambda_i \gamma(\mathbf{s}_i - \mathbf{s}_0) = \\ &= \sum_{i=1}^n \lambda_i \left\{ \gamma(\mathbf{s}_i - \mathbf{s}_0) - \sum_{j=1}^n \lambda_j \gamma(\mathbf{s}_i - \mathbf{s}_j) \right\} + \sum_{i=1}^n \lambda_i \gamma(\mathbf{s}_i - \mathbf{s}_0), \end{aligned}$$

jei λ_i - tasis svoris tenkina lygybę:

$$\sum_{j=1}^n \lambda_j \gamma(\mathbf{s}_i - \mathbf{s}_j) + \psi = \gamma(\mathbf{s}_i - \mathbf{s}_0),$$

čia $i = 1, \dots, n$, o ψ – yra Lagranžo daugiklis, kuris priklauso nuo $\mathbf{s}_0$.

Vadinasi:

$$\gamma(\mathbf{s}_i - \mathbf{s}_0) - \sum_{j=1}^n \lambda_j \gamma(\mathbf{s}_i - \mathbf{s}_j) = \psi.$$

Taigi:

$$MSPE = \sum_{i=1}^n \lambda_i \gamma(\mathbf{s}_i - \mathbf{s}_0) + \psi = \sigma_k^2(\mathbf{s}_0). \quad (2.18)$$

VKPK (*angl. MSPE – Mean Squared Prediction Error*) – tai prognozių pasiskirstymas apie tikrąją reikšmę ir yra vadinamas *prognozės dispersija* arba *klaida*:

Vietoj to, kad atlikti prognozę konkrečiame taške ar taškuose, galima atlikti prognozę tam tikro dydžio bloke (srityje) B. Tuomet *blokinio ordinaraus krigingo* tiesinis įvertinys:

$$\hat{Z}(B) = \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i). \quad (2.19)$$

Šiuo atveju blokinio ordinaraus krigingo sistema:

$$\begin{cases} \sum_{i=1}^n \lambda_i \gamma(\mathbf{s}_i - \mathbf{s}_j) + \psi(B) = \bar{\gamma}(\mathbf{s}_i, B) \\ \sum_{i=1}^n \lambda_i = 1 \end{cases}, \quad (2.20)$$

čia

$$\bar{\gamma}(\mathbf{s}_i, B) \approx \frac{1}{|B|} \sum_{i=1}^n \lambda_i \gamma(\mathbf{s}_i, \mathbf{s}), \quad (2.21)$$

o $|B|$ – bloko sritis, o $\mathbf{s}$ – taškas to bloko srityje, tad $\gamma(\mathbf{s}_i - \mathbf{s})$ – semivariograma tarp imties taškų ir prognozuojamo bloko (Rossiter 2004).

Kai duomenų vidurkis yra žinomas ir lygus tam tikrai konstantai, duomenų prognozei taikomas *paprastas krigingas*, kurio prognozė skaičiuojama taip:

$$\hat{Z}_{PK}(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i) - \mu \left(1 - \sum_{i=1}^n \lambda_i\right). \quad (2.22)$$

$\hat{Z}(\mathbf{s}_0)$ įvertinys vadinamas nepaslinktu, jei:

$$\sum_{i=1}^n \lambda_i = 1,$$

o

$$E\left[\hat{Z}(\mathbf{s}_0) - Z^*(\mathbf{s}_0)\right] = 0.$$

Paprasto krigingo sistema matricos pavidalu:

$$\begin{bmatrix} \gamma(\mathbf{s}_1 - \mathbf{s}_1) & \gamma(\mathbf{s}_1 - \mathbf{s}_2) & \dots & \gamma(\mathbf{s}_1 - \mathbf{s}_n) \\ \gamma(\mathbf{s}_2 - \mathbf{s}_1) & \gamma(\mathbf{s}_2 - \mathbf{s}_2) & \dots & \gamma(\mathbf{s}_2 - \mathbf{s}_n) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(\mathbf{s}_n - \mathbf{s}_1) & \gamma(\mathbf{s}_n - \mathbf{s}_2) & \dots & \gamma(\mathbf{s}_n - \mathbf{s}_n) \end{bmatrix} \times \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix} = \begin{bmatrix} \gamma(\mathbf{s}_1 - \mathbf{s}_0) \\ \gamma(\mathbf{s}_2 - \mathbf{s}_0) \\ \vdots \\ \gamma(\mathbf{s}_n - \mathbf{s}_0) \end{bmatrix}. \quad (2.23)$$

Vadinasi: $\mathbf{A}_{PK} \boldsymbol{\lambda}_{PK} = \mathbf{b}_{PK}$.

Prognozės klaida paprastojo krigingo atveju:

$$\sigma_{PK}^2(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i \gamma(\mathbf{s}_i - \mathbf{s}_0). \quad (2.24)$$

2.4.2. Universalus krigingas

Ordinariame kriginge mes teigiame, kad regioninis kintamasis yra stacionarus. Tačiau dažnai regioniniai kintamieji yra nestacionarūs ir jų vidutinės reikšmės kinta erdvėje.

Geostatistiniai metodai nestacionariems duomenims gali būti klasifikuojami priklausomai nuo to, kokios yra pradinių duomenų savybės:

- universalus krigingas (terminą pirmiausiai apibrėžė Matheron (1969));
- krigingas su išoriniu trendu (*angl. Kriging with External Drift*), kai išorinis trendas aprašomas kai kuriais papildomais kintamaisiais;
- regresinis arba krigingas po trendo pašalinimo, kai liekanos ir trendas gali būti atskirai priglodinti ir vėliau sudėti (Goovaerts 1997).

Universalus krigingas, krigingas su išoriniu dreifu ir regresinis krigingas priklauso „hibridinių“, t. y. nestacionarių geostatistinių metodų grupei (Waters 1997). Tačiau dažnai skirtingos procedūros vadinamos vienu vardu arba skirtingais vardais vadinama ta pati procedūra. Taip tikriausiai atsitiko dėl to, kad nagrinėjami procesai naudojami skirtingų sričių specialistų ir skirtingais tikslais.

Taigi universalus krigingo atveju nagrinėjamas erdvinis procesas $Z(\mathbf{s}_i)$ nėra stacionarus (t. y. netenkinama antros eilės stacionarumo sąlyga), tuomet proceso Z stebinio taške $\mathbf{s} \in D$ modelis aprašomas (1.2), o vidurkio funkcija turi pavidalą (1.8). Tad:

$$Z(\mathbf{s}_i) = \sum_{q=0}^Q \beta_q f_q(\mathbf{s}_o) + \varepsilon(\mathbf{s}_i), \quad (2.25)$$

(žr. Rossiter 2004).

Tuomet universalus krigingo sistema lygčių pavidalu:

$$\begin{cases} \sum_{j=1}^n \lambda_j \gamma(\mathbf{s}_i - \mathbf{s}_j) + \psi_0 + \sum_{q=1}^Q \psi_q f_q(\mathbf{s}_i) = \gamma(\mathbf{s}_o, \mathbf{s}_i) \\ \sum_{i=1}^n \lambda_i = 1 \\ \sum_{i=j}^n \lambda_i f_q(\mathbf{s}_j) = f_q(\mathbf{s}_o) \end{cases}, \quad (2.26)$$

o universalus krigingo sistema matricų pavidalu:

$$\begin{bmatrix}
\gamma(\mathbf{s}_1 - \mathbf{s}_1) & \cdots & \gamma(\mathbf{s}_1 - \mathbf{s}_n) & 1 & f_1(\mathbf{s}_1) & \cdots & f_Q(\mathbf{s}_1) \\
\vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\gamma(\mathbf{s}_n - \mathbf{s}_1) & \cdots & \gamma(\mathbf{s}_n - \mathbf{s}_n) & 1 & f_1(\mathbf{s}_n) & \cdots & f_Q(\mathbf{s}_n) \\
1 & \cdots & 1 & 0 & 0 & \cdots & 0 \\
f_1(\mathbf{s}_1) & \cdots & f_1(\mathbf{s}_n) & 0 & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
f_Q(\mathbf{s}_1) & \cdots & f_Q(\mathbf{s}_n) & 0 & 0 & \cdots & 0
\end{bmatrix} \times \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \psi_0 \\ \psi_1 \\ \vdots \\ \psi_Q \end{bmatrix} = \begin{bmatrix} \gamma(\mathbf{s}_1 - \mathbf{s}_0) \\ \vdots \\ \gamma(\mathbf{s}_n - \mathbf{s}_0) \\ 1 \\ f_1(\mathbf{s}_0) \\ \vdots \\ f_Q(\mathbf{s}_0) \end{bmatrix}. \quad (2.27)$$

Vadinasi:

$$\mathbf{A}_{uk} \boldsymbol{\lambda}_{uk} = \mathbf{b}_{uk},$$

čia $\mathbf{A}_{uk} = \begin{pmatrix} \boldsymbol{\Gamma} & \mathbf{1} & \mathbf{X} \\ \mathbf{1}^T & 0 & \mathbf{0}^T \\ \mathbf{X}^T & \mathbf{0} & \mathbf{0} \end{pmatrix}$, $\boldsymbol{\lambda}_{uk} = \begin{bmatrix} \boldsymbol{\Lambda} \\ \psi_0 \\ \boldsymbol{\Psi} \end{bmatrix}$, o $\mathbf{b}_{uk} = \begin{bmatrix} \boldsymbol{\Gamma}_0 \\ 1 \end{bmatrix}$.

Universalaus krigingo prognozės dispersija:

$$\sigma_{UK}^2 = \sum_{j=1}^n \lambda_j \gamma(\mathbf{s}_i - \mathbf{s}_0) + \psi_0 + \sum_{q=1}^Q \psi_q f_q(\mathbf{s}_0), \quad (2.28)$$

čia $\boldsymbol{\Psi} = (\psi_0, \dots, \psi_l)^T$ yra Lagranžo daugiklių vektorius. Detaliau aprašyta Cressie (1993).

Semivariograma nestacionariems procesams netenkina savybių, kurios aptartos 1.4 skyrelyje. Tuomet semivariogramą galima perrašyti tokiu būdu:

$$\gamma(\mathbf{h}) = \frac{1}{2|N(\mathbf{h})|} \left[\sum_{(\mathbf{s}_i, \mathbf{s}_j) \in N(\mathbf{h})} \left\{ (Z(\mathbf{s}_i + \mathbf{h}) - Z(\mathbf{s}_i)) - \frac{(Z(\mathbf{s}_i + \mathbf{h}) - Z(\mathbf{s}_i))}{|N(\mathbf{h})|} \right\}^2 \right]. \quad (2.29)$$

Ši semivariogramos formulė susideda iš dviejų dalių, kurių pirmoji yra skirtumas tarp stebinių poroje, o antroji – šių skirtumų vidurkis. Jei regioninis

kintamasis yra stacionarus, antroji dalis pradingsta, bet jei nėra stacionarus – tas vidurkis įgauna tam tikrą reikšmę. Todėl regioninis kintamasis gali būti pertvarkytas į dvi dalis, liekanų ir trendo. Trendas yra taškų svertinis vidurkis tarp kaimyninių taškų apie nežinomą reikšmę. Liekanos yra skirtumas tarp regioninio kintamojo ir trendo.

Taigi universalus krigingo metodas gali būti atliekamas dviem būdais:

1. Paviršius aproksimuojamas stebinių geografinių koordinačių globalių polinomų plėtimu. Polinominės funkcijos koeficientai randami mažiausių kvadratų metodu, užtikrinant, kad kvadratinų nuokrypių nuo paviršiaus trendo suma yra mažiausia. Įvertintas trendas, laikomas determinuotuoju stochastinio proceso komponentu. Pašalinus trendą gautos liekanos nagrinėjamos kaip regioninis kintamasis. Jų prognozei gali būti taikomas ordinarus krigingas. Prognozės rezultatu bus irgi liekanos. Atlikus liekanų prognozę, trendas pridedamas prie gauto rezultato.

2. Globalus trendas įtraukiamas į krigingo lygčių sistemą naudojamą tam, kad paskaičiuoti svorius, reikalingus krigingo prognozei.

Tam, kad atlikti universalų krigingą pirmuoju būdu, semivariogramoms reikalingos tikrosios liekanos, kurios negali būti žinomos, kol nėra nustatytas polinominio paviršiaus tipas. Ši problema apibūdinama kaip semivariogramų neapibrėžtumas ir nagrinėjama Chauvet bei Galli (1982).

Kadangi universaliame kriginge įvertinami ne tik svoriai, bet ir trendo koeficientai, patariama prognozei naudoti didesnę stebinių kiekį aplink prognozuojama tašką. Tačiau nėra aišku kiek kaimyninių taškų turi būti parinkta prognozei. Todėl pasirinktinai parenkami kaimyniniai taškai, kurių trendas gali būti paskaičiuotas. Tuomet liekanoms priglodinama reikiama semivariograma ir prognozuojama nežinoma (–os) reikšmė (–ės).

Universalaus krigingo metodas yra labai lankstus, tačiau kelia didelius reikalavimus papildomai informacijai. Šis krigingas pateikia tikslesnę prognozę nei ordinarus krigingas, tačiau jo rezultatai yra blogesni, jei trendo funkcija blogai apibrėžta.

2.4.3. Medianinio išlyginimo krigingas

Medianinio išlyginimo krigingas – įvertinimo metodas, skirtas sudaryti nepaslinktas prognozes, kai pradiniuose, gardelės tipo duomenyse yra trendas. Medianinio išlyginimo metodą pasiūlė Tukey (1977). Cressie šį metodą pritaikė krigingo metodui. Čia medianinio išlyginimo procedūra iš pradinių duomenų pašalinamas trendas, teigiant, kad liekanos yra stacionarios. Liekanoms parenkamas ir priglodinamas vienas iš parametrinių semivariogramų modelių bei taškiniams ar bloko duomenims pritaikomas ordinarus krigingas. Tokiu būdu, panašiai kaip ir universalaus krigingo atveju, atsitiktinė funkcija $Z(s_i)$ suskaidoma

į deterministinį trendą ir atsitiktinę liekanų komponentę. Tik medianinio krigingo atveju trendo paviršius, $\mu(\mathbf{s})$, naudojamas kriginge yra aprašomas taip:

$$\mu(\mathbf{s}) = a + r(x) + c(y), \mathbf{s} = (x, y), \quad (2.30)$$

čia a – bendras medianų efektas, c – stulpelių medianų efektas, r – eilučių medianų efektas. Pasak Cressie (1993) medianinio krigingas lyginant su universalioju yra lankstesnis ir statistiškai patikimesnis metodas, be to universalus krigingas ribojamas polinominiu trendo paviršiumi. Trendo pašalinimas iš duomenų medianiniu krigingu yra robustinis ir nejautrus išskirtinėms duomenų reikšmėms.

Medianinio išlyginimo procedūrai atlikti duomenys turi būti suskaidyti į gardelę (į stulpelius ir eilutes). Jei į tam tikrą gardelės narvelį patenka kelios reikšmės, tuomet galima paskaičiuoti tų reikšmių vidurkį, medianą ar modą. Kiekvienas stulpelis ir eilutė, kaip ir Dekarto koordinatų sistemoje, numeruojami gardelės kairiuoju apatiniu kraštu. Duomenų matrica papildoma nauju stulpeliu ir eilute, kuriuose talpinami medianinio išlyginimo procedūros rezultatai.

Medianinio išlyginimo procedūra susideda iš kelių etapų:

1. Kiekvienai matricos eilutei paskaičiuojama matricos mediana ir užrašoma eilutės pabaigoje.
2. Paskaičiuota mediana atimama iš eilutės duomenų. 1–2 punktai kartojami su kiekviena matricos eilute.
3. Nustatoma visų eilučių medianų mediana ir fiksuojama kaip bendras medianinis efektas.
4. Kiekvienam stulpeliui paskaičiuojama stulpelio mediana ir užrašoma stulpelio apačioje.
5. Paskaičiuota mediana atimama iš stulpelio duomenų. 4–5 punktai kartojami su kiekvienu matricos stulpeliu.
6. Pagaliau nustatoma visų stulpelių medianų mediana ir pridedama prie prieš tai skaičiuoto bendro medianinio efekto.

Taigi, galima įvertinti bendrą medianinį, stulpelių medianų ir eilučių medianų efektus taip:

$$\hat{a} = \frac{1}{n} \sum_{i=1}^n Z(\mathbf{s}_i),$$

$$\hat{c}(x) = \hat{c}_k = \frac{1}{q} \sum_{N_k} Z(\mathbf{s}_i), \text{ čia } k = 1, 2, \dots, q,$$

$$\hat{r}(y) = \hat{c}_l = \frac{1}{p} \sum_{M_l} Z(\mathbf{s}_i), \text{ čia } l = 1, 2, \dots, p,$$

čia N_k taškai, esantys k – tajame gardelės stulpelyje, o M_l – taškai esantys l – oje gardelės eilutėje.

Jei gardelės narvelyje nėra reikšmės, tas narvelis ignoruojamas skaičiuojant eilutės ar stulpelio, kuriame randasi tas narvelis, medianą. Jei eilutės ar stulpelio visi narveliai tušti, eilutės ar stulpelio efektas skaičiuojamas vidurkinant prieš tai buvusių eilučių ar stulpelių medianų efektus. Tai yra būtina tam, kad turėti eilučių ar stulpelių medianas visose prognozuojamos srities lokacijose.

Medianinio išlyginimo procedūrą kartais tenka pakartoti kelis kartus, siekiant įsitikinti, kad gaunamas tas pats rezultatas, nepaisant to ar pirma skaičiuojamos eilučių, ar stulpelių medianos. Procedūros pabaigoje gaunamos liekanos, bendras medianų efektas, stulpelių medianų efektas ir eilučių medianų efektas. Atlikus liekanų prognozę kriginu, liekanos sudedamos su eilučių, stulpelių bei bendrais medianiniais efektais tam, kad gauti interpoliuotą paviršių pradinių duomenų reikšmių ribose. Tad, liekanos r_e medianinio kriginio išlyginimo procedūra gaunamos taip:

$$r_e = Z(s_i) - \hat{a} - \hat{c}_k - \hat{r}_l. \quad (2.31)$$

Nustatant semivariogramos parametrus liekanoms, žinoma išskyla paklaidos (*angl. Bias*) problema. Tai ypač aktualu taikant universalų kriginą. Tačiau ši problema nėra tokia aktuali medianinio išlyginimo atveju.

2.4.4. Kokrigingas

Daugiamačiu atveju geostatistikoje įvertinimas arba prognozė atliekama naudojant kelias erdvines ir tarpusavyje koreliuotas funkcijas.

Kokrigingas (*angl. Cokriging*) – tai vienas iš prognozės metodų, kai pirminis kintamasis prognozuojamas naudojant papildomą antrinio kintamojo informaciją.

Sakykime turime du tarpusavyje susijusius regioninius kintamuosius $Z_1(s)$ ir $Z_2(s)$ ($p = 2$) ir $\hat{Z}$ prognozė atliekama remiantis tų kintamųjų matavimais. Tuomet tiesinė kokrigingo prognozė:

$$\hat{Z}_{KO}(s_0) = \sum_{k=1}^n \lambda_1^k Z_1(s_k) + \sum_{l=1}^m \lambda_2^l Z_2(s_l). \quad (2.32)$$

Gauta prognozė bus nepaslinkta tuomet, kai $\sum_{k=1}^n \lambda_1^k = 1$, o $\sum_{l=1}^m \lambda_2^l = 0$.

Reiktų pastebėti, kad kintamųjų $Z_1(s)$ ir $Z_2(s)$ matavimų kiekis gali nesutapti.

Kokrigingo sistema matricių pavidalu:

$$\mathbf{A}_{ko} \boldsymbol{\lambda}_{ko} = \mathbf{b}_{ko},$$

čia

$$\mathbf{A}_{ko} = \begin{bmatrix} & & & & & & 1 & 0 \\ & & & & & & 1 & 0 \\ & & \mathbf{\Gamma}_{kk} & \mathbf{\Gamma}_{kl} & & & \vdots & \vdots \\ & & & & & & 1 & 0 \\ & & & & & & 1 & 0 \\ & & & & & & 0 & 1 \\ & & & & & & 0 & 1 \\ & & \mathbf{\Gamma}_{lk} & \mathbf{\Gamma}_{ll} & & & \vdots & \vdots \\ & & & & & & 0 & 1 \\ & & & & & & 0 & 1 \\ 1 & \dots & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & \dots & 0 & 1 & \dots & 1 & 0 & 0 \end{bmatrix}, \quad (2.33)$$

$$\mathbf{\Gamma}_{ij} = \begin{bmatrix} \gamma_{kl}(\mathbf{s}_1 - \mathbf{s}_1) & \gamma_{kl}(\mathbf{s}_1 - \mathbf{s}_2) & \dots & \gamma_{kl}(\mathbf{s}_1 - \mathbf{s}_n) \\ \gamma_{kl}(\mathbf{s}_2 - \mathbf{s}_1) & \gamma_{kl}(\mathbf{s}_2 - \mathbf{s}_2) & \dots & \gamma_{kl}(\mathbf{s}_2 - \mathbf{s}_n) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{kl}(\mathbf{s}_n - \mathbf{s}_1) & \gamma_{kl}(\mathbf{s}_n - \mathbf{s}_2) & \dots & \gamma_{kl}(\mathbf{s}_n - \mathbf{s}_n) \end{bmatrix}, \quad (2.34)$$

o

$$\mathbf{b}_{ko} = (\boldsymbol{\gamma}_{kk}, \boldsymbol{\gamma}_{kl}, 1, 0)^T,$$

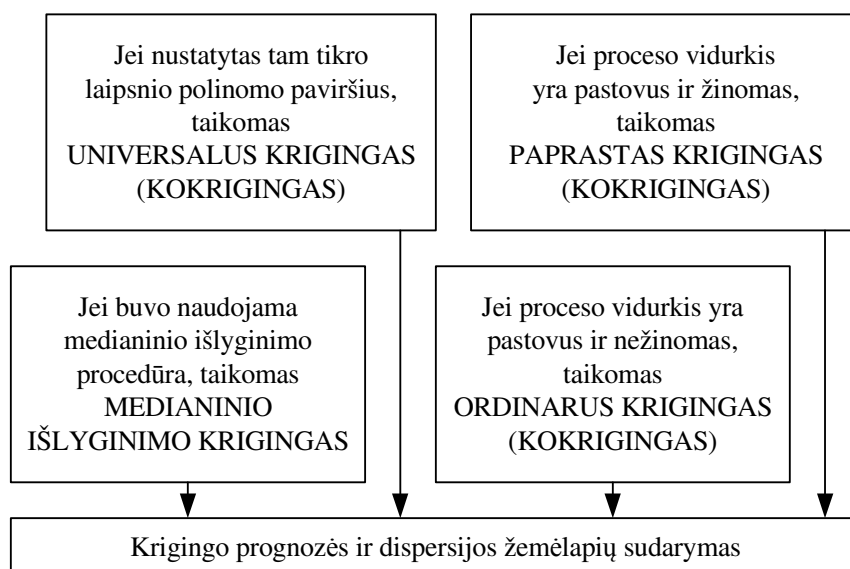
čia

$$\begin{aligned} \boldsymbol{\gamma}_{kk} &= (\gamma_{kk}(\mathbf{s}_1 - \mathbf{s}_0), \gamma_{kk}(\mathbf{s}_2 - \mathbf{s}_0), \dots, \gamma_{kk}(\mathbf{s}_n - \mathbf{s}_0))^T, \\ \boldsymbol{\gamma}_{kl} &= (\gamma_{kl}(\mathbf{s}_1 - \mathbf{s}_0), \gamma_{kl}(\mathbf{s}_2 - \mathbf{s}_0), \dots, \gamma_{kl}(\mathbf{s}_n - \mathbf{s}_0))^T, \\ \boldsymbol{\lambda}_{ko} &= (\lambda_{1k}, \lambda_{2k}, \dots, \lambda_{nk}, \lambda_{1l}, \lambda_{2l}, \dots, \lambda_{nl}, \psi_k, \psi_l)^T. \end{aligned}$$

Šiuo atveju erdvinė funkcijos struktūra – įvertinta matrica, kurios įstrižainėje – semivariogramos arba autokovariacijos, o neįstrižainėse – kryžminės semivariogramos arba kryžminės kovariacijos. Šių matricų erdvių funkcijų struktūrų modeliavimas dar sudėtingesnis nei vienmačiu atveju; matricų struktūra turi tenkinti tam tikras teigiamo apibrėžtumo sąlygas, kurių pakanka tam, kad garantuoti matricos koeficientų apverstumą kriginio lygtyje (Helterbrand, Cressie 1994).

Krigingo tipai, atliekantys netiesinį įvertinimą, tokie kaip indikatorinis, tikimybinis, izofaktorinis disjunktyvus krigingas aprašomi Vann, Guibal (1998).

Šiame skyriuje nagrinėtų krigingo tipų taikymas erdvinių duomenų prognozavimui schematiškai vaizduojamas taip (2.13 pav.):



2.12 pav. Skirtingomis sąlygomis taikomi krigingai
Fig. 2.13. Using of krigings under different conditions

2.5. Modelių patikros metodai

Klasikinėje statistikoje norint nustatyti ar parinktas modelis atitinka duomenis, tikrinama tam tikro tipo hipotezė (atliekamas statistinis testas). Dėl stebinių erdvinės priklausomybės geostatistikoje labai sudėtinga konstruoti statistinį testą. Esant dideliems atstumams lyginant su teoriniu semivariogramos modeliu empirinė semivariograma gali labai kisti ir tai riboja statistinių testų taikymą. Switzer (1984) semivariogramos mastelio ir formos parametrų patikrinimui pasiūlė keletą metodų, kurių pagrindinis yra duomenis tiesiškai transformuoti į nekoreliuotus matavimus su pastovia dispersija (Chiles, Delfiner 1999).

Kryžminės patikros metodas

Kryžminės patikros metodas – tai dar vienas iš metodų, padedančių išsirinkti, kuris iš semivariogramų modelių ar prognozavimo (įvertinimo)

procedūrų yra tinkamiausia nagrinėjamiems duomenims. Yra keli kryžminės patikros būdai:

- *Leave-One-Out* metodas, kai iš duomenų rinkinio atsitiktinai pašalinamas $Z_{[i]}(\mathbf{s}_i)$ narys; likusių duomenų pagrindu atliekama pašalintojo nario prognozė ir remiantis tikraja ir prognozuota pašalintojo nario reikšme paskaičiuojama prognozės klaida (atliekamas testavimas). Testavimas pakartojamas visiems duomenims ir paskiausiai randama vidutinė prognozės klaida. Pagrindinis metodo trūkumas – viso duomenų rinkinio testavimas reikalauja didelių laiko sąnaudų.
- *k-Fold* metodas, kai duomenų rinkinys atsitiktinai padalinamas į k klasių. Testavimas atliekamas duomenų rinkiniui, pašalinus k_i klasę bei paskaičiuojama šios klasės prognozės klaida. Patikrinus visas k_i klases paskaičiuojama vidutinė prognozės klaida. Pagrindinis metodo trūkumas – jis nėra tinkamas mažo duomenų kiekio atveju. Tuo atveju gaila dalį duomenų aukoti testavimui.
- *Test Set* metodo atveju testavimo grupei atsitiktinai atrenkama 30% duomenų rinkinio taškų. Atlikus tų taškų prognozę likusiųjų duomenų pagrindu, prognozės rezultatai palyginami. Nors šis metodas pasižymi paprastumu, kaip ir *k-Fold* metodo atveju, metodas netinkamas mažoms imtims.

Tarkim $\hat{Z}(\mathbf{s}_i)$ žymėsime prognozuotą reikšmę, $Z(\mathbf{s}_i)$ – stebimą reikšmę, o $\hat{\sigma}(\mathbf{s}_i)$ – $\mathbf{s}_i$ matavimų vietos standartinę prognozės klaidą. Jei parinktas parametrinis semivariogramos modelis yra geras, tuomet tikėtina, kad skirtumas tarp tikrosios $Z(\mathbf{s}_i)$ ir prognozuotos $\hat{Z}(\mathbf{s}_i)$ reikšmės bus mažas, t. y. įvertinta reikšmė bus artima tikrajai. „Artumas“ gali būti nustatytas tam tikromis statistikomis ir grafikais. Keli jų:

- *Vidutinė prognozės paklaida (VPK) (angl. ME – Mean Error):*

$$VPK = \frac{\sum_{i=1}^n (\hat{Z}(\mathbf{s}_i) - Z(\mathbf{s}_i))}{n}. \quad (2.35)$$

Tinkamo semivariogramos ar krigingo modelio parinkimo atveju paskaičiuota reikšmė turi būti artima nuliui.

- *Šaknis iš vidutinės kvadratinės prognozės paklaidos (ŠVKPK) (angl. RMSPE – Root Mean Squared Prediction Error):*

$$\check{VKPK} = \sqrt{\frac{\sum_{i=1}^n (\hat{Z}(s_i) - Z(s_i))^2}{n}}. \quad (2.36)$$

Pageidautina, kad šios statistikos reikšmė būtų kuo mažesnė.

- Vidutinis kvadratinis nuokrypio santykis (VKNS) (angl. MSDR – Mean Squared Deviation Ratio):

$$VKNS = \frac{\sum_{i=1}^n \left[(\hat{Z}(s_i) - Z(s_i)) / \hat{\sigma}(s_i) \right]^2}{n}. \quad (2.37)$$

Idealiu atveju paskaičiuota reikšmė artima vienetui (Rossiter 2004).

- Prognozės paklaidų (liekanų) histograma (idealiu atveju histograma turi būti beveik simetrinė, neiškreipta, savo forma panaši į Gauso skirstinį).

- $\hat{Z}(s_i)$ ir $Z(s_i)$ priklausomybės diagrama, kuri idealiu atveju parodo stiprią koreliaciją.

- $Z(s_i) - \hat{Z}(s_i)$ ir $\hat{Z}(s_i)$ priklausomybės diagrama, kuri idealiu atveju vaizduoja taškų nepriklausomybę.

Paprasto krigingo atveju, jei duomenys pasiskirstę pagal Gauso skirstinį grafiškai galima patikrinti įverčio ir paklaidos ortogonalumą bei sąlyginį nepaslinktumą. Ordinaraus krigingo atveju, šios savybės nebūtinai galioja. Standartizuotų paklaidų histograma taip pat parodo, ar krigingo paklaidų skirstinys artimas Gauso skirstiniui. Čia pirmame plane vaizduojami taškai, kurie blogai paaiškinami jų kaimynais, o šie savo ruožtu gali būti anomaliniai. Standartizuotų klaidų kontūrinis žemėlapis parodo stacionarumo trūkumą (pvz., proporcingumo efektą) bei ar klaidos reikšmė yra homogeniška erdvėje. Pašalinus nustatytas problemas būtina pakartotinai perskaičiuoti semivariogramas ir pritaikyti kryžminės patikros metodą.

AIC kriterijus

Kitas metodas, kuriuo tikrinamas parinktas statistinis modelis yra Akaike informacijos kriterijus (angl. AIC – Akaike Information Criterion) (Akaike 1973). Šis kriterijus taikytas McBratney ir Webster (1986) ir detalčiau aprašytas Webster and McBratney (1989). Akaike informacijos kriterijus yra tokio pavidalo:

$$AIC = -2\ln(L) + 2m, \quad (2.38)$$

čia m – statistinio modelio parametrų kiekis, o L – įvertinamo modelio tikėtino funkcijos maksimali reikšmė.

AIC taip pat gali būti paskaičiuotas taikant regresijos liekanų kvadratų sumą (RSS):

$$AIC = n * \ln(RSS / n) + 2m, \quad (2.39)$$

čia n – stebinių kiekis. Teigiama, kad modelis, kuriam paskaičiuota AIC reikšmė yra mažiausia, geriausias. Kai stebinių skaičius n yra mažas ($\frac{n}{m} < 40$) (o tai dažna situacija dirbant su metamodeliais), AIC reikšmė būna per maža. Ši reikšmė gali būti pakoreguota (Martin, Simpson 2005):

$$AIC_{kor} = AIC + 2m(m+1) / (n - m - 1). \quad (2.40)$$

2.6. Semivariogramos parametrų įtaka prognozei. Krigingo trūkumai ir privalumai

Išanalizavus erdvinių duomenų modeliavimo etapus, pastebėta, kad semivariogramų modeliavimo etapas reikalauja pakankamai laiko resursų bei žinių apie nagrinėjamą sritį.

Zimmerman ir Zimmerman (1991) nustatė, kad semivariogramų įvertinimas atliekamas geriausiai tuomet, kai erdvinė priklausomybė yra silpna. Taip pat jie atliko eksperimentus, kurių metu gauta krigingo prognozė atrodo pakankamai gera, nors semivariogramos parametrinis modelis buvo parinktas klaidingai. Šiuos rezultatus patvirtino Cressie ir Zimmerman (1992), gavę, kad prognozė, paremta klaidingai parinkta semivariograma yra asimptotiškai efektyvi. Tam tikras suderinamumas remiasi tikimybiniais dydžiais ir atstumo intervalais. Pavyzdžiai pateikiami Handcock (1989).

Visgi nors parinkus klaidingą parametrinį semivariogramos modelį, krigingo prognozė mažai pakinta, į šiuos netikslumus labai jautriai reaguoja krigingo dispersija. Krigingas labiausiai jautrus semivariogramos elgsenai ties 0, o ypač situacijai, kai grynuolio efektas egzistuoja/neegzistuoja, o $\gamma'(\mathbf{h})$ prie $\mathbf{h}$ artimas 0.

Taikant parametrinius semivariogramų modelius yra svarbu žinoti modelių bei jų parametrų įtaka krigingo prognozei. *Gauso* modelis lyginant su eksponentiniu paskaičiuoja mažesnes semivariogramos reikšmes mažuose atstumuose. Tai reiškia, kad taikant *Gauso* modelį taškai arti vienas kito traktuojami kaip labiau koreliuoti, todėl nustatomi didesni svoriai λ_i , nei svoriai tiems patiems taškams, kai naudojamas eksponentinis modelis. Tad, taikant

Gauso modelį prognozė tampa labiau lokali lyginant su prognoze, gauta taikant eksponentinį modelį.

Kadangi semivariogramoje pasiekiamas pastovus slenksčio atstumas prie mažesnės h reikšmės, sferinio modelio įvertinys lokalesnis atžvilgiu įvertinio, gauto taikant *Gauso* modelį.

Jei grynuolio reikšmė padidėja, kai kiti semivariogramos parametrai išlieka pastovūs, krigingo svoriai būna labiau panašūs.

Slenksčiui pasikeitus, tuo metu kai kiti semivariogramos parametrai išlieka nepakitę, krigingo svoriai nekinta, bet krigingo dispersija keičiasi tiesiškai slenksčio pokyčio atžvilgiu. Didžiausia krigingo dispersija būna tuomet, kai duomenų taškai yra nekoreliuoti.

Slenksčio atstumo pokytis, kai kiti parametrai išlieka pastovūs, nežymiai sumažina krigingo svorius (Isaaks, Srivastava 1989): jei slenksčio atstumas sumažinamas, tuomet mažiau stebinių dalyvauja prognozėje ir gauta prognozė yra labiau lokali.

Geometrinės anizotropijos atveju slenksčio atstumas kinta skirtingai skirtingomis kryptimis, o didesni svoriai suteikiami didesnio slenksčio atstumo kryptimi, tuo tarpu kai mažesni svoriai – mažesnio slenksčio atstumo kryptimi.

Nagrinėjant krigingo trūkumus, pastebėtina tai, kad krigingo metodo praktinė realizacija gana sudėtinga. Be to krigingo procedūra ignoruoja staigius reikšmių šuolius, kurie dažnai sutinkama ekologiniuose duomenyse. Procedūra realizuota daugelyje GIS ir kitų programinių paketų, tad iškyla pavojus, kad krigingas dažnai gali būti panaudotas netinkamai. Tad dauguma uždavinių, nereikalaujančių ypatingo tikslumo, turėtų būti sprendžiami taikant paprastesnius deterministinius metodus.

Visgi krigingo metodas turi nemažai privalumų lyginant su kitais interpoliacijos metodais. Pagal teoriją, nėra nė vieno tokio metodo, kuris geriausiai prognozuotų paviršių (ta prasme, kad prognozė būtų nepaslinkta ir turėtų mažiausias paklaidas).

Tam tikruose tyrimuose (Creutin, Obled, 1982; Tabios, Salas, 1985; Rouhani, 1986; Grimm, Lynch, 1991; Weber, Englund, 1994; Laslett, 1994), krigingo įvertinimo rezultatai buvo geresni nei kitų interpoliacijos metodų, tuo tarpu kituose taikymuose (Laslett *et al.* 1987; Weber, Englund, 1992; Gallichand, Marcotte, 1993) interpoliacija splainais ar atvirkštinio atstumo metodais buvo tikslesnė. Isaaks ir Srivastava (1989) savo knygoje ordinaraus krigingo įvertinį lygina su kitais keturiais taško įvertinimo metodais (daugiakampio, trianguliaciniu, lokalaus imties vidurkio ir atvirkštinio atstumo). Kaip tikėtasi ordinariu krigingu gautas įvertinys turi mažesnę įverčio dispersiją.

Kaip jau buvo minėta, praktikoje rezultatų efektyvumas priklauso nuo teisingo semivariogramos parametrų ir trendo modelio aprašymo. Tačiau, kadangi krigingas yra tikslus, net parinkus nelabai gerus semivariogramos

parametrus, krigingas paviršių prognozuos geriau nei bet kuris kitas interpoliacijos metodas. Krigingo metodas pranašesnis lyginant su deterministinėmis procedūromis, kadangi krigingu pateikiamos prognozės paklaidos kiekviename prognozuoto paviršiaus taške.

Kiti krigingo privalumai:

- Priglodinimas: krigingo regresoriai (priglodintojai), įvertinti remiantis bendra imties dispersija sudaryta iš atsitiktinio triukšmo. Labiau triukšmingas duomenų rinkinys, vaizduoja mažiau individualių imties rinkinių tapatumą ir jie labiau priglodinami.
- Deklasterizacija: krigingo svoriai priskiriami imties duomenims yra mažinami, jei jie pateikia dubliuojančių, labai koreliuotų taškų informaciją.
- Anizotropija: kai imtys yra labai koreliuotos (priklausomos) tam tikra kryptimi, krigingo svoriai nustatomi didesni ta kryptimi nutolusiems taškams.
- Tikslumas: pagal duotą semivariogramą, kuri aprašo įvertinamą paviršių, Krigingas paskaičiuos tiksliausius įvertinamo paviršiaus įverčius duotiems duomenims. Be to atliekant prognozę taške, kuriame buvo

atliktas matavimas $\hat{Z}(s_i) = Z(s_i)$ (Olea 1975).

2.7. Antrojo skyriaus apibendrinimas. Papildomų disertacijos uždavinių formulavimas.

- Bendriniai geostatistinio modeliavimo etapai yra trys: (1) statistinė erdvinių duomenų analizė, (2) duomenų erdvinės elgsenos modeliavimas semivariogramomis, (3) duomenų prognozavimas vienu iš krigingo tipų.
- Statistinės erdvinių duomenų analizės metu tikrinama turimų duomenų atributinė informacija, atliekama loginė topologinės informacijos patikra, vykdoma duomenų analizė klasikinės statistikos metodais (duomenų skirstinio parinkimas, išskirčių radimas, trendo analizė ir pan.).
- Semivariogramų modeliavimas vykdomas keliais etapais. Pirmame etape dažniausiai Matheron'o momentų metodu (MoM) apskaičiuojama ir grafiškai atvaizduojama empirinė semivariograma, o pagal jos elgseną spėjama apie duomenų anizotropiškumą, trendą ar cikliškumą.
- Antrame etape brėžiamas teorinis duomenų slenkstis, lygus duomenų dispersijai – empirinės semivariogramos taškų žymus kylimas virš teorinio slenksčio yra dar vienas trendo egzistavimo duomenyse įrodymas. Pastebėta, kad dauguma programų, skirtų erdvinių duomenų

analizei, neturi priemonių, leidžiančių brėžti teorinį slenkstį semivariogramoje.

- Anizotropijos nustatymui brėžiamas semivariogramos žemėlapis, sudaromos kryptinės semivariogramos.
- Atsižvelgus į egzistuojantį trendą bei pašalinus anizotropiją, empiriniams semivariogramos taškams parenkamas tinkamiausias parametrinis semivariogramos modelis.
- Nustatyta, kad parametrinių semivariogramos modelių parametrų įvertinimui dažnai taikomi mažiausių kvadratų – paprastas mažiausių kvadratų (PMKM), apibendrintas mažiausių kvadratų (AMKM) bei svertinis mažiausių kvadratų (SMKM) – metodai. Prie bendrų sąlygų, apibendrintas mažiausių kvadratų metodas yra asimptotiškai efektyvus lyginant su kitais mažiausių kvadratų vertinimo metodais.
- Išanalizavus semivariogramos modelių parametrų vertinimo metodus pastebėta, kad dauguma jų pateikiami per kovariacines funkcijas. Kadangi apibendrintas mažiausių kvadratų metodas prie bendrų sąlygų yra asimptotiškai efektyvus, o semivariogramos aprašo platesnę erdvių procesų klasę, vienas iš papildomų darbo uždavinių – AMKM įvertinio išraiška užrašyti per semivariogramas.
- Kadangi kitas dažnas ir svarbus uždavinys praktiniuose tyrimuose yra optimalaus imčių plano parinkimas, vienas darbo uždavinių yra AMKM įvertiniui išvesti D ir A kriterijų išraiškas per semivariogramas.

Vidurkio parametrų ir semivariogramų įvertiniai

Kadangi semivariogramos aprašo platesnę erdvinių procesų klasę, kovariacinė atsitiktinio proceso struktūra dažniausiai išreiškiama per semivariogramas. Vertinant semivariogramos parametrus vieni populiariausių yra mažiausių kvadratų metodai. Statistinės mažiausių kvadratų metodu gauto semivariogramos parametro įvertinio savybės priklauso nuo mažiausių kvadratų atstumo funkcijos. Lahiri *et al.* autorių straipsnyje (2002) įrodyta, kad iš statistinio efektyvumo pusės, prie tam tikrų sąlygų, apibendrintas mažiausių kvadratų metodas lyginant su kitais mažiausių kvadratų metodais yra optimaliausias. Taikant šį metodą būtina žinoti klasikinio semivariogramos įvertinio asimptotinę dispersijų-kovariacijų matricą, o tai gali būti sudėtinga ne Gauso skirstinio atveju. Šio skyriaus pirmoje dalyje pateikiama β parametrų AMKM įvertinio išraiška per semivariogramas (2 lema). 3 lemoje β parametrų apibendrinto mažiausių kvadratų įvertinio D ir A kriterijų funkcijos išreiškiamos per semivariogramas.

Antroje dalyje siekiant nustatyti optimalų taškų išsidėstymą taisyklingoje 2×2 gardelėje, kurios centras $(0,0)$, atliekamas eksperimentas su 5 taškais. Taškų išsidėstymo optimalumui nustatyti taikomi D ir A kriterijai.

Trečioje skyriaus dalyje stacionaraus Gauso duomenų modelio atvejui pateikiama MoM semivariogramų įvertinio dispersijų-kovariacijų matricos išraiška per teorines semivariogramas. Vėliau to pačio duomenų modelio atveju didėjančios erdvės asimptotikos erdviniams imčių planui LIDA įrodomas MoM įvertinio asimptotinis normališkumas. Asimptotinės dispersijų-kovariacijų matricos išraiška išvedama izotropinės eksponentinės semivariogramos atvejui.

Skyriaus tematika paskelbtas vienas autorės straipsnis (Dučinskas, Borisenko 2009).

3.1. Vidurkio parametrų AMKM įvertinio išraiška per semivariogramas

Kaip buvo minėta ankstesniame skyrelyje, apibendrintas mažiausių kvadratų (AMKM) metodas taikomas ir kovariacinių funkcijų, ir semivariogramų parametrų įvertinimui. Tačiau literatūroje pateikiama AMKM įvertinio išraiška per kovariacines funkcijas. Užrašysime AMKM įvertinio išraišką per semivariogramas.

Tarkime, kad erdvinio proceso su stacionaria klaida n stebėjimų modelis yra $\mathbf{Z}_n = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_n$, čia $\mathbf{Z}_n = (Z(s_1), Z(s_2), \dots, Z(s_n))^T$ – požymio vektorius, $\boldsymbol{\beta} = (\beta_0, \dots, \beta_Q)^T$ – nežinomų parametrų vektorius, $\mathbf{X} = (\mathbf{1}_n, \mathbf{F})$ – regresorių matricą, o $\boldsymbol{\varepsilon}_n = (\varepsilon(s_1), \varepsilon(s_2), \dots, \varepsilon(s_n))^T$ – klaidų vektorius.

Tuomet remiantis ryšiu tarp semivariogramos ir kovariacinės funkcijos (1.5), semivariogramų matrica $\boldsymbol{\Gamma}$ bus apibrėžiama lygtimi:

$$\boldsymbol{\Gamma} = \mathbf{C}(\mathbf{0})\mathbf{J} - \boldsymbol{\Sigma}, \quad (3.1)$$

čia $\mathbf{J} = \mathbf{1}_n \mathbf{1}_n^T$, $\boldsymbol{\Sigma}$ – dispersijų-kovariacijų $n \times n$ eilės matrica (1.11).

Pradžioje išvesime aukščiau aprašytam erdviniam modeliui su nepastoviu vidurkiu $\hat{\boldsymbol{\beta}}$ parametro AMKM įvertinio išraišką per semivariogramas atvejui, kai $\boldsymbol{\theta}$ žinomas.

Pažymėkime $\mathbf{H}_{11} = \mathbf{e}_1 \mathbf{e}_1^T$, $\boldsymbol{\Gamma}^{\cdot\cdot} = \mathbf{1}_n^T \boldsymbol{\Gamma}^{-1} \mathbf{1}_n$, $\mathbf{e}_1^T = (1, 0, \dots, 0)$ – k -matis vektorius, kurio pirmasis elementas lygus 1, o visi likusieji – 0.

2 lema. Tarkime, kad $1 - \mathbf{C}(\mathbf{0})\boldsymbol{\Gamma}^{\cdot\cdot} \neq 0$, tada:

$$\hat{\boldsymbol{\beta}}_{\text{AMKM}}(\boldsymbol{\theta}) = (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{Z}_n = (\mathbf{X}^T \boldsymbol{\Gamma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Gamma}^{-1} \mathbf{Z}_n. \quad (3.2)$$

Įrodymas.

Yra žinoma, kad:

$$\hat{\beta}_{\text{AMKM}}(\theta) = (\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma^{-1} \mathbf{Z}_n. \quad (3.3)$$

Iš (3.1) išplaukia, kad $\Sigma = -\Gamma + \mathbf{C}(\theta)\mathbf{J}$.

Tuomet pagal sutrumpintą matricių apvertimo formulę

$$\{\mathbf{A} + \mathbf{u}\mathbf{v}^T\}^{-1} = \mathbf{A}^{-1}(\mathbf{I} - \mathbf{u}\mathbf{v}^T \mathbf{A}^{-1} / (1 + \mathbf{v}^T \mathbf{A}^{-1} \mathbf{u})) \quad (3.4)$$

gauname:

$$\Sigma^{-1} = -\Gamma^{-1}(\mathbf{I} + \mathbf{C}(\theta)\mathbf{J}\Gamma^{-1} / (1 - \mathbf{C}(\theta)\Gamma^{-1})), \quad (3.5)$$

Pažymėkime $\alpha = \frac{\mathbf{C}(\theta)}{(1 - \mathbf{C}(\theta)\Gamma^{-1})}$, tuomet:

$$\mathbf{C}(\theta) = \frac{\alpha}{\alpha \Gamma^{-1} + 1}, \quad (3.6)$$

o

$$\mathbf{X}^T \Sigma^{-1} \mathbf{X} = -\mathbf{X}^T \Gamma^{-1} \mathbf{X} + \mathbf{u}\mathbf{v}^T, \quad (3.7)$$

čia $\mathbf{u} = \mathbf{X}^T \Gamma^{-1} \mathbf{1}_n$, o $\mathbf{v}^T = \alpha \mathbf{u}^T$.

Vėlgi panaudojus formules (3.5), (3.7) bei tapatybę $(\mathbf{X}^T \Gamma^{-1} \mathbf{X})^{-1} (\mathbf{X}^T \Gamma^{-1} \mathbf{1}_n)^{-1} = \mathbf{e}_1$ gauname:

$$\begin{aligned} (\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1} &= -(\mathbf{X}^T \Gamma^{-1} \mathbf{X})^{-1} + \frac{(\mathbf{X}^T \Gamma^{-1} \mathbf{X})^{-1} \mathbf{u}\mathbf{v}^T (\mathbf{X}^T \Gamma^{-1} \mathbf{X})^{-1}}{1 + \mathbf{v}^T (\mathbf{X}^T \Gamma^{-1} \mathbf{X})^{-1} \mathbf{u}} = \\ &= -(\mathbf{X}^T \Gamma^{-1} \mathbf{X})^{-1} + \alpha \frac{(\mathbf{X}^T \Gamma^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Gamma^{-1} \mathbf{1}_n \mathbf{1}_n^T \Gamma^{-1} \mathbf{X}^T (\mathbf{X}^T \Gamma^{-1} \mathbf{X})^{-1}}{1 + \mathbf{1}_n^T \Gamma^{-1} \mathbf{X}^T (\mathbf{X}^T \Gamma^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Gamma^{-1} \mathbf{1}_n} = \\ &= -(\mathbf{X}^T \Gamma^{-1} \mathbf{X})^{-1} + \alpha \frac{\mathbf{e}_1 \mathbf{e}_1^T}{1 + \alpha \mathbf{1}_n^T \Gamma^{-1} \mathbf{X} \mathbf{e}_1} = -(\mathbf{X}^T \Gamma^{-1} \mathbf{X})^{-1} + \frac{\alpha H_{11}}{1 + \alpha \Gamma^{-1}}. \end{aligned}$$

Tuomet pasinaudoję (3.6) gauname:

$$(\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1} = -(\mathbf{X}^T \Gamma^{-1} \mathbf{X})^{-1} + \mathbf{C}(\theta) H_{11}. \quad (3.8)$$

Istatę (3.4), (3.8) į formulę (3.3) ir pasinaudoję tuo, kad $H_{11} \mathbf{X}^T = \mathbf{e}_1 \mathbf{1}_n^T$ gauname:

$$(\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma^{-1} = (\mathbf{X}^T \Gamma^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Gamma^{-1}.$$

Tuo ir baigiame lemos įrodymą.

Kaip žinoma iš klasikinės statistikos D kriterijaus funkcija $\Psi_D(\cdot)$ parametrų įvertiniams lygi šių įvertinių dispersijų-kovariacijų matricos determinantui. Taigi D kriterijaus funkcija įvertiniui $\hat{\beta}_{AMKM}$ yra:

$$\Psi_D(\hat{\beta}_{AMKM}) = \det\left((\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1}\right) = \frac{1}{\det(\mathbf{X}^T \Sigma^{-1} \mathbf{X})}, \quad (3.9)$$

$$\text{kadangi } V\left(\hat{\beta}_{AMKM}\right) = (\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1}.$$

Analogiškai A kriterijaus funkcija yra:

$$\Psi_A(\hat{\beta}_{AMKM}) = \text{tr}\left((\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1}\right). \quad (3.10)$$

Sekančioje lemoje pateiksime D ir A kriterijų funkcijų išraiškas per semivariogramas.

3 lema.

$$\text{a) } \Psi_D(\hat{\beta}_{AMKM}) = (-1)^q (1 - \mathbf{C}(\mathbf{0})G_{11}) / \det(\mathbf{X}^T \Gamma^{-1} \mathbf{X}),$$

čia G_{11} matricos $\mathbf{X}^T \Gamma^{-1} \mathbf{X}$ elementas, esantis pozicijoje (1,1).

$$\text{b) } \Psi_A(\hat{\beta}_{AMKM}) = -\text{tr}(\mathbf{X}^T \Gamma^{-1} \mathbf{X})^{-1} + \mathbf{C}(\mathbf{0}).$$

Įrodymas:

Įrodyme naudosimės formule (3.8). Matricų teorijoje yra žinoma tokia formulė:

$$\det(\mathbf{A} + \mathbf{u}\mathbf{v}^T) = \det \mathbf{A} + \mathbf{v}^T \text{adj } \mathbf{A} \mathbf{u}, \quad (3.11)$$

čia $\mathbf{A}$ – neišsigimusi kvadratinė matrica, $\mathbf{u}$ ir $\mathbf{v}$ – vektoriai, o $\text{adj } \mathbf{A} = \det \mathbf{A} \cdot \mathbf{A}^{-1}$ – matricos $\mathbf{A}$ adjunktų matrica.

Pritaikę formulę (3.11) formulėje (3.8) gauname:

$$\begin{aligned} & \det\left(-(\mathbf{X}^T \Gamma^{-1} \mathbf{X})^{-1} + \mathbf{C}(\mathbf{0})\mathbf{H}_{11}\right) = \\ & = (-1)^q \left(\det(\mathbf{X}^T \Gamma^{-1} \mathbf{X})^{-1} - \mathbf{C}(\mathbf{0})\mathbf{e}_1^T \text{adj}(\mathbf{X}^T \Gamma^{-1} \mathbf{X})^{-1} \mathbf{e}_1 \right) = \\ & = (-1)^q \left(\frac{1}{\det(\mathbf{X}^T \Gamma^{-1} \mathbf{X})} - \mathbf{C}(\mathbf{0}) \frac{\mathbf{e}_1^T (\mathbf{X}^T \Gamma^{-1} \mathbf{X}) \mathbf{e}_1}{\det(\mathbf{X}^T \Gamma^{-1} \mathbf{X})} \right) = \end{aligned}$$

$$= \frac{(-1)^q}{\det(\mathbf{X}^T \mathbf{\Gamma}^{-1} \mathbf{X})} (1 - \mathbf{C}(\mathbf{0}) G_{11}).$$

Pritaikę matricų funkcijos $tr(\cdot)$ savybes lygtyje (3.8) įrodome **b**).

$$\Psi_A(\hat{\beta}_{AMKM}) = tr(\mathbf{X}^T \mathbf{\Sigma}^{-1} \mathbf{X})^{-1} = -tr(\mathbf{X}^T \mathbf{\Gamma}^{-1} \mathbf{X})^{-1} + \mathbf{C}(\mathbf{0}).$$

Pastaba 1. Tarkime, kad $\xi_n = \{\mathbf{s}_1, \dots, \mathbf{s}_n\}$, $\mathbf{s}_i \in D$, $i = 1, \dots, n$. Šią stebėjimo taškų aibę vadinsime erdvinį imčių planu EIP (*angl. Spatial Sampling Design*). Visų „leistinų“ EIP, kurių didumas lygus n , erdvę žymėsime $\wp_n = \{\xi_n\}$. Kadangi $\mathbf{C}(\mathbf{0})$ nepriklauso nuo erdvinio imties plano, tai A kriterijų galima užrašyti tokiu būdu:

$$\text{planas } \xi_n^A \text{ yra A optimalus, jeigu } \xi_n^A = \arg \max_{\{\xi_n \in \wp_n\}} tr(\mathbf{X}^T \mathbf{\Gamma}^{-1} \mathbf{X})^{-1}.$$

Pastaba 2. Pastovaus vidurkio atveju planas ξ^* bus D ir A optimalus leistinų EIP erdvėje $\wp_n$, jeigu $\xi^* = \arg \max_{\{\xi_n \in \wp_n\}} \mathbf{\Gamma}^{\cdot\cdot}$.

Išvada.

Pastovaus vidurkio atveju:

$$\Psi_D(\hat{\beta}_{AMKM}) = \Psi_A(\hat{\beta}_{AMKM}) = \frac{1}{(\mathbf{1}^T \mathbf{\Sigma}^{-1} \mathbf{1})} = -\frac{1}{(\mathbf{1}^T \mathbf{\Gamma}^{-1} \mathbf{1})} + \mathbf{C}(\mathbf{0}) = \mathbf{\Gamma}^{\cdot\cdot} + \mathbf{C}(\mathbf{0}).$$

3.2. Taškų išsidėstymo 2x2 gardelėje optimizavimas

3.1 skyrelyje pateikta nepastovaus vidurkio su laisvu nariu parametrų AMKM įvertinio išraiška per semivariogramas (2 lema). Taip pat 3 lemoje per semivariogramas išvestos D ir A kriterijų funkcijų išraiškos. Siekiant nustatyti optimalų taškų išsidėstymą taisyklingoje 2x2 gardelėje, kurios centras (0,0), atliekamas eksperimentas su 5 taškais. Eksperimento metu vieno taško pozicija su koordinatėmis (0,0) yra fiksuojama, o likusių taškų pozicijos – keičiamos.

R sistemoje buvo modeliuojamos dvi situacijos. Viena jų numato, kad semivariogramos yra izotropinės. Kitoje jų buvo atskirai nagrinėjami du pastovaus ir nepastovaus (pirmo laipsnio trendo) vidurkio modeliai.

Visų penkių taškų išsidėstymą vadinsime *erdvine konfigūracija*.

Nesunku paskaičiuoti, kad galimų erdvinų konfigūracijų skaičius yra $C_8^4 = 70$. Pagal taškų išsidėstymo simetriškumą galima išskirti 12 ekvivalenčių konfigūracijų grupių, kurios pavaizduotos 3.1 paveiksle.

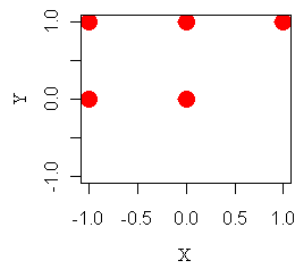
3.1 lentelėje pateikti atstumo tarp eksperimento taškų sumos, D ir A kriterijų funkcijų (pastovaus vidurkio ir pirmo laipsnio trendo modelių atvejais) rezultatai. Pagal 3 lemos išvadą pastovaus vidurkio atveju D ir A kriterijų funkcijų reikšmės sutampa.

Taip pat 3.1 lentelės rezultatai rodo, kad tiek pastovaus, tiek nepastovaus vidurkio atvejais D ir A kriterijų funkcijos minimumą įgyja toje pačioje 11 konfigūracijų grupėje. Sekančią pagal dydį minimalią reikšmę įgyja 9 konfigūracijų grupė. Tuo tarpu maksimumas pasiekiamas 1 konfigūracijų grupėje.

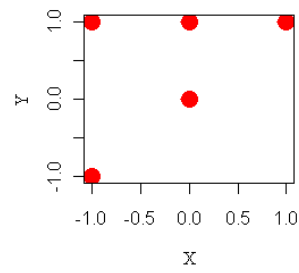
3.1 lentelė. Optimalaus taškų išsidėstymo konfigūracijos,
D ir A kriterijų funkcijų reikšmės

Table 3.1. The configurations of optimal distribution of point,
the values of D and A criterions

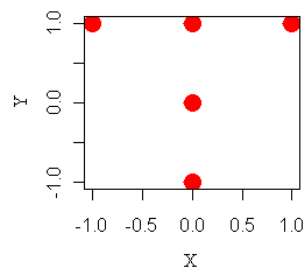
Konfigūracijos grupės numeris	Atstumo tarp taškų suma	Pastovaus vidurkio modelis	Nepastovaus vidurkio modelis (I laipsnio trendas)	
		$\Psi_D = \Psi_A$	Ψ_D	Ψ_A
1	26,95742	0,41638	0,10367	1,80112
2	32,61427	0,36777	0,04464	1,21014
3	30,60113	0,38133	0,05205	1,16669
4	31,08641	0,38896	0,08733	1,69061
5	31,90169	0,37013	0,04173	1,07123
6	29,42955	0,39204	0,06681	1,27600
7	30,60113	0,39026	0,06697	1,41091
8	32,25798	0,38164	0,07112	1,45575
9	35,08641	0,35037	0,03058	0,98145
10	31,08641	0,38422	0,06520	1,36842
11	38,62742	0,32301	0,01787	0,79346
12	27,31371	0,40329	0,07538	1,26796



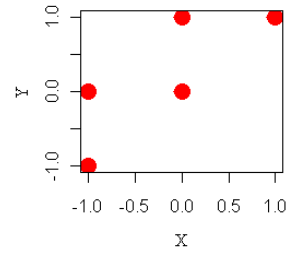
1)



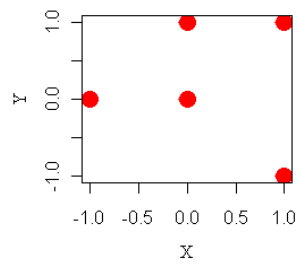
2)



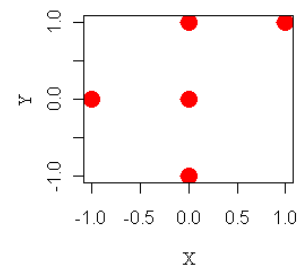
3)



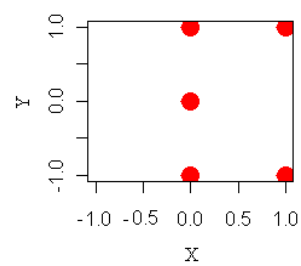
4)



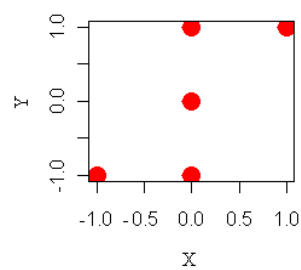
5)



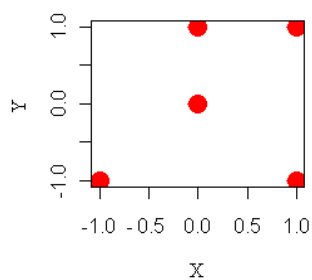
6)



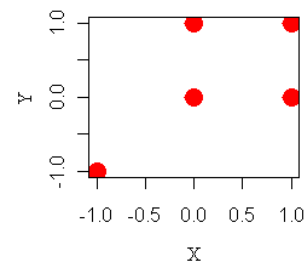
7)



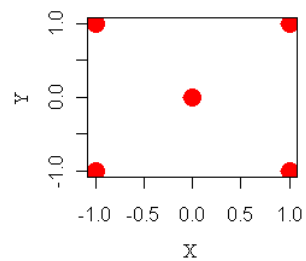
8)



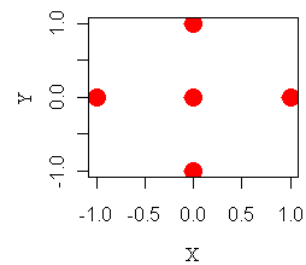
9)



10)



11)



12)

3.1 pav. Skirtingų konfigūracijų grupės
Fig. 3.1. The groups of different configurations

Galima daryti išvadą, kad A ir D kriterijų optimalumą garantuoja didžiausia atstumų suma tarp konfigūracijos taškų.

3.3. Empirinio semivariogramos įvertinio (MoM) asimptotinė analizė

Analizuojant erdvinis duomenis dažnai taikomi du asimptotikos tipai: didėjančios srities asimptotika (*angl. Increasing Domain Asymptotics*) bei pripildymo asimptotika (*angl. Infill Asymptotics*) arba kitaip dar vadinama fiksuotos srities asimptotika (*angl. Fixed-Domain Asymptotics*). Didėjančios srities asimptotika taikoma tuo atveju kai studijuojama sritis didėja, o ėminių vietos išlieka pastovios. Pripildymo asimptotika taikoma tuomet, kai ribojamoje srityje ėminių vietų kiekis yra didinamas (Cressie 1993, Smith 2001).

Chen, Simpson ir Ying (2000) nagrinėja grynuolio poveikį maksimalaus tikėtimumo metodu gautam įverčiui. Čia suderintini ir asimptotiniai įvertiniai yra gauti pripildymo asimptotikos atveju. Zhang, Zimmerman (2005) tiria abiejų asimptotikų teorinių ir empirinių rezultatų kokybę. Rezultatai rodo, kad tam tikrų nuosekliai apskaičiuotų eksponentinės kovariacinės funkcijos parametrų aproksimacija yra vienodai gera. Tačiau pripildymo asimptotika yra geresnė tiems parametrų, kurių neįmanoma įvertinti nuosekliai.

Bendri rezultatai apie asimptotines stacionarių semivariogramų modelių parametrų SMKМ įvertinių savybes aprašyti Lahiri, Lee, Cressie (2002) straipsnyje. Čia nėra pateiktos analitinės įvertinių asimptotinių kovariacijų išraiškos. Todėl tais rezultatais praktikai negali pasinaudoti (pavyzdžiui palyginant keleto erdvinį imčių planus pagal D kriterijų).

Reiktų pastebėti, kad nagrinėjant erdvinį imčių planus patogiau naudoti MoM įvertinį, išreikštą per erdvinio dizaino (Laplaso) matricą (1.17).

Šiame skyrelyje pateiksime jau žinomas asimptotines įvertinio MoM (1.15) savybes.

Semivariograma įvertinama baigtinėms $\mathbf{h} \in D$ reikšmėms $\mathbf{h}_1, \dots, \mathbf{h}_K$.

Originalūs rezultatai bus gauti atvejui, kai

$$\mathbf{Z}(\mathbf{s}) = \mathbf{f}(\mathbf{s})^T \boldsymbol{\beta} + \varepsilon(\mathbf{s}), \quad (3.12)$$

čia $\mathbf{f}(\cdot)$ – žinoma funkcija, $\boldsymbol{\beta}$ yra regresijos parametrų vektorius ir $\varepsilon(\cdot)$ – vidiniai stacionarus erdvinis procesas toks, kad:

$$E(\varepsilon(\mathbf{s})) = 0$$

ir

$$\frac{\text{var}(\varepsilon(\mathbf{s}) - \varepsilon(\mathbf{s} + \mathbf{h}))}{2} = 2\gamma(\mathbf{h}; \boldsymbol{\theta}), \quad \mathbf{s}, \mathbf{h} \in \mathfrak{R}^d.$$

Akivaizdu, kad $\gamma(\mathbf{h}; \boldsymbol{\theta})$ yra taip pat išmatuoto proceso $\mathbf{Z}(\cdot)$ semivariograma (formulė (1.3)).

Kadangi viena sudedamųjų stebimo proceso vidurkio dalių yra regresijos parametras $\boldsymbol{\beta}$, empirinis semivariogramos įvertinys paprastai apibrėžiamas identifikavus vidurkio funkciją. Tegu $\hat{\boldsymbol{\beta}}$ yra $\boldsymbol{\beta}$ įvertinys, o $\hat{\varepsilon}(\mathbf{s}_i) = \mathbf{Z}(\mathbf{s}_i) - \mathbf{f}(\mathbf{s}_i)^T \hat{\boldsymbol{\beta}}$, $i = 1, \dots, n$ vadinsime liekanomis.

Nagrinėsime semivariogramos Matheron'o įvertinį liekanoms:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2|N(\mathbf{h})|} \sum_{(\mathbf{s}_i, \mathbf{s}_j \in N(\mathbf{h}))} \left| \hat{\varepsilon}(\mathbf{s}_i) - \hat{\varepsilon}(\mathbf{s}_j) \right|^2. \quad (3.13)$$

Tarkime, kad:

- a) R_0 yra atviras $\left[-\frac{1}{2}, \frac{1}{2}\right]^d$ poaibis;
- b) $\{\lambda_n\}$ yra teigiamų skaičių seka tokia, kad $\lambda_n \rightarrow \infty$, kai $n \rightarrow \infty$;
- c) $\Delta = \text{diag}(\delta_1, \dots, \delta_d)$, čia $\delta_i > 0$, $i = 1, \dots, d$;
- d) $R_n = \lambda_n R_0$ – imčių sudarymo sritis;
- e) $Z_\Delta^d = \{\Delta \mathbf{i} : \mathbf{i} \in \mathbf{Z}_\Delta^d\} = \{(\delta_1 i_1, \dots, \delta_d i_d)^T : \mathbf{i} \in \mathbf{Z}_\Delta^d\}$.

Erdvinis imčių planas apibrėžiamas tokiu būdu (Lahiri 2002):

$$\xi_n = \{s_1, \dots, s_n\} = \{\mathbf{s} : \mathbf{s} \in Z_\Delta^d \cap R_n\}. \quad (3.14)$$

Šiame erdviniam imčių plane atstumas tarp stebinio taškų $\mathbf{s}_1, \dots, \mathbf{s}_n$ yra ne mažesnis nei $\delta_0 \equiv \min\{\delta_1, \dots, \delta_d\}$.

Akivaizdu, kad šis erdvinis imčių planas priklauso didėjančios erdvės asimptotikos tipui. Jam naudosime trumpinį LIDA.

Tarkime stebime erdvinį procesą $\{\varepsilon(\mathbf{s}) : \mathbf{s} \in D \subset \mathfrak{R}^d\}$, kuris generuoja σ – algebrą $\mathfrak{F}_\varepsilon(D) \equiv \sigma < \varepsilon(\mathbf{s}), \mathbf{s} \in D > .$

Aibės D poaibiams D_1 ir D_2 apibrėšime atstumą:

$$d(D_1, D_2) = \inf \{\|u - \mathbf{s}\| : u \in D_1, \mathbf{s} \in D_2\}.$$

1.1 apibrėžimas. $\mathfrak{Z}(D_1, D_2)$ stipraus sumaišymo (*angl. Strong-Mixing*) koeficientas $\alpha(a, b)$ procesui $\{\varepsilon(\mathbf{s}), \dots\}$ tokiu būdu:

$$\alpha(a, b) \equiv \sup \left\{ \tilde{\alpha}(D_1, D_2) : d(D_1, D_2) \geq a, \quad |D_1| \leq b, \quad |D_2| \leq b \right\}, \quad a > 0, \quad b > 0,$$

čia supremumas imamas pagal visus stačiakampius D_1, D_2 iš $\mathfrak{R}^d$.

Sąlyga 1. Visoms nykstančioms teigiamoms sekoms $\{t_n\}$, t. y. $t_n \rightarrow 0$, kai $n \rightarrow \infty$, $(\mathbf{i} + (0, 1]^d)t_n$ kubų formos laukų, kurie susikerta ir su R_0 ir su R_0^c skaičius yra $O(t_n^{d-1})$, kai $n \rightarrow \infty$.

Šioje sąlygoje dėl R_0 daroma prielaida dėl patologinių atvejų ir reiškia, kad ėminių, atliktų regiono R_n kraštuose, skaičius yra nereikšmingas, palyginus su bendru ėminių kiekiu.

Sąlyga 2. $E(\varepsilon(\mathbf{0}))^{4+\delta} < \infty$ ir su visais $a \geq 1$, $b \geq 1$, $\alpha(a, b) \leq Ca^{-\tau_1}b^{\tau_2}$, $0 < \delta \leq 4$, $C > 0$, $\tau_1 > (4 + \delta)d/\delta$ ir $0 \leq \tau_2 < \tau_1/d$.

Tai sumaišymo sąlyga atsitiktinei funkcijai $\{Z(\cdot)\}$, naudojama nustatyti įvertinio, gauto momentų metodu (3.14), asimptotinį normališkumą. Kadangi (3.13) semivariogramos įvertinys apibrėžiamas kaip kintamųjų $(\varepsilon(\mathbf{s}_i) - \varepsilon(\mathbf{s}_j))^2$ skirtumo kvadratas, vadinasi egzistuoja $(2+c)$ -tasis kintamųjų $(\varepsilon(\mathbf{s}_i) - \varepsilon(\mathbf{s}_j))^2$ momentas prie tam tikro $c > 0$.

Sąlyga 3. $\sup \left\{ \|\mathbf{f}(\mathbf{s}) - \mathbf{f}(\mathbf{s} + \mathbf{h})\| : \mathbf{s} \in \mathfrak{R}^d \right\} \leq F(\mathbf{h}) < \infty$ su visais $\mathbf{h} \in \mathfrak{R}^d$.

3 sąlyga susijusi su funkcijos $\mathbf{f}(\mathbf{s})$ augimo tempu ir yra tenkinama, jei pvz., $\mathbf{f}(\mathbf{s})$ yra ribota.

Sąlyga 4. $\lambda_n^d \left\| \hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta} \right\|^4 = o_p(1)$.

Ši sąlyga reikalauja bendro $\hat{\boldsymbol{\beta}}$ artėjimo prie regresijos parametrų vektoriaus $\boldsymbol{\beta}$ ne greičiau kaip $\lambda_n^{-d/4}$. Iš tiesų, 4 sąlygą teigia, kad, jei $\hat{\boldsymbol{\beta}}$ įvertinys yra gautas AMKM metodu:

$$\hat{\boldsymbol{\beta}}_{AMKM}(\boldsymbol{\theta}) = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Z}_n$$

remiantis 1 sąlyga ir 2 (sumaišymo) sąlyga galima parodyti, kad:

$$E_{\theta} \left\| \hat{\boldsymbol{\beta}}_{PMKM} - \boldsymbol{\beta} \right\|^2 = O(\lambda_n^{-d}),$$

kai $n \rightarrow \infty$ bet kuriai asimptotikai.

Panagrinėsime 3 sąlygą. Pirmos eilės trendo paviršiaus vidurkio modelio atveju $f(\mathbf{s}) = (1, s_x, s_y)^T$, o $f(\mathbf{s} + \mathbf{h}) = (1, s_x + h_x, s_y + h_y)^T$.

Tuomet:

$$\left\{ \sup \|f(\mathbf{s} + \mathbf{h}) - f(\mathbf{s})\| : \mathbf{s} \in R^2 \right\} = \sqrt{h_x^2 + h_y^2} = F(\mathbf{h}). \quad (3.15)$$

Tai reiškia, kad tenkinama 3 sąlyga, kai $\mathbf{h} = \mathbf{h}_1, \dots, \mathbf{h}_k$.

Naudosimės 3.3 teorema iš (Lahiri *et al.* 2002):

1 teorema. Tarkime, kad galioja 1 – 4 sąlygos. Tada su visais $\mathbf{h}_1, \dots, \mathbf{h}_K \in Z_{\Delta}^d$

$$n^{1/2} \left(2\hat{\gamma}(\mathbf{h}_1) - 2\gamma(\mathbf{h}_1; \boldsymbol{\theta}), \dots, 2\hat{\gamma}(\mathbf{h}_K) - 2\gamma(\mathbf{h}_K; \boldsymbol{\theta}) \right)^T \xrightarrow{d} N(\mathbf{0}, \sum_{2\gamma}(\boldsymbol{\theta})),$$

čia (i, j) -asis matricos $\sum_{\gamma}(\boldsymbol{\theta})$ elementas yra:

$$\left(\sum_{2\gamma}(\boldsymbol{\theta}) \right)_{ij} = \sum_{i \in Z_{\Delta}^d} \text{cov}_{\theta} \left\{ (\varepsilon(\mathbf{0}) - \varepsilon(\mathbf{h}_i))^2, (\varepsilon(\Delta \mathbf{i}) - \varepsilon(\Delta \mathbf{i} + \mathbf{h}_j))^2 \right\}. \quad (3.16)$$

kai $1 \leq i, j \leq K$.

Aprašytam erdvių imčių planui LIDA nagrinėsime atvejį, kai erdvinis procesas ε yra stacionarus Gauso erdvinis procesas.

Sąlyga 5. Spektrinis stacionaraus erdvinio proceso $Z(\mathbf{s})$ tankis $g(\mathbf{w})$, apibrėžtas formulėje (1.4), yra analitinė funkcija.

Aprašytam atvejui įrodysime asimptotinį (1.15) semivariogramos įvertinio normališkumą.

4 lema. Tarkime, kad $\mathbf{Z}_n \sim N(\mathbf{X}\boldsymbol{\beta}, \Sigma)$ galioja 1, 3, 4 ir 5 sąlygos. Tada su visais $\mathbf{h}_1, \dots, \mathbf{h}_K \in Z_{\Delta}^2$,

$$n^{1/2} \left(\hat{\gamma}(\mathbf{h}_1) - \gamma(\mathbf{h}_1; \boldsymbol{\theta}), \dots, \hat{\gamma}(\mathbf{h}_K) - \gamma(\mathbf{h}_K; \boldsymbol{\theta}) \right)^T \xrightarrow{d} N_K(\mathbf{0}, \sum_{\gamma}(\boldsymbol{\theta})), \quad (3.17)$$

čia (i, j) -asis asimptotinės kovariacijų matricos $\sum_{\gamma}(\boldsymbol{\theta})$ elementas yra:

$$\begin{aligned} & (\text{cov}(\varepsilon(\mathbf{0}), \varepsilon(\Delta \mathbf{i})) - \text{cov}(\varepsilon(\mathbf{0}), \varepsilon(\Delta \mathbf{i} + \mathbf{h}_i)) - \\ & - \text{cov}(\varepsilon(\mathbf{h}_i), \varepsilon(\Delta \mathbf{i} + \mathbf{h}_j)) + \text{cov}(\varepsilon(\mathbf{h}_i), \varepsilon(\Delta \mathbf{i} + \mathbf{h}_j)))^2 / 2. \end{aligned} \quad (3.18)$$

Irodymas. Lemoje nagrinėjamas stacionarus erdvinis procesas yra Gausinis ir tenkina 5 sąlygą. To pakanka tam, kad galiotų 2 sąlyga (Lahiri 2004). Tuo pačiu galioja ir 1 teoremos (Lahiri 2004) tvirtinimas. Pritaikę Gauso skirstinio savybes gauname:

$$\begin{aligned} & \text{cov}((\varepsilon(\mathbf{0}) - \varepsilon(\mathbf{h}_i))^2 (\varepsilon(\Delta \mathbf{i}) - \varepsilon(\Delta \mathbf{i} + \mathbf{h}_j))^2) = \\ & = 2(\text{cov}(\varepsilon(\mathbf{0}), \varepsilon(\Delta \mathbf{i})) - \text{cov}(\varepsilon(\mathbf{0}), \varepsilon(\Delta \mathbf{i} + \mathbf{h}_i)) - \\ & - \text{cov}(\varepsilon(\mathbf{h}_i), \varepsilon(\Delta \mathbf{i} + \mathbf{h}_j)) + \text{cov}(\varepsilon(\mathbf{h}_i), \varepsilon(\Delta \mathbf{i} + \mathbf{h}_j)))^2. \end{aligned} \quad (3.19)$$

Išraišką (3.19) įstatę į (3.16) gauname (3.17), (3.18).

Irodymas baigtas.

Pastaba 1. Tarkime, kad Gauso skirstinio erdvinio proceso $\{\varepsilon(\mathbf{s}) : \mathbf{s} \in D \in R^2\}$ erdvinis pokytis apibrėžiamas izotropine eksponentine semivariograma:

$$\gamma(\mathbf{h}, \boldsymbol{\theta}) = \gamma_0(|\mathbf{h}|, \boldsymbol{\theta}) = \theta_0 + \theta_1 \left(1 - e^{-\frac{|\mathbf{h}|}{\theta_2}} \right), \quad (3.20)$$

o dispersijų-kovariacijų matricos $\sum_{\gamma}(\boldsymbol{\theta})$ elementas (i, j) apibrėžiamas tokiu būdu:

$$\left(\sum_{\gamma}(\boldsymbol{\theta}) \right)_{ij} = \left(\frac{\theta_1^2}{2} \right) \sum_{i \in Z_{\Lambda}^2} (e^{-|\Delta \mathbf{i}|/\theta_2} - e^{-|\Delta \mathbf{i} + \mathbf{h}_j|/\theta_2} - e^{-|\Delta \mathbf{i} - \mathbf{h}_i|/\theta_2} + e^{-|\Delta \mathbf{i} + \mathbf{h}_j - \mathbf{h}_i|/\theta_2})^2. \quad (3.21)$$

Irodymas.

Šis erdvinis procesas turi spektrinį tankį:

$$g(w) = (\theta_1, \theta_2^2 / 2\pi) / (1 + w^2 \theta_2^2)^{3/2}, \quad w \geq 0. \quad (3.22)$$

Akivaizdu, kad (3.22) yra analitinė funkcija taško $w = w_0$ aplinkoje, kai $0 < w_0 < +\infty$.

Formulę (3.21) gauname panaudoję (3.20) ir (3.18) išraiškas.

3.4. Trečiojo skyriaus apibendrinimas ir išvados

1. Išvesta β parametrų AMKM įvertinio išraiška per semivariogramas, kai semivariogramos parametrų vektorius θ yra žinomas. Nesunku pastebėti, kad gauta galutinėje išraiškoje dispersijų-kovariacijų matrica atitinkamai pakeista semivariogramos matrica.
2. Nepastovaus vidurkio su laisvu nariu atvejui išvestos D ir A kriterijų AMKM įvertinio išraiškos per semivariogramas. Pastebėta, kad pastovaus vidurkio atveju šios išraiškos yra lygios.
3. MoM semivariogramos įvertinio asimptotinės dispersijų-kovariacijų matricos išraiška išvesta ekponentinės izotropinės semivariogramos, kuri vėliau naudojama praktiniuose skaičiavimuose, atvejui.
4. Siekiant nustatyti optimalų taškų išsidėstymą taisyklingoje 2x2 gardelėje išsiaiškinta, kad kaip ir buvo teigiama 3 lemoje pastovaus vidurkio atveju D ir A kriterijų funkcijų reikšmės sutampa, o D ir A kriterijų optimalumą garantuoja didžiausia atstumų suma tarp konfigūracijos taškų.

Siūlomos erdvinių duomenų modeliavimo ir prognozavimo metodikos taikymas

Šiame skyriuje pasiūlyta erdvinių duomenų modeliavimo per semivariogramas ir prognozavimo ordinariu, universaliu, medianinio išlyginimo krigingu ir ordinariu kokrigingu metodika taikoma Kuršių marių duomenų masyvams.

Skyriaus tematika paskelbti šeši autorės straipsniai (Krūminienė (Borisenko), Dučinskas, Garška 2004; Krūminienė (Borisenko), Garška 2004; Krūminienė (Borisenko), Dučinskas 2004; Krūminienė (Borisenko), Dučinskas 2004; Krūminienė (Borisenko) 2004; Krūminienė (Borisenko) 2006).

Aukščiau aprašytiems geostatistinių duomenų modeliavimo etapams atlikti buvo naudojami Kuršių marių duomenys. Kaip teigiama Jokšas K., Galkus A., Stakėnienė R. (1998) straipsnyje „Kuršių marios – sekli, beveik 93 kilometrų ilgio, gėlavandė lagūna atskirta nuo Baltijos jūros siaura smėlėta Kuršių nerija. Į pietus marios driekiasi iki Sambijos pusiasalio, o šiaurėje siauru Klaipėdos sąsiauriu jungiasi su centrine Baltijos jūros dalimi. Kuršių marių vidutinis gylis tesiekia 3,8 m, o 3 m. izobata juosia daugiau negu du trečdalius jų ploto. Maksimalus marių gylis yra 5,8 m. ir tik Klaipėdos sąsiauryje jis dirbtinai yra

padidintas iki 14 m. Didžiausi gyliai, siekiantys 4,5 metrus, yra centrinėje ir pietinėje marių dalyse.“

„Upėmis atnešama, pačiose mariose produkuojama bei iš atmosferos patenkanti smulkiadispersinė nuosėdinė medžiaga sorbuoja apie 70,98% įvairių cheminių ingredientų. Patyrusi sedimentacinius–geocheminius virsmus ji nusėda marių dugne, suformuodama vienokios ar kitokios granulinės, medžiaginės ir cheminės sudėties nuosėdas. Dalis šios medžiagos per Klaipėdos sąsiaurį iškeliauja į atvirą jūrą (Jokšas, Galkus, Stakėnienė 1998).

Disertaciniame darbe naudotas Kuršių marių duomenis sudaro du masyvai.

- *I masyvas* – Kuršių marių gylio ir dugno nuosėdų (granulimetrinės sudėties frakcijų) duomenys (autorius S. Gulbinskas, 1990 metai). Gylio duomenys rinkti 263 lokacijose, dugno nuosėdų – 213 lokacijose. Dugno nuosėdų mėginiai imami gruntosemiu. Dugno granulimetrinė sudėtis nustatoma nuosėdas sijojant per sietus. Priklausomai nuo nuosėdų diametro (mm), nuosėdos suskirstytos į 7 kategorijas (frakcijas): (1) > 0,5 mm; (2) 0,5 – 0,25 mm; (3) 0,25 – 0,125 mm; (4) 0,125 – 0,063 mm; (5) 0,063 – 0,01 mm; (6) 0,01 – 0,004 mm; (7) < 0,004 mm. Dugno nuosėdų reikšmės pateiktos procentais. Bendra visų 7 kategorijų procentų sumą konkrečioje lokacijoje sudaro 100%.
- *II masyvas* – Kuršių marių gylio duomenys 2990 lokacijose (duomenis pateikė Baltijos pajūrio tyrimų ir aplinkos planavimų institutas, matavimai atlikti 2005 metais).

Praktikoje taikant geostatistinius metodus susiduriama su programinių priemonių, sukurtų būtent geostatistinių metodų taikymui, trūkumu.

Vienuose programiniuose paketuose geostatistiniai įrankiai yra integruoti, kai kurie paketai yra tikslingai sukurti geostatistikoje naudoti. Tačiau ir tarp jų yra tik keletas nemokamų ir laisvai prieinamų geostatistinių programų.

Šiame darbe geostatistinių metodų taikymui Kuršių marių duomenų masyvams naudotos nemokamos programos *Gstat* bei *R*.

Gstat yra atviro kodo programa sukurta 1997 metais (autorius E. Pebesma) bei skirta geostatistiniam modeliavimui bei prognozei. Pradinė *Gstat* programos versija buvo autonomiška, turinti sąsają su kai kuriomis GIS programomis. Nuo 2003 metų *Gstat* funkcijos tapo prieinamos *R* sistemoje arba kaip *S-Plus* biblioteka. *Gstat* funkcijų integracija į *R* sistemą, sudarė galimybę vienoje programoje duomenims taikyti ne tik geostatistinius, bet ir įvairius kitus statistinius metodus.

R sistema yra sudaryta iš įvairios paskirties paketų. Geostatistinių duomenų analizei *R* sistemoje skirti šie paketai: *akima*, *tripack*, *sgeostat*, *splancs*, *spatstat*, *stats*, *stats4*, *geor*, *gstat*, *sp*, *spatial* ir kiti. Kai kurie iš šių paketų yra įdiegiami kartu su *R* sistema (*stats*, *stats4*, *spatial*), kitus reikia įdiegti iš CRAN svetainės (*gstat*, *sp* ir kitus paketus). *R* sistemos *gstat* paketo funkcijos buvo

dažniausiai naudojamos šio disertacinio darbo duomenų modeliavimui ir prognozei įgyvendinti.

Kaip ir *Gstat* programoje, taip ir *gstat* pakete, integruotame į *R* sistemą, nagrinėjamiems duomenims galima sudaryti empirinę visakryptę, kryptines bei kryžmines semivariogramas, semivariogramos žemėlapi; empirinės semivariogramos taškams parinkti vieną iš kelių parametrinių semivariogramos modelių, paskaičiuoti parinkto modelio semivariogramos parametrų bei PKS reikšmes.

Taip pat *gstat* pakete duomenų prognozei atlikti yra galimybė naudoti (taškinį arba blokinį) paprastą, ordinarų, universalų krigingą ar kokrigingą. Universalus krigingo atveju galima nurodyti trendo paviršiaus laipsnį arba konkrečius paviršiaus funkcijos koeficientus. Kokrigingo atveju kiekvieno iš nagrinėjamų kintamųjų trendo paviršiaus funkcija gali skirtis. Norint palyginti kelių taikytų krigingų paklaidas yra galimybė naudoti kryžminę (*angl. Leave-One-Out*) patikrą.

4.1. 1990 metų gylio ir dugno nuosėdų duomenų modeliavimas ir prognozė

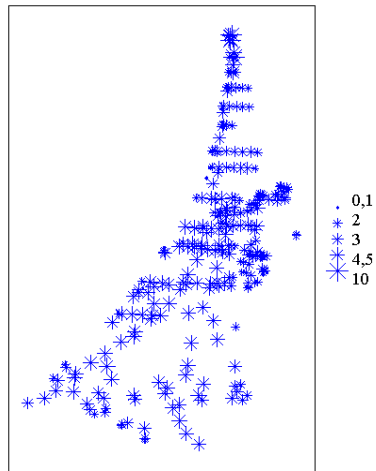
Pradinė duomenų analizė. Atlikus pradinę 1990 metų gylio duomenų analizę (4.1 lentelė) išsiaiškinta, kad *x* koordinatės duomenų intervalas apima 278199 ir 333376 metro intervalą, o *y* koordinatės: 6088178 – 6172784 metro intervalą. Mažiausia gylio duomenų reikšmė lygi 0,1 metro, o didžiausia gylio duomenų reikšmė – 10 metrų.

4.1 lentelė. Gylis duomenų pagrindinės statistikos
Table 4.1. Statistics of depth data

	X koordinatė (metrais)	Y koordinatė (metrais)	Gylis (metrais)
MIN	278199	6088178	0,100
Q1	310170	6115148	2,000
Mediana	318407	6129108	3,000
Vidurkis	314983	6128941	3,161
Q3	322332	6139799	4,500
MAX	333376	6172784	10,000

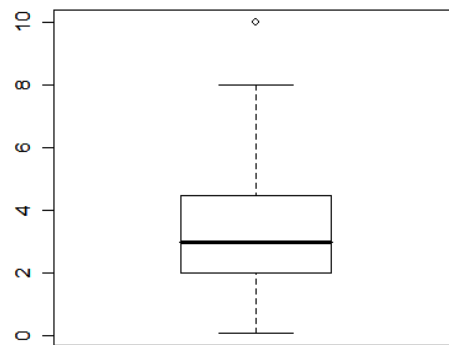
čia MIN – minimumas, Q1 – pirmasis kvartilis, Q3 – trečiasis kvartilis, MAX – maksimumas.

Gylio duomenų išsibarstymo Kuršių marių srityje vaizdas pateiktas 4.1 paveiksle.



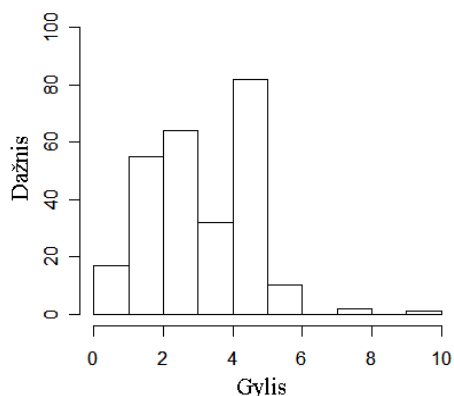
4.1 pav. 1990 metų gylio duomenų išsibartymas visoje Kuršių marių teritorijoje

Fig. 4.1. The distribution of depth data (1990 year) across Curonian Lagoon territory



4.2 pav. Gylio reikšmių (1990 metai) penkiaskaitė suvestinė

Fig. 4.2 Box-plot of depth data (1990 year)



4.3 pav. Gylio reikšmių (1990 metai) histograma

Fig. 4.3. Histogram of depth data (1990 year)

Gylio duomenų histograma (4.2 pav.) parodė, kad kaip ir dauguma ekologinių duomenų, Kuršių marių gyliai turi dešinės pusės asimetriją. Atsižvelgus į 2.2 skyriuje nurodytus duomenų transformacijos trūkumus, tolesniam gylio duomenų modeliavimui atlikti, taikoma prielaida, kad daugialypis duomenų skirstinys yra Gauso ir duomenys netransformuojami.

Dugno nuosėdų pagrindinės statistikos pateiktos 4.2 lentelėje.

Ordinaraus kriginio ir kokrigino taikymas. Siekiant atlikti Kuršių marių I masyvo duomenų modeliavimą ir prognozę apibrėžtos kelios prielaidos:

- nagrinėjamas erdvinis procesas yra vidiniai stacionarus (1.3 apibrėžimas, psl. 11) bei izotropinis (1.5 apibrėžimas, psl. 12),
- proceso vidurkis yra pastovus, bet nėra žinomas (ordinaraus kriginio atvejis, 2.4.1 skyrelis, psl. 50).

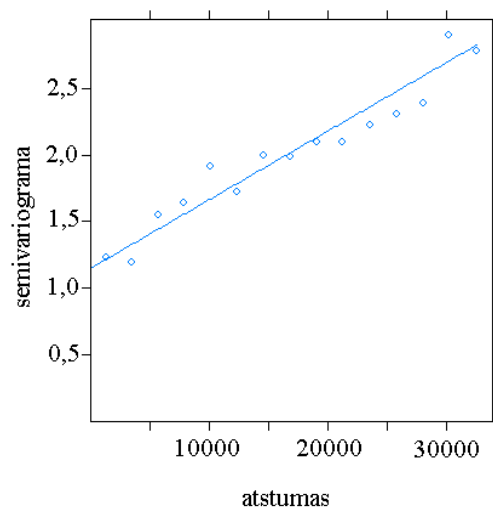
Kuršių marių gylio duomenų erdvinei sklaidai aprašyti naudoti tiesinis (4.4 pav.), sferinis (4.5 pav.), eksponentinis (4.6 pav.) ir *Gauso* (4.7 pav.) parametriniai semivariogramų modeliai (1.4 skyrelis, psl. 15–27). Naudotų parametrinių semivariogramų modelių parametrai ir PKS reikšmės pateiktos 4.3 lentelėje. Kadangi mažiausia PKS reikšmė gauta sferinio modelio atveju, tolesnei gylio duomenų prognozei ordinaraus kriginio atveju pasirinktas sferinis semivariogramos modelis. Ordinaraus kriginio prognozės ir prognozės dispersijų rezultatai žemėlapių pavidalu pavaizduoti 4.8 paveiksle ir 4.9 paveiksle.

4.2 lentelė. Pagrindinės I duomenų masyvo dugno nuosėdų statistikos
Table 4.2. Statistics of sediments of I data array

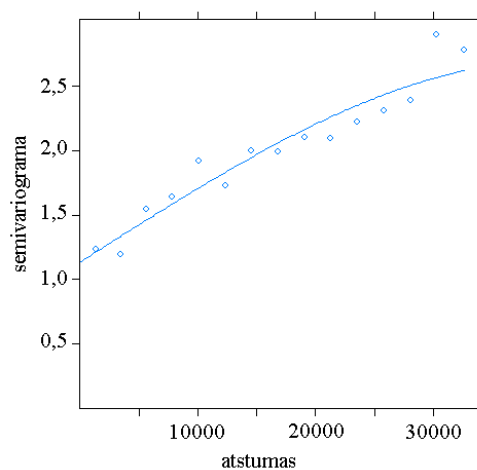
		MIN	Q1	Mediana	Vidurkis	Q3	MAX
Koordinatės	X m	278199	307983	318281	314517	322296	333376
	Y m	6088178	6115088	6128898	6129117	6140968	6172784
Dugno nuosėdų kategorijos	(1) 0,5	0,000	0,000	0,240	1,663	1,050	29,970
	(2) 0,5–0,25	0,00	0,00	4,10	12,98	15,38	92,18
	(3) 0,25–0,125	0,00	5,27	42,43	41,90	73,38	94,84
	(4) 0,126–0,063	0,08	2,46	8,28	15,2	24,08	73,74
	(5) 0,063–0,01	0,01	0,26	1,79	18,30	36,02	75,85
	(6) 0,01–0,004	0,000	0,000	0,000	4,694	7,500	34,640
	(7) <0,004	0,000	0,000	0,000	5,213	8,490	53,370

4.3 lentelė. Gylio reikšmių parametrinių semivariogramų modelių parametrai, PKS reikšmės
Table 4.3. SSE values and parameters of parametric semivariogram models of depth data

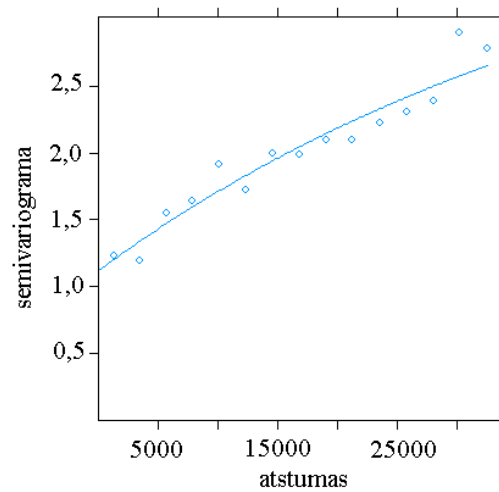
Parametrinis semivariogramos modelis	θ_0	$\theta_0 + \theta_1$	θ_2	PKS
Tiesinis	1,149956	2,84608	32988,09	0,0000039
Sferinis	1,131616	2,69877	40228,95	0,0000036
Eksponentinis	1,119583	4,12937	45759,61	0,0000039
Gauso	1,201574	2,25094	11281,43	0,0000038



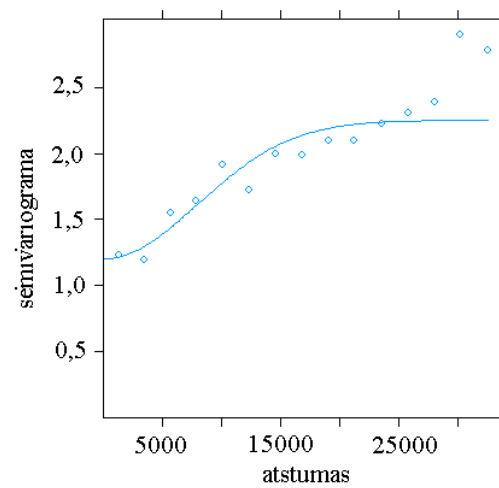
4.4 pav. Gylio reikšmių tiesinė semivariograma
Fig. 4.4. The linear semivariogram of depth data



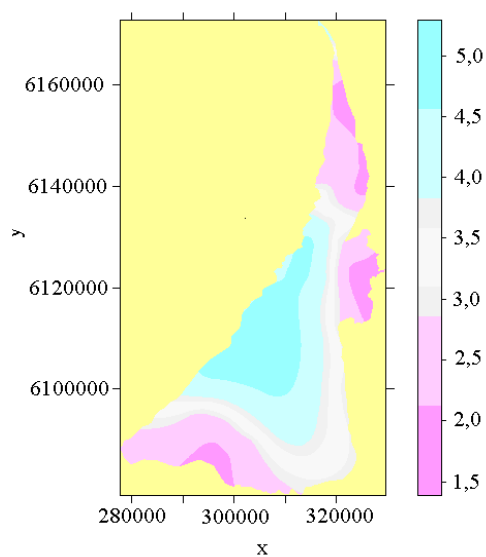
4.5 pav. Gylio reikšmių sferinė semivariograma
Fig. 4.5. The spherical semivariogram of depth data



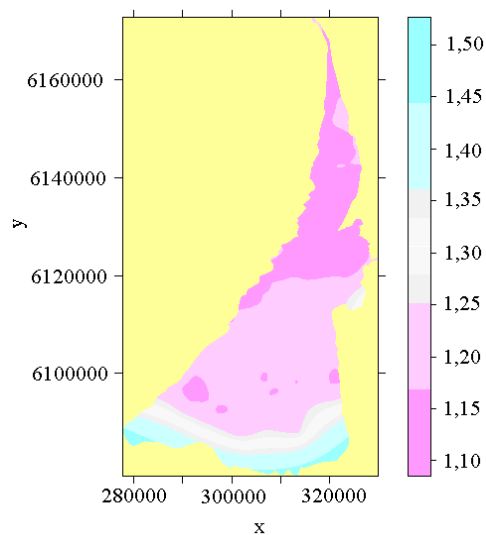
4.6 pav. Gylio reikšmių eksponentinė semivariograma
Fig. 4.6. The exponential semivariogram of depth data



4.7 pav. Gylio reikšmių *Gauso* semivariograma
Fig. 4.7. The *Gaussian* semivariogram of depth data



4.8 pav. Gylio prognozės rezultatai (ordinarus krigingas)
Fig. 4.8. The prediction results of depth data (ordinary kriging)



4.9 pav. Gylio dispersijos (ordinarus krigingas)
Fig. 4.9. Variance of depth data (ordinary kriging)

Duomenų prognozei atlikti buvo suformuotas atskiras visą Kuršių marių teritoriją apibrėžiantis masyvas (kitai dar vadinamas „kauke“ (*angl. Mask*)) pavadinimu *marios.txt*. Duomenų masyvą sudaro 3 kintamieji: x koordinatė (minimali – 277807, maksimali – 329707), y koordinatė (minimali – 6079080, maksimali – 6172680) bei z (tariamo gylio) kintamasis, sudarytas iš vienetinių reiškinių. Atliekant prognozę krigingu, vienetukai pakeičiami atitinkamomis prognozės ar prognozės dispersijų reikšmėmis. Šis masyvas–„kaukė“ darbe naudotas įvairių prognozių procedūrose, kas leidžia palyginti kelių prognozės procedūrų rezultatus.

Yra žinoma, kad dugno nuosėdų išsibirstymas dugne priklauso nuo dugno reljefo, t. y. nuo gylio. Šia priklausomybe pasinaudota kiekvienos dugno nuosėdų kategorijos duomenis prognozuojant kokrigingu (2.4.3 skyrelis).

Pirma visų dugno nuosėdų kategorijų erdvinei sklaidai aprašyti sudarytos empirinės semivariogramos; joms priglodintas tiesinis semivariogramos modelis. Vėliau sudarytos atitinkamų dugno nuosėdų kategorijų ir gylio reikšmių kryžminės semivariogramos. Kadangi pagal aukščiau apibrėžtas prielaidas (psl. 87) erdvinio proceso vidurkis nėra žinomas, tuomet ir gylio, ir dugno nuosėdų kategorijų duomenims prognozuoti visoje Kuršių marių teritorijoje taikytas ordinarus kokrigingas. Rezultatai aprašyti Krūminienė (Borisenko); Dučinskas, Garška (2003) straipsnyje.

Vėliau dugno nuosėdų kategorijų sklaidai aprašyti naudoti sferinis ir eksponentinis semivariogramų modeliai. Kiekvienos dugno nuosėdų kategorijos sferinio ir eksponentinio semivariogramų modelių parametrai ir PKS reikšmės pateiktos 4.4 lentelėje.

4.4 lentelė. Dugno nuosėdų kategorijų sferinio ir eksponentinio semivariogramų modelių parametrai ir PKS reikšmės
Table 4.4. SSE values and parameters of spherical and exponential semivariogram models of sediments in different category

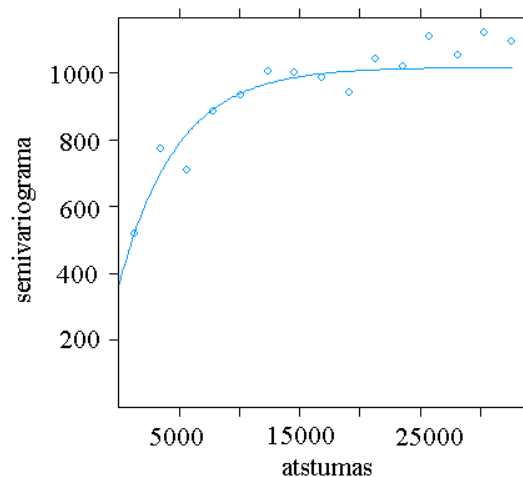
Dugno nuosėdų kategorija	Parametrinis semivariogramos modelis	θ_0	$\theta_0 + \theta_1$	θ_2	PKS
(1)	Sferinis	16,76314	38,54266	231105,40	0,006369
	Eksponentinis	14,09064	131,7139	575064,20	0,011744
(2)	Sferinis	310,8018	553,2946	51950,45	0,361729
	Eksponentinis	311,0166	1442,365	161229,90	0,362127
(3)	Sferinis	441,6102	996,6513	12300,99	0,690219
	Eksponentinis	361,8325	1016,539	4763,74	0,510723

4.4 lentelės tęsinys.

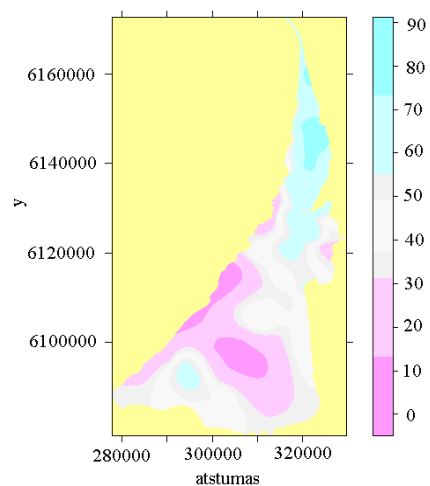
Dugno nuosėdų kategorija	Parametrinis semivariogramos modelis	θ_0	$\theta_0 + \theta_1$	θ_2	PKS
(4)	Sferinis	265,9736	303,8986	7843,18	0,012749
	EkspONENTINIS	254,3599	303,7208	2433,61	0,013228
(5)	Sferinis	208,8978	575,9207	17232,86	0,035182
	EkspONENTINIS	186,6983	626,149	8919,54	0,011944
(6)	Sferinis	15,6694	709,8204	575065,00	0,001723
	EkspONENTINIS	16,1060	3161,659	1717138,00	0,001815
(7)	Sferinis	14,2324	77,072	20551,03	0,001554
	EkspONENTINIS	12,3298	95,0431	13630,17	0,001152

Gautos PKS reikšmės rodo, kad (1), (2), (4) ir (6) dugno nuosėdų kategorijų sklaidą geriausiai aprašo sferinis semivariogramos modelis, kai tuo tarpu (3), (5) ir (7) dugno nuosėdų kategorijų sklaidą – eksponentinis semivariogramos modelis.

Ordinarus kriginis ir ordinarus kokriginis procedūrų rezultatams palyginti pasirinkta (3) dugno nuosėdų kategorija, kur dugno nuosėdų diametras patenka į intervalą 0,25 – 0,125 mm.

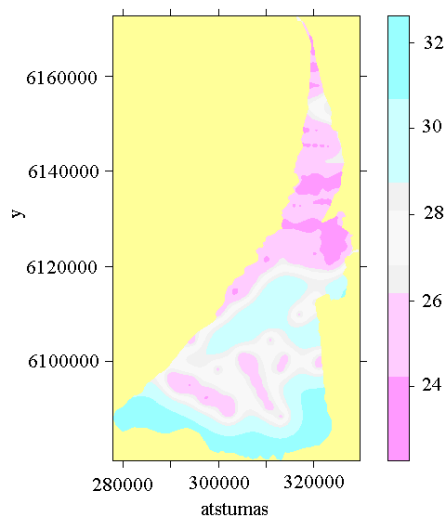


4.10 pav. (3) dugno nuosėdų kategorijos eksponentinė semivariograma
Fig. 4.10. The exponential semivariogram of (3) category of sediments



4.11 pav. (3) dugno nuosėdų kategorijos prognozės rezultatai
(ordinarus krigingas)

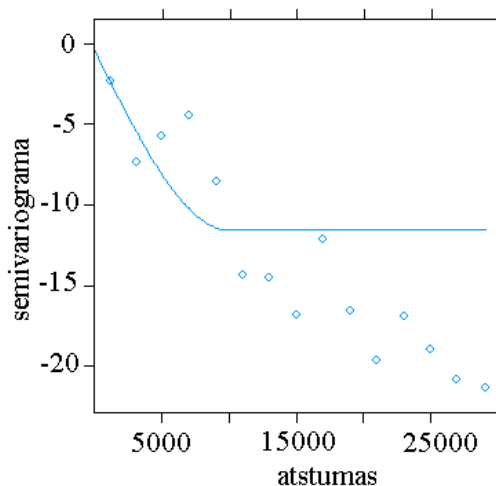
Fig. 4.11. The prediction results of (3) category of sediments
(ordinary kriging)



4.12 pav. (3) dugno nuosėdų kategorijos dispersijos
(ordinarus krigingas)

Fig. 4.12. Variance of (3) category of sediments
(ordinary kriging)

Pirma (3) dugno nuosėdų kategorijos duomenų erdvinei variacijai aprašyti buvo naudotas eksponentinis semivariogramos modelis, kuris pavaizduotas 4.10 paveiksle, o duomenų prognozei taikytas ordinarus krigingas (prognozės ir dispersijų žemėlapiui pateikti atitinkamai 4.11 ir 4.12 paveiksluose).



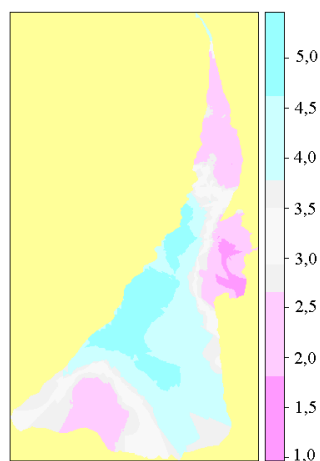
4.13 pav. (3) dugno nuosėdų kategorijos/gylio reikšmių sferinė kryžminė semivariograma

Fig. 4.11. The spherical cross-semivariogram of (3) category of sediments/depth data

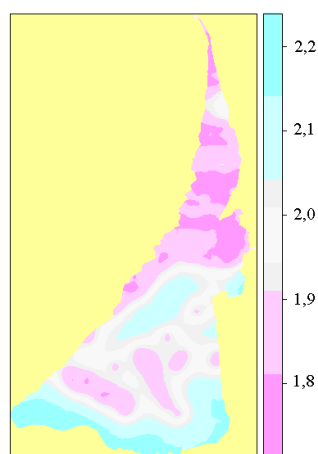
4.5 lentelė. Kryžminės patikros statistikos ordinario krigingo ir kokrigingo atvejais

Table 4.5. The statistics of cross-validation method in the case of ordinary kriging and cokriging

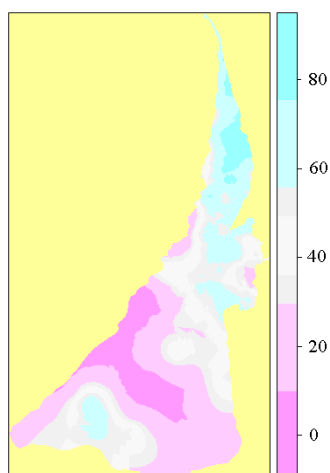
Krigingo tipas	Duomenys	VPK	ŠVKPK	VKNS
Ordinarus krigingas	gylis	-0,0005201	1,297655	0,972133
Kokrigingas		-0,0224748	1,562771	0,864518
Ordinarus krigingas	(3) dugno nuosėdų kategorija	-0,2401935	575,6072	0,906008
Kokrigingas		-0,2254713	561,8491	0,901150



4.14 pav. Gylio prognozės rezultatai (ordinarus kokrigingas)
Fig. 4.14. The prediction results of depth data (ordinary cokriging)

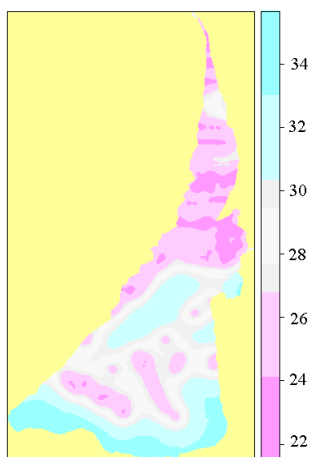


4.15 pav. Gylio dispersijos (ordinarus kokrigingas)
Fig. 4.15. Variance of depth data (ordinary cokriging)



4.16 pav. (3) dugno nuosėdų kategorijos prognozės rezultatai
(ordinarus kokrigingas)

Fig. 4.16. The prediction results of (3) category of sediments
(ordinary cokriging)



4.17 pav. (3) dugno nuosėdų kategorijos dispersijos
(ordinarus kokrigingas)

Fig. 4.17. Variance of (3) category of sediments
(ordinary cokriging)

Kadangi gylio reikšmių (4.3 lentelė) ir daugelio dugno nuosėdų kategorijų sklaidai aprašyti tinkamiausias sferinis semivariogramos modelis, (3) dugno nuosėdų kategorijos /gylio reikšmių kryžminei semivariogramai priglodintas sferinis semivariogramos modelis (4.13 pav.). Vėliau ordinaraus kokrigingo procedūros dėka atlikta abiejų kintamųjų prognozė (gylio reikšmių prognozės ir dispersijų žemėlapiai pateikti 4.14 ir 4.15 paveikslėliuose, o (3) dugno nuosėdų kategorijos reikšmių prognozės ir dispersijų žemėlapiai atitinkamai pavaizduoti 4.16 ir 4.17 paveiksluose).

Siekiant patikrinti ar ordinariu krigingu, ar ordinariu kokrigingu (3) dugno nuosėdų kategorijos bei gylio duomenys buvo prognozuoti tiksliau, taikytas kryžminės patikros metodas ir skaičiuotos 2.5 skyrelyje aprašytos statistikos. Statistikų rezultatai pateikti 4.5 lentelėje.

Išanalizavus 4.5 lentelę galima pastebėti, kad gylio prognozės rezultatai ordinaraus krigingo atveju lyginant su gautais rezultatais ordinaraus kokrigingo atveju yra tikslesni (vidutinė klaida artima nuliui, vidutinė kvadratinė prognozės klaida mažesnė, vidutinis kvadratinis nuokrypio santykis – artimas vienetui). Tuo tarpu vidutinė klaida bei vidutinė kvadratinė prognozės klaida rodo, kad (3) dugno nuosėdų kategorijos rezultatai tikslesni ordinaraus kokrigingo atveju.

Universalaus krigingo taikymas. Atliekant 1990 metų gylio duomenų semivariogramų analizę pastebėta, kad taškai monotoniškai kyla žemyn ir nėra aiški slenksčio riba. Tai trendo egzistavimo duomenyse požymis. Siekiant nustatyti, kurio laipsnio polinomo trendas geriausiai aprašo nagrinėjamą paviršių, buvo taikytas ANOVA metodas. Tam panaudotos *spatial* paketo funkcija *surf.ls* (mažiausių kvadratų metodu skaičiuojamas tam tikro laipsnio polinomo paviršius) ir *base* paketo funkcija *summary* (ANOVA lentelės sudarymas). Daliniai skaičiavimų rezultatai pateikti 4.6 lentelėje.

4.6 lentelė. Gylio duomenų trendo paviršiui nustatyti ANOVA rezultatai

Table 4.6. Results of ANOVA method to determine the trend type of depth data

Polinominio paviršiaus tipas		Tiesinis	Kvadratinis	Polinominis
Daugialypis R^2 :		0,1094	0,2109	0,3581
Priglodintų duomenų	Minimumas	2,44	1,188	0,114
	1Q	2,776	2,713	2,477
	Mediana	2,983	3,164	3,1062
	3Q	3,401	3,738	3,761
	Maksimumas	4,66	4,117	4,944

4.6 lentelės tęsinys.

Polinominio paviršiaus tipas		Tiesinis	Kvadratinis	Polinominis
Liekanų	Minimumas	-3,7772	-3,8545	-3,9994
	1Q	-0,9186	-0,8444	-0,7266
	Mediana	0,0419	0,2073	0,1211
	3Q	1,0554	0,9061	0,8106
	Maksimumas	5,2859	5,2662	4,0401

Lentelės rezultatai rodo, kad III laipsnio polinomo daugialypis R^2 yra didžiausias, tad galima daryti išvadą, kad būtent šios laipsnio polinomas nagrinėjamų duomenų paviršių aprašo geriausiai. III laipsnio polinomo paviršiaus trendo koeficientai:

β_1 :	β_2 :	β_3 :	β_4 :	β_5 :	β_6 :	β_7 :	β_8 :	β_9 :	β_{10} :
2,919	3,326	-3,840	-2,329	-6,866	8,102	0,526	-3,253	5,155	4,011

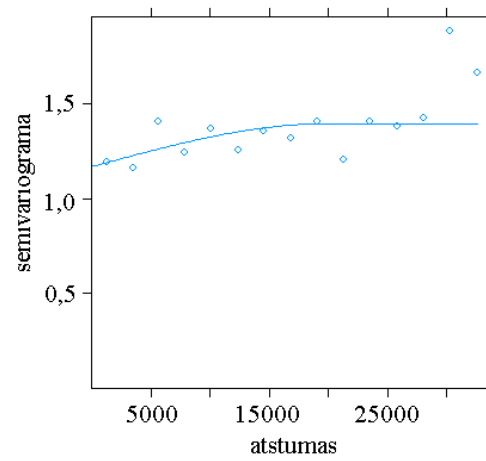
Apibrėžus III laipsnio polinomo paviršių, sudarant visakryptę semivariogramą taškų kryptis pasikeitė ir jie įgavo pastovumą. Visakryptei semivariogramai priglodinti naudoti trys parametriniai semivariogramų modeliai: sferinis, eksponentinis ir *Gauso*. Naudotų parametrinių semivariogramos modelių parametrų ir PKS reikšmės pateiktos 4.7 lentelėje.

4.7 lentelė. Semivariogramų, priglodintų gylio liekanoms, parametrinių semivariogramų modelių parametrai ir PKS reikšmės

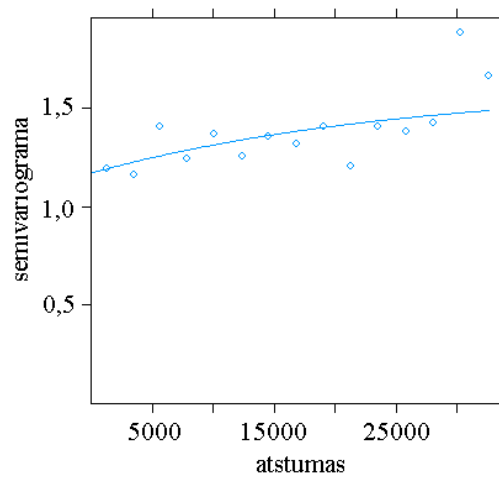
Table 4.7. SSE values and parameters of semivariogram models of depth residuals

Parametrinis semivariogramos modelis	θ_0	$\theta_0 + \theta_1$	θ_2	PKS
Sferinis	1,1677973	1,39555	19984,85	0,00000205
Eksponentinis	1,1692036	1,60942	25361,17	0,00000195
<i>Gauso</i>	1,1765091	1,36609	6113,77	0,00000197

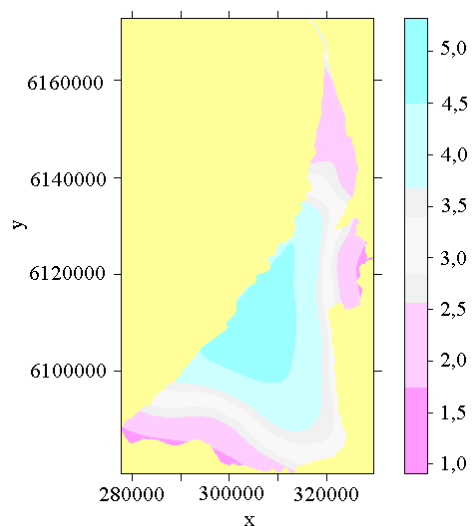
PKS reikšmė eksponentinio semivariogramos modelio atveju yra mažiausia, tačiau sferinis modelis buvo pasirinktas kaip tinkamiausias atvejui, kai gylio duomenų vidurkio pokyčiui aprašyti nėra taikomas III laipsnio polinomo trendas. Tad duomenų prognozei atlikti buvo parinkti ir sferinis, ir eksponentinis semivariogramų modeliai.



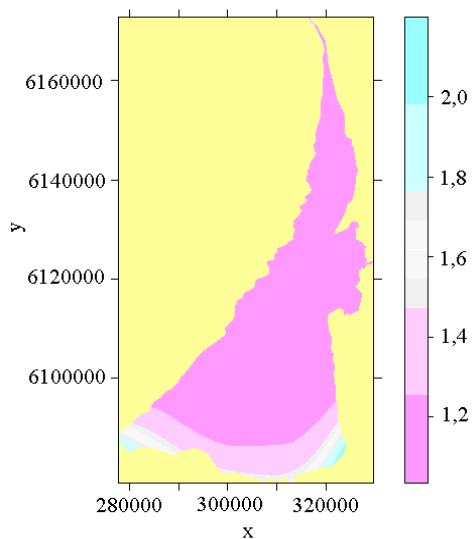
4.18 pav. Liekanų sferinis semivariogramos modelis
Fig. 4.18. The spherical semivariogram of residuals



4.19 pav. Liekanų eksponentinis semivariogramos modelis
Fig. 4.19. The exponential semivariogram of residuals



4.20 pav. Gylio prognozės žemėlapis (universalus krigingas)
Fig. 4.20. The prediction map of depth data (universal kriging)



4.21 pav. Gylio dispersijos (universalus krigingas)
Fig. 4.21. Variance of depth data (universal kriging)

4.8 lentelė. Kryžminės patikros statistikos universalaus krigingo atveju
Table 4.8. Statistics of cross-validation method in the case of universal kriging

Parametrinis semivariogramos modelis	VPK	ŠVKPK	VKNS
Sferinis	-0,02717368	2,678192	2,038265
Eksponentinis	-0,05203854	4,700589	2,861926

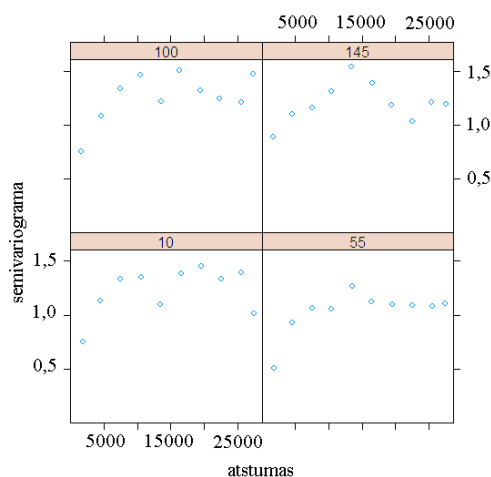
Kaip rodo kryžminės patikros statistikų rezultatai (4.8 lentelė), sferinis semivariogramos modelis III laipsnio polinominio paviršiaus atveju gylio duomenų sklaidą aprašo geriau. Prognozės ir prognozės dispersijos rezultatai pateikti atitinkamai 4.20 paveiksle ir 4.21 paveiksle.

Anizotropijos ir išskirties įtakos 1990 metų gylio prognozei analizė.

Apart trendo, viename iš tyrimų buvo siekiama nustatyti:

- ar duomenų prognozę įtakoja pradiniam duomenų analizės etape rasta išskirtis (10 metrų gylis);
- ar 1990 metų gylio duomenims būdinga anizotropija.

Tam buvo sudarytos kryptinės robastinės semivariogramos (žr. 2.2. skyrelis) keturiomis 4 pagrindinėmis kryptimis: 10° , 55° , 100° ir 145° laipsnių kryptimi kiekvienam vėlinimo vektoriui su $22,5^\circ$ tolerancijos kampu (4.22 pav.). Išanalizavus kryptines semivariogramas, nustatyta, kad duomenims būdinga geometrinė anizotropija. Visakryptėms semivariogramoms taikyti sferinis ir *Gauso* parametriniai semivariogramų modeliai (4.9 lentelė).



4.22 pav. 1990 metų gylio empirinės kryptinės semivariogramos

Fig. 4.22. The empirical directional semivariogram of depth data (1990 year)

4.9 lentelė. Kryptinių sferinio ir *Gauso* semivariogramų modelių parametrai
Table 4.9. The parameters of directional spherical and *Gaussian* semivariogram models

	Sferinis semivariogramos modelis			<i>Gauso</i> semivariogramos modelis		
Kryptis	θ_0	$\theta_0 + \theta_1$	θ_2	θ_0	$\theta_0 + \theta_1$	θ_2
10°	0,7082016	1,290837	8315,83	0,7948496	1,291901	4104,56
55°	0,1516071	1,117112	7440,11	0,2829290	1,126218	3300,22
100°	0,6233273	1,409745	11750,59	0,7420671	1,499498	6978,11
145°	0,6713363	1,395457	12868,58	0,7269854	1,318329	4642,59

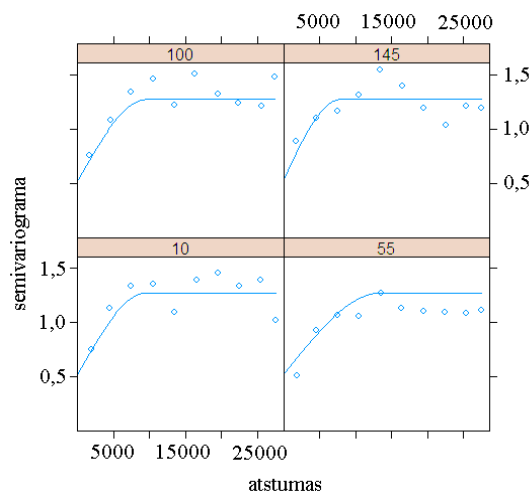
Didžiausias slenksčio atstumas sferinio modelio atveju aptiktas 145° krypties semivariogramoje, *Gauso* modelio atveju – 100° krypties semivariogramoje. Tuo tarpu mažiausias slenksčio atstumas abiejų semivariogramų modeliuose aptiktas 55° kryptimi. Atitinkamai pagal šias kryptis kiekvienam semivariogramos modeliui skaičiuotas anizotropijos santykis (4.10 lentelė) bei kryžminės patikros statistikos (4.11 lentelė) (žr. 2.4. skyrelyje).

4.10 lentelė. Anizotropijos santykis sferinio ir *Gauso* semivariogramų modelių atveju
Table 4.10. Anisotropy ratio in the cases of spherical and *Gaussian* semivariogram models

Semivariogramos modelis	Anizotropijos santykis	θ_0	$\theta_0 + \theta_1$	θ_2
Sferinis	0,578	0,51164	1,26972	13501,39
<i>Gauso</i>	0,711	0,61988	1,27835	5142,48

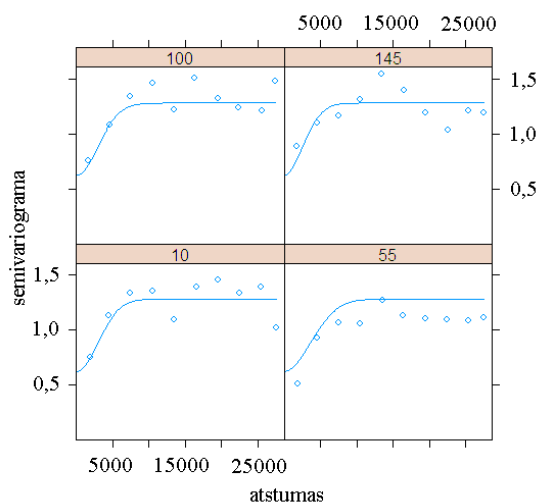
4.11 lentelė. Kryžminės patikros statistikos sferinio ir *Gauso* semivariogramų modelių atveju
Table 4.11. The statistics of cross-validation methods in the cases of spherical and *Gaussian* semivariogram model

Semivariogramos modelis	VPK	ŠVKPK	VKNS
Sferinis	-0,03867	1,47572	1,96029
<i>Gauso</i>	0,01509	1,45692	1,81656



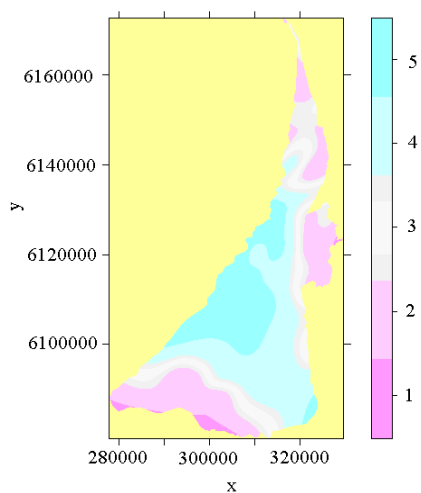
4.23 pav. Kryptinės semivariogramos priglodintos sferiniu semivariogramos modeliu

Fig. 4.23. The directional semivariogram fitted by spherical semivariogram model

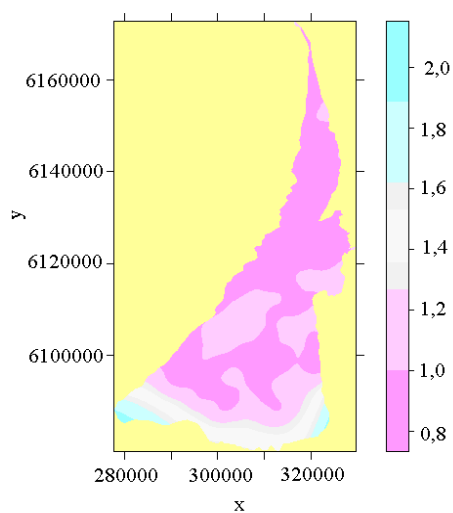


4.24 pav. Kryptinės semivariogramos priglodintos *Gauso* semivariogramos modeliu

Fig. 4.24. The directional semivariogram fitted by *Gaussian* semivariogram model



4.25 pav. Gylio prognozės rezultatai pašalinus išskirtį ir eliminavus geometrinę anizotropiją
Fig. 4.25. The prediction result of depth data without outlier after elimination of geometric anisotropy



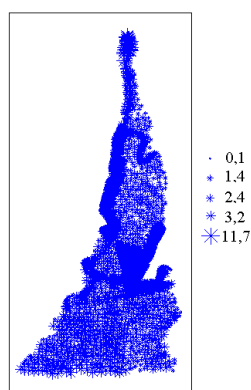
4.26 pav. Gylio dispersijos pašalinus išskirtį bei geometrinę anizotropiją (universalus krigingas)
Fig. 4.26. Variance of depth data without outlier after elimination of geometric anisotropy

Kadangi *Gauso* modelio atveju visos kryžminės patikros statistikos tenkina geriausio modelio kriterijus, gylio duomenims prognozuoti pašalinus išskirtį, III laipsnio polinomo trendą bei geometrinę anizotropiją naudotas *Gauso* semivariogramos modelis bei universalus krigingas. Prognozės ir prognozės dispersijų rezultatai pateikti atitinkamai 4.25 paveiksle ir 4.26 paveiksle.

4.2. 2005 metų gylio duomenų modeliavimas ir prognozė

Gavus 2005 metais matuotų gylio duomenų masyvą, atsirado galimybė prognozės rezultatus palyginti su realiais duomenimis. Tam buvo atrinkta 15 naujo masyvo matavimo taškų, kurių koordinatės artimos prognozuotų reikšmių koordinatėms. Siekiant praktiškai palyginti deterministines prognozės/interpoliacijos procedūras bei krigingo metodus, 1990 metų gylio duomenims pritaikyti trendo paviršiaus interpoliacijos bei atvirkštinio atstumo (kai $p=2^3$) metodai. Tiksliausiai prognozės/interpoliacijos procedūrai atrinkti taikytas PKS metodas. Lyginamosios analizės rezultatai pateikti 4.12 lentelėje.

Pradinė duomenų analizė. 2005 metų Kuršių marių duomenų išsibarstymas (4.27 pav.) ir pagrindinės 2005 metų gylio duomenų statistikos (4.13 lentelė) rodo, kad masyvą sudaro tik Lietuvai priklausančios teritorijos Kuršių marių gyliai.



4.27 pav. 2005 metų gylio duomenų išsibarstymas
Fig. 4.27. The distribution of the 2005 year depth data

³ Atvirkštinio atstumo interpoliacijos metodu kaimyniniams taškams suteikiamas svoris, kuris skaičiuojamas pagal formulę: $w(d) = 1/d^p$, čia d - atstumas, p - laipsnis.

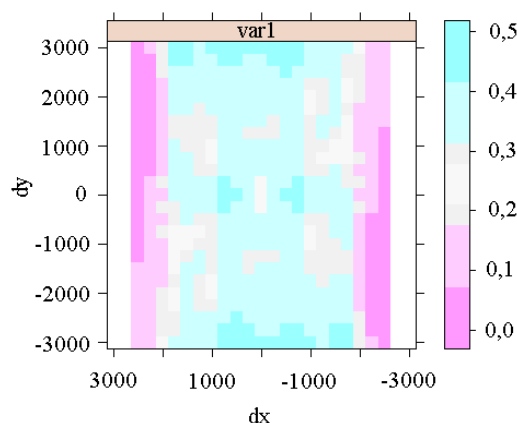
4.12 lentelė. Prognozės/interpoliacijos procedūrų lyginamoji analizė
Table 4.12. The comparative analysis of prediction/interpolation methods

2005 metų duomenys			Masyvo, kuris naudojamas prognozei ar interpoliacijai, koordinatės		Trendo paviršiaus interpoliacijos metodas	Atvirkštinio atstumo metodas, kai $p = 2$	Ordinarus krigingas	Universalus krigingas	Universalus krigingas, atsižvelgiant į anizotropiją	Ordinarus krigingas
X	Y	Gylis	X	Y	Gylis	Gylis	Gylis	Gylis	Gylis	Gylis
318890	6123675	4,6	318907	6123680	3,2343	3,3270	3,5179	3,4968	3,4870	3,3270
327114	6123874	0,4	327107	6123880	1,9314	1,2096	1,7842	1,5951	1,9676	1,2096
326105	6124297	2,8	326107	6124280	2,1557	2,5009	1,9325	1,8257	1,9572	2,5009
317891	6124481	4,6	317907	6124480	3,3187	3,8188	3,9199	3,8069	4,0276	3,8188
325621	6126288	1,6	325607	6126280	2,2817	2,5724	2,2608	2,1208	2,2936	2,5724
314058	6126579	5	314107	6126580	3,5606	4,2056	4,6877	4,4882	5,0236	4,2056
318623	6128179	4,6	318607	6128180	3,2088	3,9517	3,8186	3,7806	3,7531	3,9517
324834	6128577	3,2	324807	6128580	2,4459	2,7142	2,5648	2,4733	2,5668	2,7142
323581	6128885	2,8	323607	6128880	2,6316	2,5923	2,6369	2,7003	2,5155	2,5923
325327	6129171	3,6	325307	6129180	2,3693	2,6291	2,5747	2,4192	2,7056	2,6291
314914	6129278	5,2	314907	6129280	3,4707	4,4119	4,6046	4,3697	5,0185	4,4119
326135	6130487	2,6	326107	6130480	2,2502	2,6218	2,6194	2,3377	2,9450	2,6218
316286	6130772	4,8	317307	6130780	3,2843	3,8156	4,1543	4,0157	4,3952	3,8156
318801	6131188	4,4	318807	6131180	3,1592	4,0626	3,8164	3,7524	3,9143	4,0626
320811	6131478	3	320807	6131480	2,9696	3,2658	3,3700	3,3756	3,4135	3,2658
				PKS:	1,320103	0,531303	0,545162	0,619619	0,529168	0,531303

4.13 lentelė. Pagrindinės 2005 gylio duomenų statistikos**Table 4.13.** Statistics of depth data (2005 year)

	X koordinatė (metrais)	Y koordinatė (metrais)	Gylis (metrais)
Minimumas	303622	6123405	0,100
Pirmas kvartilis	317867	6137221	1,400
Mediana	320373	6141446	2,400
Vidurkis	320073	6143390	2,503
Trečias kvartilis	322911	6149732	3,200
Maksimumas	328290	61755434	11,700

Medianinio išlyginimo kringo taikymas 2005 metų gylio duomenų prognozei. Kadangi ankstesniuose tyrimuose nustatyta, kad 1990 metų Kuršių marių gylio duomenims būdinga dešinės pusės asimetrija, trendas bei anizotropija, nuspręsta naujesniems Kuršių marių gylio duomenims taikyti medianinio išlyginimo bei universalų kringą, ir nustatyti, kuris šių metodų naudojamus duomenis prognozuoja tiksliau.

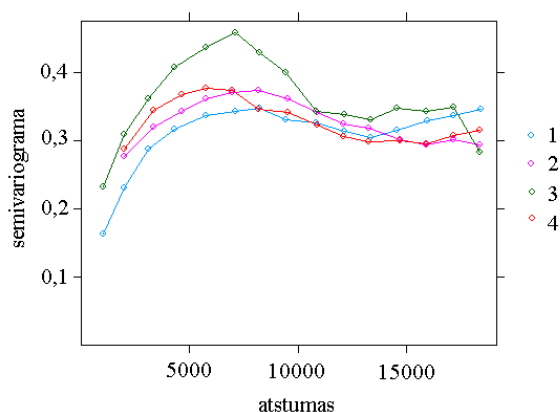
**4.28 pav.** Liekanų semivariogramos žemėlapis**Fig. 4.28.** The semivariogram map of residuals

Kaip jau buvo rašyta 2.3.3. skyrelyje medianinio išlyginimo kringo metodas taikomas tinklo tipo duomenims, kuriems būdingas trendas. Tad Kuršių marių gylio duomenys (2005 metai) buvo suskaidyti į 27 eilučių ir 52 stulpelių

tinklą. Toks eilučių ir stulpelių kiekis parinktas tam, kad tinklą sudarytų kvadratiniai 1800:1922 dydžio langeliai. Suskaidytų į tinklą duomenų minimumas – 0 m., o maksimumas – 9,69 m. Duomenų tinklo sudarymui panaudota *R* sistemos EDA paketo *medpolish* funkcija.

Po medianinio išlyginimo procedūros gauti eilučių, stulpelių medianų efektai bei liekanos. Anizotropijos tipui nustatyti, po medianinio išlyginimo funkcijos gautų liekanų pagrindu sudaryti semivariogramos žemėlapis (4.28 pav.) ir 4 kryptinių semivariogramų (10° , 55° , 100° ir 145° laipsnių kryptimi) diagrama (4.29 pav.) rodo, kad liekanų duomenys nėra izotropiniai.

Taip pat bandyta išanalizuoti, kurio laipsnio polinomo funkcija geriausiai aprašo liekanų, gautų po medianinio išlyginimo procedūros, ypatybes. Kaip parodė ANOVA rezultatai (4.14 lentelė), liekanoms nėra būdingas trendas.



4.29 pav. Liekanų kryptinės semivariogramos
Fig. 4.29. The directional semivariogram of residuals

Duomenų erdvinei sklaidai aprašyti naudoti sferinis, eksponentinis, *Gauso* semivariogramų modeliai, o zoninei anizotropijai pašalinti – 0,01 bei 0,0009 anizotropijos santykiai (4.15 lentelė). Tokie anizotropijos santykiai buvo parinkti todėl, kad *R* sistemos *Gstat* pakete (kaip ir kai kuriuose kituose programiniuose geostatistinės analizės paketuose) zoninė anizotropija pašalinama nurodant nuliui maksimaliai artimą anizotropijos santykį. Tokiu būdu ties begalybe semivariogramų su maksimaliu ir minimaliu slenksčiu atstumai tampa vienodi ir zoninė anizotropija virsta geometrine.

4.14 lentelė. Trendo nustatymo ANOVA rezultatai liekanoms gautoms po medianinio išlyginimo procedūros
Table 4.14. The results of ANOVA methods of detecting trend type for residuals after median polishing procedure

Trendo tipas	Daugialypis regresijos koeficientas R^2	AIC
I laipsnio polinomas	0,02431	(df = 3) -1473,368
II laipsnio polinomas	0,03714	(df = 6) -1485,956
III laipsnio polinomas	0,05016	(df = 10) -1497,066

4.15 lentelė. Liekanų semivariogramų parametrai, PKS bei kryžminės patikros statistikų rezultatai

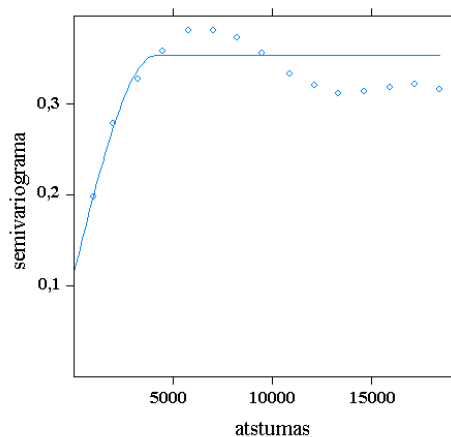
Table 4.15. The parameters of residuals semivariogram, the values of SSE and cross-validation methods

	Anizotropija egzistuoja		Anizotropija neegzistuoja
Semivariogramos modelis	Sferinis	EkspONENTINIS	Sferinis
Anizotropijos santykis	0,0080	0,0009	-
θ_0	0,11628	0,00000	0,11419
$\theta_0 + \theta_1$	0,35308	0,35517	0,35308
θ_2	89264,3	23817,9	4103,1
PKS	0,0000133	0,0000133	0,0000035
VPK	0,00021	0,00060	0,00043
ŠVKPK	0,44711	0,44686	0,45982
VKNS	0,8152	0,8281	1,0303

Kaip rodo PKS rezultatai, sferinis semivariogramos modelis, liekanų, gautų po medianinio išlyginimo procedūros, sklaidą aprašo geriau (kai teigiame, kad anizotropijos duomenyse nėra). Tad šis semivariogramos modelis buvo priglodintas prie visakryptės empirinės semivariogramos taškų (4.30 pav.) ir duomenų prognozei panaudotas ordinarus krigingas (4.31 pav. ir 4.32 pav.).

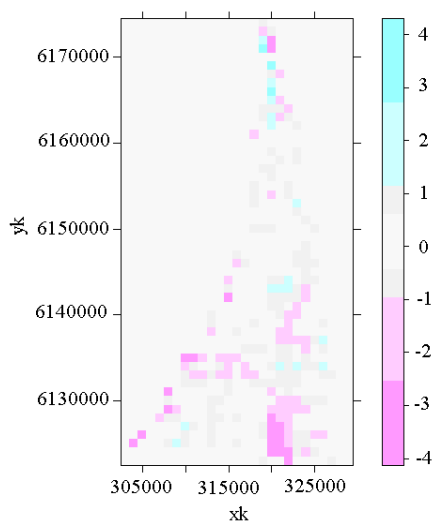
Įvertinus aukščiau minėtas semivariogramų analizės išvadas, liekanoms buvo pritaikytas ordinarus krigingas. Prie liekanų prognozės reikšmių pridėjus bendrą medianų efektą, stulpelių ir eilučių efektus gautos prognozuotos gylio

reikšmės. Paskaičiavus prognozuotų reikšmių minimumą ($\min = -1,065$ m.) bei maksimumą ($6,328$ m.) pastebėta, kad duomenų aibės plotis apima $\sim 7,393$ metro.



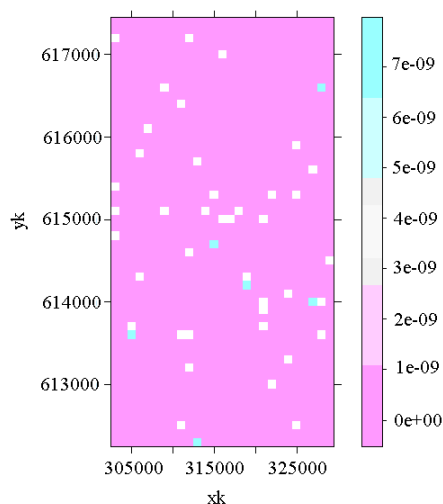
4.30 pav. Visakryptė sferinė semivariograma

Fig. 4.30. Omnidirectional spherical semivariogram



4.31 pav. Medianinio išlyginimo kringo prognozės rezultatai

Fig. 4.31. The prediction results of median polishing kriging



4.32 pav. Medianinio išlyginimo kringo dispersijos
Fig. 4.32. Variance of median polishing kriging

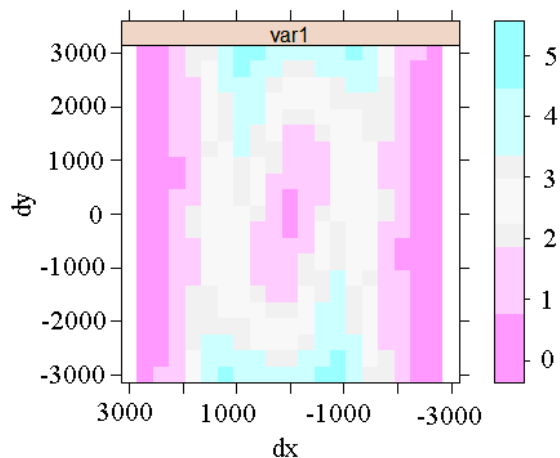
Universalaus kringo taikymas 2005 metų gylio duomenų prognozei. Prieš taikant tinklo duomenims universalų kringą ANOVA metodu bandyta nustatyti kokio tipo polinomo paviršius geriausiai aprašo gylio erdvinio vidurkio pasiskirstymą. Kaip ir 1990 metų gylio duomenų atveju, rezultatai rodo (4.16 lentelė), kad III laipsnio polinomas geriausiai aprašo gylio duomenis.

4.16 lentelė. Trendo nustatymo ANOVA rezultatai
 (2005 metų gylio duomenys)

Table 4.16. The results of ANOVA for detecting trend type
 (depth data of 2005 year)

Trendo tipas	Daugialypis regresijos koeficientas R^2	AIC
I laipsnio polinomas	0,2438	(df = 3) 1065,854
II laipsnio polinomas	0,4664	(df = 6) 582,221
III laipsnio polinomas	0,5786	(df = 10) 258,935

Siekiant nustatyti, ar duomenims būdinga anizotropija, sudarytas semivariogramos žemėlapis (4.33 pav.).



4.33 pav. Liekanų semivariogramos žemėlapis
Fig. 4.33. The semivariogram map of residuals

4.17 lentelė. Liekanų semivariogramų parametrai, PKS bei
 kryžminės patikros statistikų rezultatai
Table 4.17. The values of SSE, statistics of cross-validation method,
 the parameters of residual semivariograms

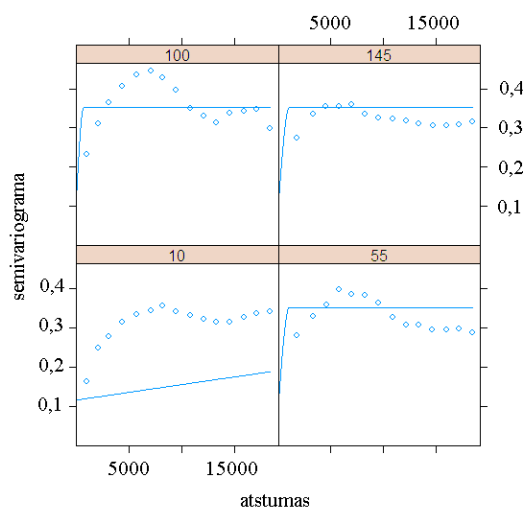
Trendo tipas	I laipsnio polinomas	II laipsnio polinomas	III laipsnio polinomas
Semivariogramos modelis	Sferinis	Sferinis	Sferinis
Anizotropijos santykis	0,0009	0,01	0,0009
θ_0	0,18594	0,19877	0,16729
$\theta_0 + \theta_1$	2,13249	1,47847	1,12792
θ_2	2154872	137423	1066097
PKS	0,001598	0,000742	0,000381
VPK	-0,0124	-0,0033	-0,0188
ŠVKPK	0,89325	0,61086	0,84505
VKNS	2,0068	0,6635	1,6905

Jis leido nustatyti, kad duomenų anizotropija vyksta šiaurės-rytų – pietvakarių kryptimis (apie 10 laipsnių pagal laikrodžio rodyklę). Kryptinėms semivariogramas 4 pagrindinėmis kryptimis pagal laikrodžio rodyklę pradedant 10° buvo pritaikyti sferinį, eksponentinį bei *Gauso* parametriniai semivariogramų modeliai. Taip pat šiems semivariogramų modeliams kiekvieno polinominio trendo atveju paskaičiuotos PKS bei kryžminių statistikų reikšmės (4.17 lentelė).

Rezultatai parodė, kad sferinis modelis visais atvejais geriausiai aprašo duomenų priklausomybę erdvėje (4.34 pav.) Tad šis semivariogramos modelis su 4.17 lentelė nurodytu anizotropijos santykiu naudotas 2005 metų gylio duomenis prognozuoti universaliu krigingu.

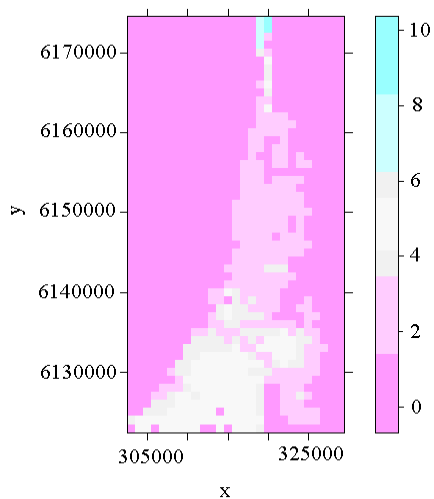
Paskaičiavus prognozuotų reikšmių minimumą (~0,000 m.) bei maksimumą (~9,689 m.) pastebėta, kad duomenų aibės plotis apima 9,689 m. Tai atitinka tinklo duomenų minimumo ir maksimumo reikšmes, bei įrodo, kad krigingas tose pačiose lokacijose sudaro nepaslinktą prognozę.

Siekiant 2005 metų gylio duomenų pagrindu prognozuoti gylio reikšmes visoje Kuršių marių srityje, kaip ir anksčiau buvo naudotas duomenų masyvas „kaukė“.



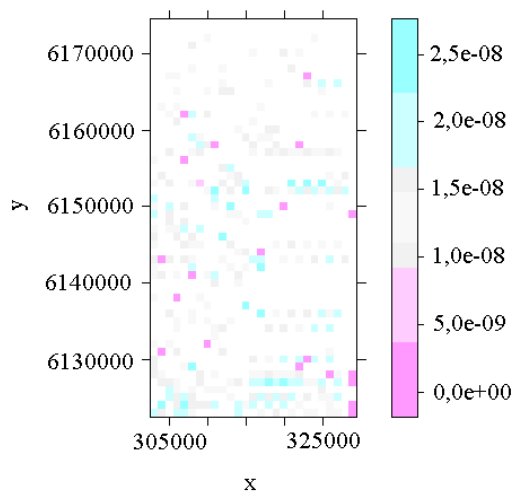
4.34 pav. Kryptinės semivariogramos pašalinus trečiojo laipsnio polinomo trendą

Fig. 4.34. The directional semivariograms after elimination of third-degree polinomial trend



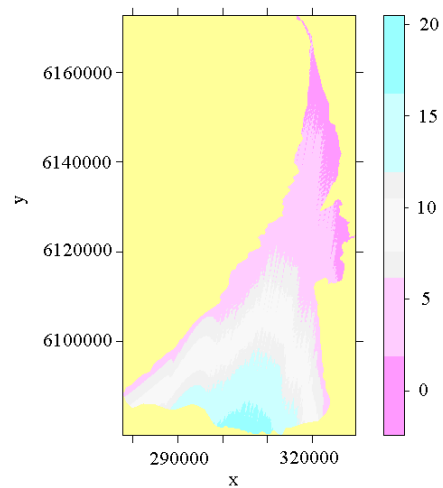
4.35 pav. Gylio liekanų prognozės rezultatai
(universalus krigingas)

Fig. 4.35. The prediction results of residuals
(universal kriging)



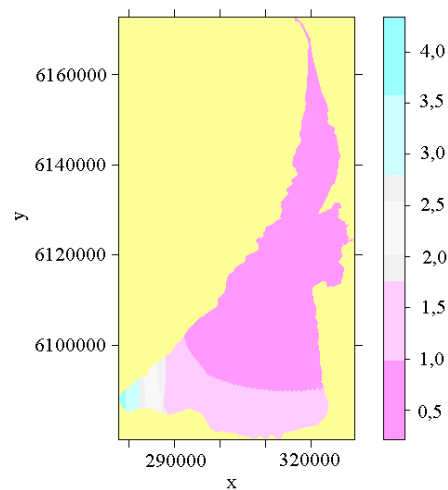
4.36 pav. Gylio liekanų dispersijos
(universalus krigingas)

Fig. 4.36. The variance of residuals of depth data
(universal kriging)



4.37 pav. 2005 metų gylio duomenų prognozės rezultatai
(universalus krigingas)

Fig. 4.37. The prediction results of depth data of year 2005
(universal kriging)



4.38 pav. 2005 metų gylio duomenų dispersijos
(universalus krigingas)

Fig. 4.38. Variance of depth data of year 2005
(universal kriging)

Visoje Kuršių marių srityje prognozuotų gylio duomenų minimumas = 1,017 m., o maksimumas – 19,070 m. Reikšmių plotis atitinkamai ~20 m. Šis rezultatas ženkliai skiriasi nuo pradinių duomenų reikšmių intervalo pločio. Tad šiuo atveju galima teigti, kad medianinio išlyginimo krigingo prognozės rezultatai lyginant su rezultatais, gautais taikant universalų krigingą, yra tikslesni.

4.3. Ketvirtojo skyriaus apibendrinimas ir išvados

Pritaikius siūlomą erdvinių duomenų modeliavimo ir prognozavimo metodiką išsiaiškinta, kad:

1. Kaip ir daugumai ekologinių duomenų, Kuršių marių gylio pasiskirstymams būdinga dešinės pusės asimetrija.
2. Nagrinėjant erdvines duomenų savybes Kuršių marių gylio duomenų erdvinę sklaidą daugeliu atveju tiksliausiai aprašo sferinis semivariogramos modelis.
3. Gylio duomenims būdingas trendas bei anizotropija. Tikėtina, kad šios nestacionarumo savybės tokio pobūdžio erdviniam procese atsiranda dėl įtekančių upių srovių.
4. Kelių prognozavimo procedūrų rezultatų ir 2005 metais išmatuotų duomenų lyginamoji PKS pagrindu analizė parodė, kad mažiausia PKS reikšmė gaunama duomenis prognozuojant krigingu, kai atsižvelgiama į duomenų trendą ir anizotropiją.
5. Medianinio išlyginimo krigingo prognozės rezultatai lyginant su universalaus krigingo rezultatais labiau artimi duomenų vidurkiui, dėl prognozės metodo robastiškumo.
6. *Gstat* programoje ir *R* sistemos *gstat* pakete nesukurta funkcija, realizuojanti medianinio krigingo išlyginimo procedūrą. Šiame darbe medianinio krigingo išlyginimo procedūros realizacija atlikta naudojant *base*, *spatial* ir *GSTAT* paketų *R* sistemoje funkcijų grupę.

Bendrosios išvados

Darbe apibendrinta ir realizuota vieninga erdvinių duomenų su stacionariomis klaidomis modeliavimo metodika. Gautus darbo rezultatus apibendriname ir pateikiame išvadas:

1. Išvesta nepastovaus vidurkio su laisvu nariu parametro $\hat{\beta}$ AMKM įvertinio išraiška per semivariogramas, kai semivariogramos parametrų vektorius θ yra žinomas. Pastabėta, kad galutinėje išraiškoje dipersijų-kovariacijų matricą atitinkamai keičia semivariogramos matrica.
2. Nepastovaus vidurkio su laisvu nariu atvejui išvestos D ir A kriterijų AMKM įvertinio išraiškos per semivariogramas. Išsiaiškinta, kad pastovaus vidurkio atveju šios išraiškos yra lygios.
3. Eksperimentas, kurio metu buvo siekiama nustatyti optimalų taškų išsidėstymą taisyklingoje 2x2 gardelėje, įrodė, kad D ir A kriterijų optimalumą garantuoja didžiausia atstumų tarp konfigūracijos taškų suma.
4. Taikant darbe siūlomą vieningą erdvinių duomenų su stacionariomis klaidomis modeliavimo metodiką atliktas skirtingais laiko periodais (1990 ir 2005 metais) matuotų Kuršių marių gylio duomenų erdvinis statistinis modeliavimas. Gylis duomenų erdvinei variacijai aprašyti taikyti tiesinis,

sferinis, eksponentinis ir/arba Gauso semivariogramų modeliai. Nustatyta, kad Kuršių marių gylio duomenų erdvinę sklaidą prie įvairių stacionarumo sąlygų daugeliu atveju tiksliausiai aprašo sferinis semivariogramos modelis.

5. Kuršių marių gylio erdvinio statistinio modeliavimo metu, išsiaiškinta, kad gylio duomenims būdingas trendas bei anizotropija. Tikėtina, kad šios nestacionarumo savybės tokio pobūdžio erdviniam procese atsiranda dėl į Kuršių marias įtekančių upių srovių.

6. 1990 metais matuotų gylio duomenų prognozei/interpoliacijai atlikti naudoti ordinarus krigingas, universalus krigingas, ordinarus kokrigingas (čia pagalbinio kintamuoju naudoti dugno nuosėdų duomenys), atvirkštinio atstumo metodas, trendo paviršiaus interpoliacijos metodas. Šių prognozavimo/interpoliavimo procedūrų rezultatų ir 2005 metais išmatuotų duomenų lyginamoji PKS pagrindu analizė parodė, kad mažiausia PKS reikšmė gaunama duomenis prognozuojant krigingu, kai atsižvelgiama į duomenų trendą ir anizotropiją.

7. Atlikus 2005 metų gylio duomenų prognozę medianinio išlyginimo krigingu bei universaliu krigingu bei palyginus prognozių paklaidas, nustatyta, kad medianinio išlyginimo prognozės rezultatai labiau artimi duomenų vidurkiui dėl prognozės metodo robastiškumo.

Literatūros sąrašas

- Akaike, H. 1973. Information theory as an extension of the maximum likelihood principle, in *2nd International Symposium on Information Theory*, 267–281.
- Arbia, G.; Lafratta, G. 2002. Anisotropic spatial sampling designs for urban pollution, *Applied Statistics* 51: 223–234.
- Barry, J.; Crowder, M.; Diggle, P. 1997. *Parametric Estimation of the Variogram*. Preprint. Department of Mathematics and Statistics. Lancaster University, Lancaster.
- Blyth, A.; Cake, D. *et al.* 2008. *Erdvinė analizė ir modeliavimas (GII-07)*: mokomoji knyga. Vilnius.
- Bogaert, P.; Russo, D. 1999. Optimal spatial sampling design for the estimation of the variogram based on a least-squares approach, *Water Resources Research* 35: 1275–1289.
- Brooker, P. A. 1991. *Geostatistical Primer*. Singapore: World Scientific.
- Chen, H.; Simpson, D.G.; Ying Z. 2000. Infill asymptotics for a stochastic process model with measurement error, *Statistica Sinica* 10: 141–156.
- Chauvet, P.; Galli, A. 1982. *Universal Kriging*. Tech. rep. C-96, Centre de Geostatistique, Ecole des Mines de Paris, Fontainebleau.
- Chiles, J. P., Delfiner, P. 1999. *Geostatistics Modeling Spatial Uncertainty*. New York: Wiley. 695 p.

- Christakos, G. 1984. On the problem of permissible covariance and variogram models, *Water Resources Research* 20(2): 251–265.
- Chou, Y. 1997. *Exploring Spatial Analysis in Geographic Information Systems*. Santa Fe, OnWord Press.
- Cliff, A. D.; Ord, J. K. 1973. *Spatial Autocorrelation*. Pion, London.
- Cliff, A. D.; Ord, J. K. 1981. *Spatial Processes; Models and Applications*. Pion, London.
- Cressie, N. 1985. Fitting variogram models by weighted least squares, *Mathematical Geology* 17: 693–702.
- Cressie, N. 1993. *Statistics for Spatial Data*. Wiley&Sons, New York.
- Cressie, N.; Hawkins, D. H. 1980. Robust estimation of the variogram: I, *Mathematical Geology* 12(2): 115–125.
- Cressie, N.; Zimmerman, D. L. 1992. On the stability of the geostatistical method, *Mathematical Geology* 24 (1): 45–49.
- Creutin, J. D.; Obled, C. 1982. Objective analyses and mapping techniques for rainfall fields: an objective comparison, *Water Resources Research* 18(2): 413–431.
- David, M. 1977. *Geostatistical Ore Reserve Estimation*. Amsterdam: Elsevier.
- Deutsch, C. V.; Journel, A. G. 1998. *GSLIB: Geostatistical Software Library and User's Guide (2nd ed.)*. Oxford University Press, New York.
- Diggle, P. J.; Ribeiro, Jr. P. J.; Christensen, O. F. 2003. *An Introduction to Model-based Geostatistics. Spatial Statistics and Computational Methods* [interaktyvus]. Springer-Verlag, New York. [žiūrėta 2009-09-25]. Prieiga per internetą: <<http://leg.est.ufpr.br/~paulojus/pub/MBGchapter.pdf>>
- Diggle, P. J. 2003. *Statistical Analysis of Spatial Point Patterns*, Second Edition. London: Arnold Publishers.
- Doraj-Raj, S. S. 2001. *First- and Second-Order Properties of Spatiotemporal Point Processes in the Space-Time and Frequency Domains*. Blacksburg, Virginija.
- Dučinskas, K.; Šaltytė-Benth, J. 2003. *Erdvinė statistika*. KU, Klaipėda.
- Fedorov, V.V. 1996. *Design of spatial experiments: model fitting and prediction*. In *Handbook of Statistics* 3.
- Fortin, M. J., Dale, M. R. T., Ver Hoef, J. M. 2001. Spatial analysis in ecology, *Encyclopedia of Environmetrics* 3: 2051–2058.
- Gallichand, J.; Marcotte, D. 1993. Mapping clay content for subsurface drainage in the Nile Delta, *Geoderma* 58: 165–179.

- Genton, M. G. 1997. *Variogram Fitting by Generalized Least Squares Using an Explicit Formula for The Covariance Structure*. Preprint. Massachusetts Institute of Technology, Cambridge.
- Genton, M. 1998. Highly robust variogram estimation, *Mathematical Geology* 30(2): 213–221.
- Goodchild, M. F. 1987. A spatial analytical perspective as geographical information systems, *International Journal of Geographical Information Systems* 1(4): 327–334.
- Goovaerts, P. 1997. *Geostatistics for Natural Resources Evaluation*. Oxford Univ. Press, New-York.
- Gringarten, E.; Deutsch, C. V. 2001. Teacher's aide semivariogram interpretation and modeling, *Mathematical Geology* 33(4): 507–534.
- Grimm, J. W.; Lynch, J. A. 1991. Statistical analysis of errors in estimating wet deposition using five surface estimation algorithms, *Atmospheric Environment* 25(2): 317–327.
- Handcock, M. S. 1989. *Inference for Spatial Gaussian Random Fields when the Objective is Prediction*. PhD. Thesis. Department of Statistics, University of Chicago, Chicago, Illinois.
- Handcock, M. S.; Stein, M. L. A 1993. Bayesian analysis of kriging, *Technometrics* 35: 403–410.
- Handcock, M. S.; Wallis, J. R. 1994. An approach to statistical spatio-temporal modeling of meteorological fields, *Journal of The American Statistical Association* 89: 368–378.
- Haining, R. 1990. *Spatial Data Analysis in the Social and Environmental Sciences*, Cambridge, U.K.: Cambridge University Press.
- Heikkinen, J. *Spatial Statistics* [interaktyvus]. University of Helsinki. Rolf Nevanlinna institute. Paskaitų konspektas. [žiūrėta 2009-09-25]. Prieiga per internetą: <http://www.rni.helsinki.fi/~jmh/ss03/contents.html>.
- Helterbrand, J. D.; Cressie, N. 1994. Universal cokriging under intrinsic coregionalization, *Mathematical Geology*, 205–226.
- Hillier, G.; Martellosio, F. 2006. Spatial design matrices and associated quadratic forms: structure and properties, *Journal of Multivariate Analysis* 97: 1–18.
- Hohn, M. 1999. *Geostatistics and Petroleum Geology* (2nd ed.). Dordrecht: Kluwer.
- Isaaks, E. H.; Srivastava, R. M. 1989. *An Introduction to Applied Geostatistics*. Oxford University Press.
- Jokšas, K.; Galkus, A.; Stakėnienė, R. 1998. Geocheminiai Kuršių marių dugno nuosėdų ypatumai ir juos formuojantys veiksniai, *Geografijos metraštis* 31: 123–144,
- Journel, A.; Huijbregts, C. 1978. *Mining Geostatistics*. New York: Academic Press.

- Kaluzny, S.; Vega, S.; Cardoso, T.; Shelly, A. 1996. *S+SpatialStats user's manual*. MathSoft Inc., Seattle.
- Kitanidis, P. K. 1983. Statistical estimation of polynomial generalized covariance functions and hydrologic applications, *Water Resources Research* 19: 909–921.
- Kumetaitis, A. 2004. Skaitmeninio reljefo modelio sudarymas ir tobulinimas geostatistiniais metodais, *Geodezija ir kartografija* 1: 29–34.
- Kumetaitienė, A. 2006. Skaitmeninio reljefo modelio sudarymas skirtingais geostatistiniais reljefo modeliavimo metodais, *Geografija* 42(1): 28–32.
- Kumetaitienė, A. 2006. Skaitmeninio reljefo modelio tikslumo priklausomybė nuo ląstelės matmenų ir atstumo tarp modeliavimui naudojamų pradinų duomenų taškų, *Geodezija ir kartografija* 1: 14–18.
- Lahiri, S. N.; Lee, Y.; Cressie, N. 2002. On asymptotic distribution and asymptotic efficiency of least squares estimators of spatial variogram parameters, *Journal of Statistical Planning and Inference* 103: 65–85.
- Lark, R. M. 2000. Estimation of the variograms of soil properties by the method-of-moments and maximum likelihood; a comparison, *European Journal of Soil Science* 51: 717–728.
- Lark, R. M.; Cullis, B. R. 2004. Model-based analysis using reml for inference from systematically sampled data on soil, *European Journal of Soil Science* 55: 799–813.
- Laslett, G. A.; McBratney, A. B.; Pahl, P. J.; Hutchinson M. F. 1987. Comparison of several spatial prediction methods for soil pH, *Journal of Soil Science* 38: 325–41.
- Laslett, G. 1994. Kriging and splines: an empirical comparison of their predictive performance in some applications, *Journal of The American Statistical Association* 89(426): 391–405.
- Lee, Y. D.; Lahiri, S. N. 2002. Least squared variogram fitting by spatial subsampling, *Royal Statistical Society* 64: 837–854.
- Lophaven, S. N. *Reconstruction of Data From the Aquatic Environment* [interaktyvus] Master's thesis. LYNGBY, 2001. [žiūrėta 2009-09-25]. Prieiga per internetą: < http://www2.imm.dtu.dk/pubdb/views/edoc_download.php/730/pdf/imm730.pdf >
- Mardia, K.V.; Marshall, R. J. 1984. Maximum likelihood estimation of models for residual covariance in spatial regression, *Biometrika* 71: 135–146.
- Martin, J. D.; Simpson, T. W. 2005. Use of kriging models to approximate deterministic computer models, *AIAA Journal* 43(4): 853–863.
- Matheron, G. 1963. Principles of geostatistics, *Economic Geology* 58: 1246–1266.
- Matheron, G.; Armstrong, M. 1987. *Geostatistical Case Studies (Quantitative Geology and Geostatistics)*, Dordrecht, Holland.

- McBratney, A. B.; Webster, R.; Burgess, T. M. 1981. The design of optimal sampling schemes for local estimation and mapping of regionalized variables, *Computer & Geosciences* 7(4): 331–334.
- McBratney, A. B. Webster, R. 1986. Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates, *Soil Sciences* 37: 617–639.
- Meckesheimer, M.; Barton, R. R.; Simpson, T. W.; Booker, A. J. 2002. Computationally inexpensive metamodel assessment strategies, *AIAA Journal* 40(10): 2053–2060.
- Myers, D. E. 1991. On variogram estimation, in *Proceedings of the First Inter. Conf. Stat. Comp. American Sciences Press II*: 261–281.
- Myers, D. E. 1988. Multivariate geostatistics for environmental monitoring, *Sciences de la Terre* 27: 411–427.
- Myers, D. E. 1992a. Cokriging, radial basis functions, and the role of positive definiteness, *Computers and Mathematical Applications* 24: 139–148.
- Myers, D. E. 1995. The linear coregionalization and simultaneous diagonalization of the semivariogram matrix function, *Sciences de la Terre* 32: 125–139.
- Muller, W. G.; Zimmerman, D. L. 1999. Optimal designs for variogram estimation, *Environmetrics* 10: 23–37.
- Muller, W. G. 2001. *Collecting data: Optimum Design of Experiments for Random Fields*. Springer: New York.
- Olea, R. A. 1975. *Optimum Mapping Techniques Using Regionalized Variable Theory*, Lawrence. KS: Kansas Geological Survey.
- Pannatier, Y. 1996. *VARIOWIN: Software for Spatial Data Analysis*, New York: Springer-Verlag.
- Pebezma, E. J. 1999. *Gstat User's Manual* [interaktyvus]. De. pt. of Physical Geography, Utrecht University. [žiūrėta 2009-09-25]. Prieiga per internetą: <<http://www.gstat.org/GSTAT.pdf>>
- Pebezma, E. J.; Wesseling, C. G. 1998. Gstat: a program for geostatistical modelling, prediction and simulation, *Computers & Geosciences* 24(1): 17–31.
- Pebezma, E. J. 2006.09.15 *The Meuse Data Set: A Tutorial for The Gstat R Package* [interaktyvus]. [žiūrėta 2009-09-25]. Prieiga per internetą: <<http://cran.r-project.org/web/packages/gstat/vignettes/gstat.pdf>>
- Pukelsheim, F. 1993. *Optimum Design of Experiments*. Wiley: New York.
- Rossiter, D. G. 2004. *An Introduction to Applied Geostatistics*. Department of Earth Systems Analysis. International Institute for Geo-information Science & Earth Observation.

- Rouhani, S. 1986. Comparative study of groundwater mapping techniques, *Groundwater* 24(2): 207–216.
- Roustant, O.; Dupuy, D.; Helbert, C. 2007. *Robust Estimation of The Variogram in Computer Experiments*, ENBIS 7 Dortmund.
- Ritter, K. 1996. Asymptotic optimality of regular sequence designs, *Annals of Statistics* 24: 2081–2096.
- Sain, S. R. 2004. *Geostatistical Data and Semivariogram Models* [interaktyvus]. Department of Mathematics University of Colorado at Denver. [žiūrėta 2009-09-25]. Prieiga per internetą: <www-math.cudenver.edu/~ssain/spatstat>.
- Schabenberger, O.; Pierce, F. J. 2001. *Contemporary Statistical Models for the Plant and Soil Sciences* [interaktyvus]. CRC Press, LLC. [žiūrėta 2009-09-25]. Prieiga per internetą: <<http://filebox.vt.edu/s/schabenb/book/>>
- Simpson, T. W.; Mauery, T. M.; Korte, J. J.; Mistree, F. 2001. Kriging metamodels for global approximation in simulation-based multidisciplinary design optimization, *AIAA Journal* 39(12): 2233–2241.
- Smith, R. L. 2000. Spatial statistics in environmental science, *Nonlinear and Nonstationary Signal Processing*, 152–183.
- Smith, R. L. 2001. *Environmental Statistics*. The University of North Carolina Chapel Hill, NC 27599–3260, USA.
- Srivastava, R. M. A. 1990. Geostatistical paradox, *Geostatistics* 4(1): 9–10.
- Su, Y.; Cambanis, S. 1993. Sampling design for estimation of a random process, *Stochastic Process Applications* 46: 47–89.
- Switzer, P. 1984. Inference for spatial autocorrelation functions, *Geostatistics for Natural Resources Characterization* 1: 127–140.
- Tabios, G. Q.; Salas, J. D. 1985. A comparative analysis of techniques for spatial interpolation of precipitation, *Water Resources Bulletin* 21: 365–380.
- Tukey, J. W. 1977. *Exploratory Data Analysis*, Reading, MA: Addison-Wesley.
- Tumas, R. 2006. *Aplinkos geoinformacijos sistemas*. Vilnius, 263 p.
- van Beers, W. C. M.; Kleijnen, J. P. C. 2004. Kriging interpolation in simulation: a survey, *Simulation Conference* 1: 113–121.
- Vann, J.; Guibal, D. 1998. *Beyond Ordinary Kriging - an Overview of Non-Linear Estimation*. Geostatistical Association of Australasia. Monograph 1. Perth, Western Australia.
- Venables, W. N.; Smith, D. M. The R Development Core Team. 2003. *An Introduction to R* [interaktyvus]. [žiūrėta 2009-09-25]. Prieiga per internetą: <<http://cran.r-project.org/doc/manuals/R-intro.pdf>>

- Zhang, X.; van Eijkeren, J.; Heemink, A. 1995. On the weighted least-squares method for fitting a semivariogram model, *Computers & Geosciences* 21: 605–608.
- Zhang, H.; Zimmerman, D. L. 2005. Towards reconciling two asymptotic frameworks in spatial statistics, *Biometrika* 92(4): 921–936.
- Zhu, Z. 2002. *Optimal Sampling Design and Parameter Estimation of Gaussian Random Fields*. PhD Thesis. Department of Statistics, University of Chicago.
- Zimmerman, D. L.; Zimmerman, M. B. 1991. A Comparison of Spatial Semivariogram Estimators and Corresponding Ordinary Kriging Predictors, *Technometric* 33(1): 77–91.
- Waters, N. M. 1997. *Spatial Interpolation I*. University of Calgar. [žiūrėta 2009-09-25]. Paskaitų konspektas. Prieiga per internetą: <<http://www.geog.ubc.ca/courses/klink/gis.notes/ncgia/u40.html#SEC40.2>>
- Wackernagel, H. 2003. *Multivariate Geostatistics*. Springer.
- Warner, J. J.; Ripley, B. D. 1987. Problems with the likelihood estimation of covariance functions of spatial gaussian processes, *Biometrika* 74: 640–642.
- Warrick, A. W.; Myers, D. E. 1987. Optimization of sampling locations for variogram calculations, *Water Resources Research* 23: 496–500.
- Weber, D. D.; Englund, E. L. 1992. Evaluation and comparison of spatial interpolators, *Mathematical Geology* 24: 381–391.
- Weber, D. D.; Englund, E. J. 1994. Evaluation and comparison of spatial interpolators II, *Mathematical Geology* 5(26): 589–603.
- Webster, R.; McBratney, A. B. 1989. On the Akaike information criterion for choosing models for variograms of soil properties, *Journal of Soil Sciences* 40: 493–496.
- Webster, R.; Oliver, M. 1992. Sample adequately to estimate semivariograms of soil properties, *Journal of Soil Sciences* 43: 177–192.

Autoriaus mokslinių publikacijų disertacijos tema sąrašas

Straipsniai recenzuojamuose mokslo žurnaluose

Krūminienė (Borisenko), I.; Dučinskas, K.; Garška, R. 2003. Applying of kriging and cokriging methods for prediction of Curonian Lagoon data, *Lietuvos matematikos rinkinys*. Spec. nr., 43: 504–508, ISSN 0132-2818.

Krūminienė (Borisenko), I.; Garška, R. 2004. Spatial analysis and prediction of Curonian Lagoon data with Gstat, *Mathematical Modelling and Analysis* 9(1): 39–50. ISSN 1392-5113.

Krūminienė (Borisenko), I.; Dučinskas, K. 2004. Generalized product-sum semivariogram model of the data of the center of marine research, *Lietuvos matematikos rinkinys*, Spec. nr. 44: 736–741. ISSN 0132-2818.

Krūminienė I. 2006. Analysis of anisotropic semivariogram models for prediction of the Curonian Lagoon data, *Mathematical Modelling and Analysis* 11(1): 73–86. ISSN 1392-5113 (Thomson ISI Web of Science).

Krūminienė (Borisenko), I.; Dučinskas, K. 2004. Comparison ordinary and universal kriging of Curonian Lagoon by cross-validation method, *Technologijos mokslo darbai vakarų Lietuvoje IV*, 185–191. ISBN 9955-585-53-6.

Krūminienė (Borisenko) I. 2004. Using universal block kriging to prediction of the Curonian Lagoon depth data with gstat package of R system, *Vadyba* 1(4):46–50. ISSN 1648-7974.

Straipsniai kituose leidiniuose

Krūminienė (Borisenko), I. 2004. Using gstat package of R program for spatial data prediction, in *9-osios magistrantų ir doktorantų konferencijos „Informacinė visuomenė ir universitetinės studijos“ pranešimų medžiaga*, 213–219.

Dučinskas, K.; Borisenko, I. 2009. Optimal sampling design for the GLS estimators of the mean parameters for the stationary spatial random process, in *Proceedings of XIII International Conference “Applied Stochastic Models and Data Analysis”*.

Priedai

A priedas. Pradinės statistinės duomenų analizės komandos

Šių komandų dėka atlikta pradinė 1990 gylio ir dugno nuosėdų duomenų prognozė *R* programoje: apskaičiuotos pagrindinės statistikos; duomenų skirstinio normalumo patikrinimui sudaryta histograma ir penkiaskaitė suvestinė, skirstinio normalumo grafikas; duomenų išsidėstymui nagrinėjamoje Kuršių marių teritorijoje vizualizuoti sudaryta burbulinė diagrama. Šios komandos analogiškai taikytos ir 2005 metų gylio duomenų pradinei analizei.

```
#Įkeliamas 1990 metų gylio duomenų masyvas
gyliai<-read.table("c:/tyrimas/h.txt", header = TRUE)
#Skaičiuojamos pagrindinės statistikos
summary(gyliai)
#Įkeliamas dugno nuosėdų masyvas
frakcijos<-read.table("c:/tyrimas/g.txt", header = TRUE)
summary(frakcijos)
#Įkeliamas visos Kuršių marių srities masyvas
marios<-read.table("c:/research/tyrimas0/marios.txt", header = TRUE)
summary(marios)
```

```

#1990 metų gylio duomenų masyvas išskaidomas į atskirus stulpelius
attach(gyliai)
#Sudaroma histograma
hist(z, main = "Gylio histograma", xlab = "Gylis", ylab = "Daznis", ylim = c(0,
100))
#Sudaroma penkiaskaitė suvestinė
boxplot(z, main = "Gylio penkiaskaitė suvestinė")
#Sukuriamas skirstinio normalumo grafikas
qqnorm(z, main = "Q-Q grafikas", xlab = "Teoriniai kvantiliai", ylab = "Imties
kvantiliai")
#Sudaroma logaritmuotų gylio duomenų histograma
hist(log(z), main = "Gylio logaritmuotu duomenų histograma", xlab = "Gylis",
ylab = "Daznis", xlim = c(-3, 3))
#Sudaroma logaritmuotų gylio duomenų penkiaskaitė suvestinė
boxplot(log(z), main = "Gylio logaritmuotu duomenų penkiaskaitė suvestinė")
#Sudaroma 1990 metų gylio reikšmių burbulinė diagrama
coordinates(gylis) <- c("x", "y")
bubble(gylis, zcol = 3, maxsize = 1, pch = 8, col = c(3, 4), main = "Kursiu mariu
gylis (1990 metai)")

```

B priedas. 1990 gylio duomenų prognozė ordinariu krigingu

Šių komandų dėka atliktas 1990 metais matuotų gylio duomenų erdvinės elgsenos modeliavimas tiesine semivariograma bei prognozė universalium krigingu. Komandos vykdomos R programoje *gstat* pakete.

```
#gstat paketo aktyvavimas
library(gstat)
#Empirinės semivariogramos sudarymas
v1<- variogram(z ~1, loc=~x+y, data=h)
# Semivariogramos grafikas
plot(v1)
#Sferinio semivariogramos modelio priglodinimas empiriniams
duomenims
fv1<-fit.variogram(v1, vgm(1.13, "Sph",40000, 1.5))
#Empirinės semivariogramos prigludinto semivariogramos modelio
grafikas
plot(v1, model = fv1)
#Marių kontūro duomenų masyvo įkėlimas
marios<-read.table("c:/tyrimas/marios.txt", header = TRUE)
#Duomenų prognozė ordinariu krigingu
x1 <- krige(z ~1, ~x+y, model = fv1, data = h, newdata = marios)
#Prognozės rezultatų išsaugojimas kintamuosiuose
p1<-levelplot(var1.pred~ x+y, x1, aspect = mapasp(x1), main = "ordinary kriging
predictions")
p2<-levelplot(var1.var~ x+y, x1, aspect = mapasp(x1), main = "ordinary kriging
variance")
#Prognozės ir dispersijų žemėlapių atvaizdavimas
print(p1, split = c(1, 1, 2, 1), more = TRUE)
print(p2, split = c(2, 1, 2, 1))
```

C priedas. 1990 gylio duomenų prognozė universalium krigingu

Šios R programos *spatial* ir *gstat* paketų komandos leidžia nustatyti polinominio trendo paviršiaus laipsnį ANOVA metodu, nustatytą polinomą pašalinti iš duomenų ir liekanų pagrindu sudaryti semivariogramą, duomenis prognozuoti universalium krigingu.

```
#I, II, ir III laipsnio polinomų taikymas duomenims mažiausių kvadratų
metodu
h2.surf1 <- surf.ls(1, h2)
h2.surf2 <- surf.ls(2, h2)
h2.surf3 <- surf.ls(3, h2)
#ANOVA lentelės sudarymas vienam ar keliems polinominiams paviršiams
summary(h2.surf1)
summary(h2.surf2)
summary(h2.surf3)
#Robustinės semivariogramos sudarymas, pašalinus III laipsnio polinomo
trendą
g <- gstat(id = "gylis ", form = z ~
x+y+I(x^2)+I(x*y)+I(y^2)+I(x^3)+I(y^3)+I(x^2*y)+I(y^2*x), loc=~x+y, da-
ta=h)
v2 <- variogram(g, alpha = c(10, 55, 100, 145), cressie = TRUE, cutoff = 28000,
width = 3000)
model_var2<-fit.variogram(v2, vgm(1.0, "Gau",10000, 0.5, anis = c(55, 0.9))
+ )
#Duomenų prognozė universalium krigingu
x3 <- krige(z x+y+I(x^2)+I(x*y)+I(y^2)+I(x^3)+I(y^3)+I(x^2*y)+I(y^2*x),
~x+y, model = fv3, data = h2, newdata = marios)
#Prognozės rezultatų išsaugojimas kintamuosiuose
p1<-levelplot(var1.pred~ x+y, x3, aspect = "iso", main = "Universal kriging
predictions")
p2<-levelplot(var1.var~ x+y, x3, aspect = "iso", main = "Universal kriging
variance")
#Prognozės ir dispersijų žemėlapių atvaizdavimas
print(p1, split = c(1, 1, 2, 1), more = TRUE)
print(p2, split = c(2, 1, 2, 1))
```

D priedas. 2005 gylio duomenų, atvaizduotų tinklu, prognozė medianinio išlyginimo kriginu

```
# Aktyvuojamas sp paketas
library(sp)
# Sudaromas gylio duomenų tinklas iš 27 stulpelių ir 52 eilučių
marios<-read.table("c:/tyrimas/marios.txt", header = TRUE)
grd_marios <- GridTopology(c(303500,6123500), c(1000,1000), c(27,52))
sgrd_marios <- SpatialGrid(grd_marios)
bbox(sgrd_marios)
# Tikrinama, ar žemėlapiu rėmas apima x ir y minimalią ir maksimalią
koordinates
coordinates(marios) <- c("x", "y")
m <- overlay(sgrd_marios, marios)
t00 <- lapply(split(as(marios, "data.frame"), m), mean)
t01<-vector(mode = "list", length=prod(slot(slot(sgrd_marios, "grid"),
"cells.dim"))))
nafill0 <- rep(as.numeric(NA), ncol(as(marios, "data.frame")))
t01 <- lapply(t01, function(x) nafill0)
t01[as.integer(names(t00))] <- t00
t02<- data.frame(t(data.frame(t01)))
sgrd_marios01 <- SpatialGridDataFrame(slot(sgrd_marios, "grid"), t02)
write.asciigrid(sgrd_marios01, "c:/tyrimas/mtinklas.txt", attr = 3, na.value = NA)
#Ištrinama ascii vektoriaus antraštė
mt<-read.table("c:/tyrimas/mtinklas.txt", header = FALSE)
mtinklas<-as.matrix(mt)
mvektorius<-as.vector(mtinklas)
#write(mtinklas2, "c:/tyrimas/mtinklas2.txt")
gyliai<-read.table("c:/tyrimas/gyliai_2006.txt", header = TRUE)
grd_gyliai <- GridTopology(c(303500,6123500), c(1000,1000), c(27,52))
sgrd_gyliai <- SpatialGrid(grd_gyliai)
bbox(sgrd_gyliai)
coordinates(gyliai) <- c("x", "y")
m2 <- overlay(sgrd_gyliai, gyliai)
t0 <- lapply(split(as(gyliai, "data.frame"), m2), mean)
t1<-vector(mode = "list", length=prod(slot(slot(sgrd_gyliai, "grid"),"cells.dim"))))
nafill <- rep(as.numeric(NA), ncol(as(gyliai, "data.frame")))
t1 <- lapply(t1, function(x) nafill)
t1[as.integer(names(t0))] <- t0
t2<- data.frame(t(data.frame(t1)))
```

```

sgrd_gyliai1 <- SpatialGridDataFrame(slot(sgrd_gyliai, "grid"), t2)
# Medianoms suskaičiuoti, NA masyvo reikšmė pakeičiama nuliu
write.asciigrid(sgrd_gyliai1, "c:/research/medianinis/gyliaitinklas.txt", attr = 3,
na.value = 0)
gvektorius2<-as.vector(sgrd_gyliai1)
is.vector(gvektorius2)
ntinklas<-read.table("c:/tyrimas/gyliaitinklas.txt", header = FALSE)
matrica<-as.matrix(ntinklas)
is.matrix(matrica)
mpmatrica<-medpolish(matrica, eps = 0.01, na.rm = TRUE)
mpmatrica$overall
mpbendra = mpmatrica$overall
mpcol = mpmatrica$col
mprow = mpmatrica$row
#Skaičiuoklėje 9999 pakeičiami į 0 ir tinklas perdaromas į vieną stulpelį.
ntinklas<-read.table("c:/tyrimas/gyliaitinklas.txt", header = FALSE)
matrica<-as.matrix(ntinklas)
is.matrix(matrica)
# Tikrinama, ar gautas tinklas pertvarkytas į matricą
mpmatrica<-medpolish(matrica, eps = 0.01, na.rm = TRUE)
mpbendra = mpmatrica$overall
mpcol = mpmatrica$col
mprow = mpmatrica$row
liekanos<-mpmatrica$residuals
write.table(liekanos, "c:/tyrimas/res.txt")
res<-read.table("c:/tyrimas/res.txt", header = TRUE)
coordinates(res) <- c("xk", "yk")
bubble(res, zcol = 3, maxsize = 0.5,pch = 8, col = c(3, 4))
resvar <- semivariogram(resg ~ 1, loc = ~xk+yk, data = res, map = TRUE, cutoff
= 30000, width = 2500)
plot(resvar)
gr <- gstat(id = "resg ", form = resg ~ 1, loc=~xk+yk, data=res)
vr <- semivariogram(gr)
plot(vr)
gr <- gstat(id = "resg ", form = resg ~ 1, loc=~xk+yk, data=res)
vr <- semivariogram(gr)
plot(vr)
model_resvar<-fit.variogram(vr, vgm(0.2, "Sph", 5000, 0.1))
model_resvar
plot(vr, model = model_resvar)
attr(model_resvar, "SSErr")

```

```

#Liekanų tinklo įkėlimas prieš duomenų prognozavimą
res_grid<-read.table("c:/tyrimas/res_grid.txt", header = TRUE)
#Ordinaraus kriginio taikymas duomenų prognozei
gr_krige <- krige(formula = resg~1, loc = ~xk+yk, model = model_resvar, data =
res, newdata = res_grid)
library(lattice)
pl1 <- levelplot(var1.pred ~ xk + yk, gr_krige, aspect = "iso", main = "Liekanu
OK prognoze")
pl2 <- levelplot(sqrt(var1.var) ~ xk + yk, gr_krige, aspect = "iso", main = "Lie-
kanu OK klaidos")
plot(pl1)
plot(pl2)
x_gr_or <- krige.cv(formula = resg~1, loc = ~xk+yk, model = model_resvar, data
= res, nmax = 40)
rmse <- function(xv.obj){
resid <- xv.obj$residual
return(sqrt(sum((resid)^2)/length(resid)))}
rmse(x_gr_or)
me <- function(xv.obj){
resid <- xv.obj$residual
return(sum(resid)/length(resid))}
me(x_gr_or)
msdr <- function(xv.obj){
resid <- xv.obj$residual
var<- xv.obj$var1.var
return(sum(resid^2/var)/length(resid))}
msdr(x_gr_or)
liek<-gr_krige$var1.pred
liekanos<-as.matrix(liek)
is.matrix(liek)
write.table(liekanos, "c:/tyrimas/liekanos.txt")
#Sutvarkoma gauta eilutė i matricą
liekanos<-read.table("c:/tyrimas/krige_pred.txt", header = FALSE)
liek<-data.matrix(liekanos)
is.matrix(liek)
ntinklas<-read.table("c:/tyrimas/gyliaitinklas.txt", header = FALSE)
matrica<-as.matrix(ntinklas)
is.matrix(matrica)
mpmatrica<-medpolish(matrica, eps = 0.01, na.rm = TRUE)
mpmatrica$overall
mpbendra = mpmatrica$overall

```

```
mpcol = mpmatrica$col
mprow = mpmatrica$row
rc<-matrix(0, nrow = length(mprow), ncol = length(mpcol))
rcb<-rc+mpbendra
zp0<-rcb+liek
marios_tinklas<-read.table("c:/tyrimas/mtinklas.txt", header = FALSE)
mtinklas<-as.matrix(marios_tinklas)
is.matrix(mtinklas)
prediction<-zp0*mtinklas
write.table(prediction, "c:/tyrimas/prediction.txt")
```

Ingrida BORISENKO

ingrida@ik.ku.lt

ERDVINIŲ DUOMENŲ MODELIAVIMAS NAUDOJANT STACIONARIŲ ERDVINIŲ
PROCESŲ SEMIVARIOGRAMAS

Daktaro disertacija

Fiziniai mokslai,
matematika (01P)

MODELLING OF SPATIAL DATA USING SEMIVARIOGRAMS OF STATIONARY
SPATIAL PROCESSES

Doctoral Dissertation

Physical Sciences,
Mathematics (01P)

2009 09 29. 14,0 sp. l. Tiražas 20 egz.
Vilniaus Gedimino technikos universiteto leidykla „Technika“,
Saulėtekio al. 11,
10223 Vilnius,
<http://leidykla.vgtu.lt>
Spausdino UAB „Baltijos kopija“,
Kareivių g. 13B, 09109 Vilnius
<http://www.kopija.lt>