



**Vilniaus universitetas  
Duomenų mokslo ir skaitmeninių  
technologijų institutas  
L I E T U V A**



---

INFORMATIKA (N009)

---

**FRAKTALINIŲ BRAUNO LAUKŲ  
TYRIMAS IR TAIKYMAS DAUGIAMAČIŲ  
DUOMENŲ MODELIAVIMUI**

**Neringa Urbonaitė**

2020 m. spalio

Mokslinė ataskaita DMSTI-DS-N009-20-14

VU Duomenų mokslo ir skaitmeninių technologijų institutas, Akademijos g. 4,  
Vilnius LT-08412

[www.mii.lt](http://www.mii.lt)

## **Santrauka**

Duomenų chaotiškumas ir netvarka tampa sudėtinga problema kai norima analizuoti ir prognozuoti ypač jei duomenys daugiamačiai. Daugelis darbų, taikant fraktalinį vektorinį Brauno lauko (fvBl) modelį, atlikta, kai duomenys yra vienmačiai. Šios ataskaitos tikslas - pritaikyti fvBl'o modelį vektoriniams duomenims, ir palyginti fvBl'o paviršių nusakančio Hursto parametro vertinimo metodus. FvBl'o modeliui buvo įvesta atstumų matrica ir nauja kovariacinė funkcija. Modeliui analizuoti atliktas Monte Karlo statistinis eksperimentas: sudarytas fvBl'o generavimo algoritmas ir metodai Hursto parametrai įvertinti. Pasirinkti du parametro vertinimo metodai: didžiausio tikėtimumo ir plačiai naudojamas variogramos. Šių metodų palyginimas, įvertinant Hursto parametrai, leidžia spręsti, kad fvBl'o modelis sudarytas teisingai. Šioje ataskaitoje pateikiamas naujas fvBl daugiamačio atvejo apibrėžimas, su nauja kovariacine funkcija, kuriai nėra svarbus koordinatų pradžios taškas, bei pagrindžiama, kad didžiausio tikėtimumo metodas veikia geriau nei variograma. Atliekamas praktinis didžiausio tikėtimumo metodo taikymas realioms temperatūros ir kritulių duomenims tiriant jų priklausomybes. Tai žingsniai leidžiantys supaprastinti daugiamačių duomenų modeliavimą ir kartu giliau nagrinėti sudėtingus duomenis bei prognozuoti jų anomalijas.

**Reikšminiai žodžiai: Fraktalinis vektorinis Brauno laukas, Didžiausio tikėtimumo metodas, Variograma, Hursto indeksas**

# Turiny

---

1	Įvadas .....	4
2	Apžvalga .....	5
3	Metodai .....	6
4	Fraktalinis Vektorinis Brauno Laukas .....	8
4.1	Kovariacinė matrica .....	8
4.2	Savipanašumas.....	9
4.3	Fraktalinio vektorinio Brauno lauko realizacijų imitavimas .....	9
4.4	Variogramos metodas .....	10
4.5	Didžiausio tikėtinumo metodas .....	10
5	Monte Karlo eksperimento rezultatai.....	11
6	Situacijos analizė (praktinis taikymas) .....	15
7	Išvados .....	16
8	Literatūra.....	17

# 1 Įvadas

Netvarkingas išsibarstymas yra būdinga daugeliui daugiamačių paviršių, kurie gali būti modeliuojami taikant  $f_vBl$  o modelį. Būtent šis modelis pasižymi koreliacija ir toluma priklausomybe nusakoma indeksu ( $H$ ) dar dažnai vadinamu Hursto parametru. Žinant daugiamačių duomenų rinkiniui būdingą Hursto parametą galima nuspėti chaotines duomenų išsidėstymo savybes, o vėliau atlikti prognozę ir interpoliaciją tų taškų, kurie nežinomi. Šioje ataskaitoje atlikta analizė pasitelkiant Hursto parametro vertinimo metodus ir pateiktas pagrįstas  $f_vBl$  o apibrėžimas.

Dauguma tyrimų atliekama kai  $f_vBl$  yra skaliarinis ir vienmatis t.y. fraktalinis Brauno judesio ( $fB_j$ ) atvejis. Šiame darbe siekiama modeliuoti kai erdvės topologinė dimensija didesnė, t. y. ne vienmačiu atveju, o daugiamačiu todėl naudojama sąvoką „laukas“ vietoj „procesas“ ar „judėjimas“. Modeliuojant daugiau dimensijų turinčius duomenis atsiranda daugiau iššūkių. Viena iš problemų, kad  $f_vBl$  o kovariacinė matrica tampa sudėtinga ir ją sudarant svarbu žinoti koordinatinių pradžių tašką, kurį nevisada galima tiksliai įvertinti. Taipogi  $f_vBl$  o Hursto indeksas dažnai nėra žinomas todėl ieškoma metodų kurie šį parametą gali efektyviai nuspėti ir įvertinti.

Šioje ataskaitoje  $H$  parametru vertinti pasirinkti du metodai didžiausio tikėtimumo ir variogramos. Didžiausio tikėtimumo metodas, kuris taikomas ir prognozei sudaryti, dėl savo sudėtingumo, yra rečiau taikomas šiai problemai spręsti. Variogramos metodas plačiai taikomas geostatistikoje. Šis modelis yra vienas iš krigingo modelio dalių, kuri analizuoja duomenų panašumą [1]. Tai svarbi ypatybė, kuria šis modelis remiasi, nes kituose algoritmuose kaip neuroninis tinklas, atraminių vektorių klasifikatorius, rezultatas gaunamas taikant vidinius skaičiavimus, neatskleidžiant detalai duomenų variacijos. Variogramos metodą kaip ir didžiausio tikėtimumo siekiama pritaikyti vertinant Hursto parametą daugiamačiams duomenims.

Ataskaitos tikslas įvesti naują  $f_vBl$  o apibrėžimą pagrindžiant naują kovariacinę funkciją, ir pritaikyti du metodus daugiamačiams skaičiavimams, kurie įvertina  $f_vBl$  o Hursto parametą vektoriniams duomenims.  $f_vBl$  o nauja kovariacijos funkcija, remiasi atstumų matrica paskaičiuota tarp duomenų taškų. Pateikiamas jos matematinis išvedimas. Metodų veikimas tikrintas Monte Karlo eksperimentu. Statistiniam eksperimentui sudarytas metodas, kuris generuoja  $f_vBl$  o daugiamačius duomenis su iš anksto parinktu Hursto parametru. Abu algoritmai veikė tinkamai pagal teorines

prielaidas. Didžiausio tikėtimumo metodas parodė geresnį parametro vertinimo rezultatą nei variograma.

Likusiose šios ataskaitos dalyse atitinkamai detaliai ir nuosekliai pateikiamos notacijos, sprendimai ir atlikti eksperimentų rezultatai. Antrame skyriuje pateikiama detali literatūros apžvalga. Trečias skyrius skirtas metodologijai: Monte Carlo eksperimento šiame darbe iššaiškinimui ir pagrindinei analizei. Ketvirtas skyrius skirtas pateikti svarbias išraiškas, taikomų metodų detalizavimą, apibrėžti pagrindines su fVBl'o modeliu susijusias sąvokas. Penktame skyriuje pateikiami eksperimentų rezultatai. Šeštame skyriuje atliktas praktinis taikymas. Septintame skyriuje pateiktos ataskaitos išvados, apibendrinimai, idėjos tolimesniems darbams taip pat šios ataskaitos pabaigoje pateiktos nuorodos leidžia skaitytojui išstudijuoti daugiau šios iššukių temos aspektų.

## 2 Apžvalga

Modeliuojant išsibarstymą duomenyse fBj, kuri charakterizuoja Hursto indeksas, yra vienas populiariausių modelių [2]. Taikymo sritys apima finansinį modeliavimą [3], hidrodinamiką [4], geoinformatiką [5], signalų apdorojimą [6]. FvBl'o modelis daugiausiai randamas vaizdų analizės darbuose [7, 8, 9], paviršių modeliavime [10, 11].

FvBl'u šioje ataskaitoje modeliuojama kai daugiamačiai duomenys yra erdviniai. Tokius duomenis nagrinėja geostatistikos mokslas, kuris remiasi Gauso atsitiktiniais laukais. FvBl'o modelis priklauso Gauso atsitiktiniam laukams [12]. Duomenų sąryšiai Gauso laukuose modeliuojami atstumų tarp taškų porų matrica ir kovariacijos funkcija arba populiaria variograma [13]. Atstumų matricos detalūs tyrimai buvo atlikti darbuose [14, 15]. Daugelis atliktų modeliavimų taikant fVBl'o modelį remiasi kovariacijų matrica, kuri buvo įvesta A. Kolmogorov'o darbe [16]. Šis fVBl'o modelis tapo žinomas po Mandelbrot'o ir van Ness [17] publikuoto darbo. Nors modelis daugiausiai taikomas kaip fBj, tačiau yra darbų kurie siekė šį modelį taikyti daugiamačių duomenų modeliavimui įvesdami jo apibendrinimą. Vienas iš tokių pristatytų fBj modelių skirtas didesnei topologinei dimensijai buvo aprašytas pasitelkiant spektrą. Darbe teigiama, kad fBj kovariacinė funkcija gali būti gaunama iš stochastinio spektro [32]. Poul Levy [18] fBj apibendrino erdvėje  $\mathfrak{R}^1$  į fraktalinį Brauno lauką erdvėje  $\mathfrak{R}^d$ . Pastarasis modelis taip pat apibendrintas šių dienų darbe [19]. Šis

modelis po to pavadintas Levio Brauno judesiu. Daugiamačius duomenis modeliuojant kovariacinė funkcija paprastai išreiškiama su euklidine norma. Neeuklidinėmis normomis paremtas apibendrinimas atliktas [20].

Taikant  $f_{vBl}$  daugiamatims duomenims yra svarbu taip pat atsižvelgti į svarbią šiam modeliuoj būdingą dalį - Hursto indeksą, kuris yra tarp 0 ir 1. Hursto parametras parodo duomenų koreliaciją. Kuo parametro reišmė mažesnė tuo paviršius daugiau chaotiškas.

Svarbu šį parametras tinkamai įvertinti. Yra klasikinis R/S statistikos metodas nagrinėtas [21, 22]. Taip pat dažnai taikomas variogramos metodas, kurio veikimas pagrįstas ir šiame darbe. Šie metodai paremti geografiniais sprendimais, todėl nėra itin tikslūs. Ieškoma tikslesnių metodų, kaip didžiausio tikėtumo metodas, kuris pasižymi asimptotinėmis statistinėmis savybėmis [23]. Daugiausiai darbuose įvertinant parametras  $H$  taikomas  $f_{Bj}$  ne lauko modelis [24]. Šis parametras dažnai taikomas kaip laiko eilučių klasifikatorius ir yra aktualus modeliuojant ilgą atminties priklausomybę [25, 26] ir savipanašumą [27].

Teoriniai modelį sudarantys aspektai kaip kovariacijos funkcija, atstumų matrica ir Hursto parametras yra pagrindinės dalys nusakančios duomenų ryšius šiame modelyje. Modeliuojant daugiau dimensijų turinčius duomenis modelis sudėtingėja, todėl  $f_{vBl}$  apibendrinimo ir Hursto parametro vertinimo darbai svarbūs aiškinantis modelio galimybes modeliuojant erdvinius duomenis.

### 3 Metodai

Šiuo tyrimu siekiama įvesti naują  $f_{vBl}$  apibrėžimą ir  $f_{vBl}$  analizei pritaikyti Hursto parametro vertinimo didžiausio tikėtumo ir variogramos metodus daugiamatims duomenims. Dažnai daugiamatiai duomenys yra suprantami kaip erdviniai duomenys kurie vieni su kitais yra susiję. Darbe parodoma kad įvedus naują kovariacijos funkciją su atstumų matrica ji teisingai apibrėžia  $f_{vBl}$  modelį.

Toliau sudaryti didžiausio tikėtumo ir variogramos metodai Hursto parametrai daugiamatims  $f_{vBl}$  duomenims įvertinti. Vienas iš galimų būdų patikrinti sudarytų metodų veikimą yra palankios situacijos modeliams sudarymas. Tam atliekamas statistinis Monte Karlo eksperimentas. Monte Karlo metodas yra vienas iš dažniausiai taikomų statistinių modeliavimų būdų. Metodas yra stochastinis, t. y. naudoja atsitiktinius generuotus skaičius [28]. Taikant šį metodą duomenų imties skirstiniai gali

būti aproksimuoti normaliuoju dėsnio remiantis centrine ribine teorema. Įvertintų parametrų paklaida priklauso nuo duomenų imties dydžio, t.y. kuo imtis didesnė tuo rezultatai artėja prie tikrųjų reikšmių [29]. Įvardijant, kas sudaro statistinį Monte Karlo modeliavimą, tai, atstiktinių dydžių pasiskirsčiusių pagal tam tikrą skirstinį generavimas (šiam darbe taikomas normalusis skirstinys), vertinimo metodų sudarymas, sudarytų metodų statistinė analizė.

Šioje ataskaitoje eksperimento užduotis su sukurtu fVBI'o realizacijų imitavimo metodu generuoti fVBI'o duomenis su iš anksto parinktu  $H$  parametru ir taikant sudarytus vertinimo metodus – didžiausio tikėtinumo ir variogramą – įvertinti  $H$  parametras kiekvienai generuotai duomenų grupei, skaičiuoti atstiktinių grupių vidurkius ir standartinius nuokrypius; rezultate palyginti įvertinimus su realiu Hursto parametru. Įgyvendinti šį Monte Karlo statistinį eksperimentą sudarytas fVBI'o generavimo algoritmas su iš anksto parenkamu Hursto indeksu, kuris remiasi naujai įvesta fVBI'o kovariacijų matricos išraiška. Hursto indeksas  $H \in [0,1)$ , kuris lemia fVBI'o koreliacinę struktūrą, eksperimentinėje procedūroje nustatomas atitinkamai. Generavimui pasirinktas Choleckio algoritmas, kuris remiasi Choleckio dekompozicija [30, 31]. Nors šio metodo sudėtingumas  $O(N^3)$ , tačiau anksčiau atlikti eksperimentai rodo kad algoritmas generuoja tikslius rezultatus.

Pateikiami Monte Karlo eksperimento procedūros šiame darbe etapai:

- 1) *su iš anksto parenkamu Hursto indeksu  $H = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$ . generuotos trys duomenų grupės, po 100, kai  $k = 20; 50; 100$  (eilučių skaičius). Lokacijų stulpelių du  $m = 2$ , požymių stulpeliai  $p = 2; 5; 10$ ;*
- 2) *kiekvienai duomenų grupei taikyti variogramos ir didžiausio tikėtinumo metodai ir įvertintinami Hursto parametrai kai  $p = 2$  ; kai  $p = 5; 10$ ; modeliuota tik taikant didžiausio tikėtinumo metodą;*
- 3) *gautiems parametrų įverčiams skaičiuoti vidurkiai, standartiniai nuokrypiai.*
- 4) *parametrų įverčių vidurkiai lyginti su iš anksto parinktomis  $H$  reikšmėmis;*
- 5) *eksperimento procedūra kartojama  $K = 100$  kartų, kiekvienai duomenų grupei  $k$  atskirai.*

Eksperimentai buvo atlikti taikant R programinę įrangą kompiuteryje kurio procesoriaus pagrindiniai parametrai Intel(R) CORE(TM) i7-2600 CPU@3.40GHZ ir atmintis 8GB.

Jei algoritmai – variograma, didžiausio tikėtinumo metodas - tinkamai veikė fvBl' o duomenims, tai realiems duomenims metodai gebės nuspėti, ar tai fvBl, ir įvertinti jį nusakantį parametą  $H$ . Eksperimentas parodys, kad didžiausio tikėtinumo metodas veikia taip pat gerai kaip ir variograma ir pagrįs sukurto naujo fvBl' o modelio korektiškumą.

## 4 Fraktalinis Vektorinis Brauno Laukas

Šioje dalyje pirmiausia apibrėžiamas fvBl' o modelis jo pagrindimas. Pateikiamas fvBl' o generavimas daugiamatėje erdvėje. Sistemiskai sudaromi Hursto parametro vertinimo metodai.

### 4.1 Kovariacinė matrica

FvBl' a charakterizuoja kovariacijų matricą ir Hursto parametras. Tarkime, turime  $k$  taškų  $X = (x_1, x_2, \dots, x_k)$ ,  $X \in \mathfrak{R}^d$ ,  $d \geq 1$ ,  $1 \leq i \leq k$ . Dažniausiai tai yra dvimačiai taškai aprašomi dekartio ar geofinėmis koordinatėmis  $(x, y)$ . Kiekvienoje šių erdvės lokacijų vyksta procesai. Atlikus stebėjimus juose gaunamos reikšmės, kurios sudarant modelį atitinka vektorines tikslo funkcijos reikšmes  $Z = (z_1, z_2, \dots, z_k)$ ,  $Z \in \mathfrak{R}^m$ .

Pateikiamas fvBl' o apibrėžimas pagal A. Kolmogorovą.

Tarkime, turime duomenų rinkinį  $X = (x_1, x_2, \dots, x_k)$ ,  $X \in \mathfrak{R}^d$ ,  $d \geq 1$ ,  $1 \leq i \leq k$ . Gauso atsitiktinis laukas  $Z(X)$ , kurio kovariacinė funkcija  $C(x_1, x_2)$  yra fvBl:

$$C(x_1, x_2) = \frac{1}{2} (|x_1|^{2H} + |x_2|^{2H} - |x_1 - x_2|^{2H}). \quad (1)$$

Formulę (1) taip pat atitinka  $E(|x_1 - x_2|^2) = |x_1 - x_2|^{2H}$ .

Funkcija  $C(x_1, x_2)$  yra neneigiamai apibrėžta kiekvienam  $k \geq 1$ .

Toliau pateikiamas fvBl' o apibrėžimas, kuris remiasi atstumų matrica (2). Atstumai apskaičiuojami tarp  $X$  rinkinio vektorių porų, taip gaunama atstumų matrica  $A = A_{i,j}$  (1)

$$A(X) = \left( (x_i - x_j)^T \cdot (x_i - x_j) \right)^H \quad (2)$$

$x_i \in R^d$ ,  $x_i \neq x_j$ ,  $i \neq j$ ,  $1 \leq i, j \leq k$ , ir  $H \in [0,1)$ . Pažymimas  $k$ -matis vektorius -  $\mathbf{1} = (1,1,\dots,1)^T$



Apibrėžiama nauja fVBl'o kovariacijų matricos išraišką (3).

$$B_H(X) = 2 \frac{\mathbf{1}\mathbf{1}^T}{\mathbf{1}^T A^{-1} \mathbf{1}} - A \quad (3)$$

Analizuojant atsitiktinius laukus svarbios stacionarumo, arba homogeniškumo sąvokos, todėl apibrėžtas homogeninis fVBl, kurio vidurkių matrica, lygi

$$EB_H(x) = \mu \cdot \mathbf{1}.$$

FBj vadinamas fVBl  $B_H$  erdvėje  $R^d$  pasižymi charakteristikomis:

(1) Kiekvienam  $x_1, x_2, \dots, x_n \in R^d$ , atsitiktinis vektorius atitinkamas  $(B_H(x_1), B_H(x_2), \dots, B_H(x_n))$ ;

(2)  $B_H(0) = 0, EB_H(0) = 0$ ;

(3)  $E\{|B_H(x) - B_H(x - \Delta)|^2\} = \sigma^2 |\Delta|^{2H}$

(4)  $B_H(x)$  turi Gauso pasiskirstymą.

## 4.2 Savipanašumas

Statistinis savipanašumas pasireiškia keičiant proceso  $B_H$  mastelį: kiekvienam  $\lambda > 0$ ,  $E\{B_H(\lambda x)B_H(\lambda x')\} = E\{\lambda^H B_H(x)\lambda^H B_H(x')\}$ . Kadangi  $B_H(\lambda x)$  ir  $\lambda^H B_H(x)$  yra Gauso procesai turintys tą patį vidurkį ir kovariaciją, jie turi ir vienodą pasiskirstymą

$$B_H(\lambda x) \equiv \lambda^H B_H(x)$$

## 4.3 Fraktalinio vektorinio Brauno lauko realizacijų imitavimas

Pristatomas fVBl'o generavimo metodas. Kadangi Brauno Laukas kaip ir Brauno paviršius atitinka Gauso procesą su kovariacijų matrica galima taikyti Choleckio metodą. Šis algoritmas remiasi fVBl'o kovariacijų matricos Choleckio dekompozicija. Tikslas generuoti fVBl'ą su iš anksto nustatytu Hursto parametru.

Pirmiausia parenkamas Hursto parametras. Paskaičiuojama pagal (2) fVBl'o matrica  $B_H(X)$  (3). Toliau generuojamos atsitiktinių dydžių, tolygiai pasiskirsčiusių atkarpoje  $[0,1]$ , realizacijos,  $\vartheta = (\vartheta_1, \vartheta_2, \dots, \vartheta_k)$ . Įvedama matrica  $b$

$$b = \begin{pmatrix} 1.1 & -0.1 \\ 0.7 & 1.1 \end{pmatrix}.$$

Kadangi matricos  $B, b$  simetrinės ir teigiamai apibrėžtos galima taikyti Choleckio dekompoziciją taip perskaičiuojant matricas į trikampį pavidalą.

$$\beta = chol(b) \quad (4)$$

$$C = chol(B_H(X)) \quad (5)$$

Taikant matricas  $\beta, C$  ir atsitiktinių dydžių vektorių  $\vartheta$  pagal (5) paskaičiuojamos fVBl'o realizacijos.

$$Z = \beta\vartheta C^T \quad (6)$$

Atlikus  $p$  perskaičiavimų gaunama atsitiktinių dydžių matrica  $Z, i = 1, 2, \dots, p; j = 1, 2, \dots, m$  (6). taškuose  $X$ . Atitinkamai  $Z$  yra nepriklausomi Gauso procesai, kurie sudaro fVBl'ą.

#### 4.4 Variogramos metodas

Vienas iš metodų įvertinti parametą  $H$  yra taikant variogramą. Tai fundamentalus modelis skirtas erdviųjų porcesų modeliavimui. Geostatistikos darbuose žinoma, kad variograma iš atsitiktinės funkcijos gali būti suskaidyta į atskirus taškus.

Sudarant variogramos metodą skaičiuojama empirinė variograma ir parametrinis variogramos modelis.

$$\begin{aligned}
 V(H) &= \frac{2}{k(k-1)} \left( \sum_{i=1}^{k-1} \sum_{j=i}^k \left( (z_i - z_j)^T (z_i - z_j) \right)^2 \right) \\
 &= \frac{\sum_{i=1}^m \sum_{j=1}^m \left( \sum_{i=1}^{k-1} \sum_{j=i}^k (z_{ii,i} - z_{ii,j})(z_{jj,i} - z_{jj,j}) \left( (x_i - x_j)^T \cdot (x_i - x_j) \right)^H \right)^2}{\sum_{i=1}^{K-1} \sum_{j=i}^K \left( (x_i - x_j)^T \cdot (x_i - x_j) \right)^{2H}} \quad (9)
 \end{aligned}$$

Iš formulės (9) gaunama variogramos funkcija, ją minimizavus įvertinama optimali parametro  $H$  reikšmė.

Variogramos metodas remiasi tik atstumų matrica, jam nereikia kovariacinės funkcijos, pati variograma atspindi duomenų priklausomybę.

#### 4.5 Didžiausio tikėtinumo metodas

Didžiausio tikėtinumo metodas leidžia apskaičiuoti fVBl'o parametro  $H$  reikšmę, kuri minimizuoja tikėtinumo funkciją.

Sudarant tikėtinumo funkciją į formulę (7) įstatome kovariacijų matricą (3)

$$L = \frac{1}{k} \left( ZB_H^{-1}Z^T - \frac{(ZB_H^{-1}\mathbf{1})(ZB_H^{-1}\mathbf{1})^T}{\mathbf{1}^TB_H^{-1}\mathbf{1}} \right). \quad (7)$$

Toliau (7) išraiška ir fvBl'o kovariacijų matrica (3) įstatoma į didžiausio tikėtinumo lygtį (DT) (8),  $m$  – nurodo požymių skaičių.

$$DT(H) = \frac{1}{m} \ln(|L|) + \frac{1}{k} \ln(|B_H|). \quad (8)$$

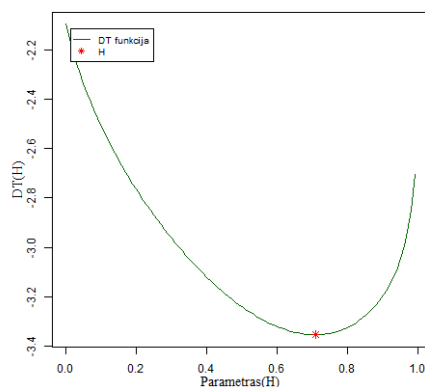
Iš formulės (8) gaunama didžiausio tikėtinumo funkcija, ją minimizavus įvertinamas Hursto indeksas.

Taikant Didžiausio tikėtinumo metodą įvertinamos tokios parametrų reikšmės, su kuriomis gaunami rezultatai yra labiausiai tikėtini duotajam modeliui. Tokiu būdu siekiama įvertinti Hursto parametą.

## 5 Monte Karlo eksperimento rezultatai

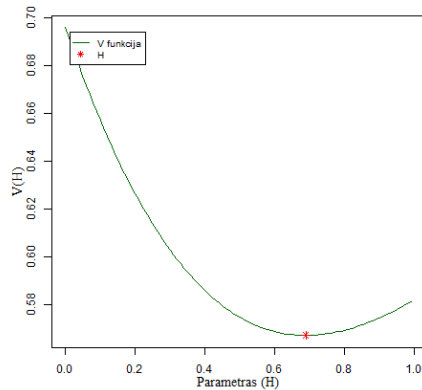
Šioje dalyje pristatomi rezultatai, kurie buvo gauti atlikus Monte Karlo statistinį eksperimentą.

Pateikiami pavyzdžiai didžiausio tikėtinumo ir variogramos funkcijų, gautų Monte Karlo eksperimento metu modeliuojant duomenims, kurie buvo generuoti kai  $H = 0.7$ ,  $k = 100$ ,  $m = 2$ ,  $p = 2$ .



pav. 1 DT funkcija. DT funkcijos minimali reišmė

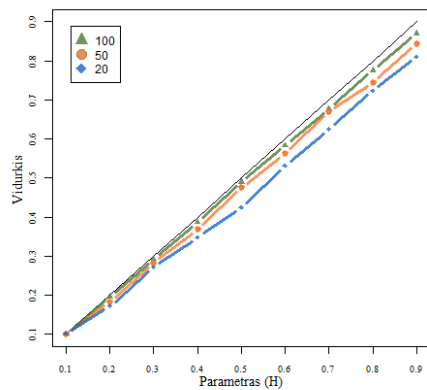
Didžiausio tikėtimumo funkciją vaizduojamą (pav. 1) minimizavus buvo gautas  $H$  įvertis 0.74, kuris sutampa su anksčiau parinktu  $H$  indeksu.



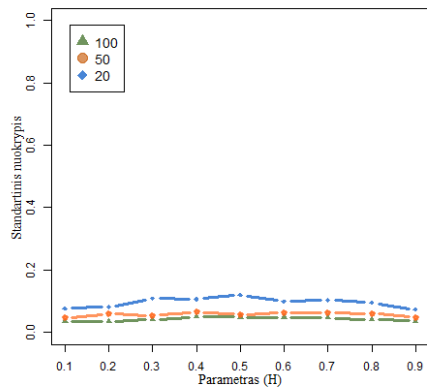
pav. 2 V funkcija. V funkcijos minimali reišmė

Variogramos funkciją vaizduojamą (pav. 2) minimizavus buvo gautas  $H$  įvertis 0.71, kuris sutampa su anksčiau parinktu  $H$  indeksu. Šiuo atveju abu metodai panašiai įvertino Hursto parametą.

Toliau pristatomi rezultatai visų generuotų duomenų, kai  $p = 2$ ;  $k = 20, 50, 100$ . Sugeneravus fvBl'o duomenis ( $k$ ), su iš anksto parinktu Hursto parametru, vertinamas  $H$  įvertis taikant didžiausio tikėtimumo metodą. Įvertintiems įverčiams  $H$  skaičiuojamas vidurkis ( $\mu_{DT}$ ) ir standartinis nuokrypis ( $\sigma_{DT}$ ).



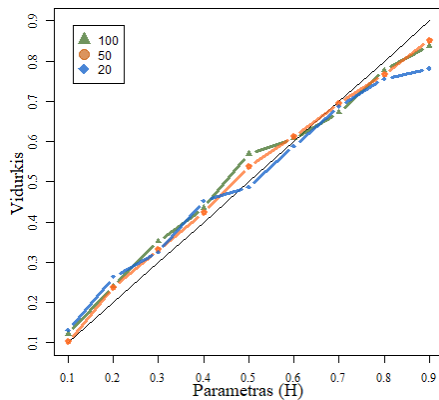
pav. 3 Didžiausio tikėtimumo (DT) metodu įvertintų parametų  $\mu_{DT}$



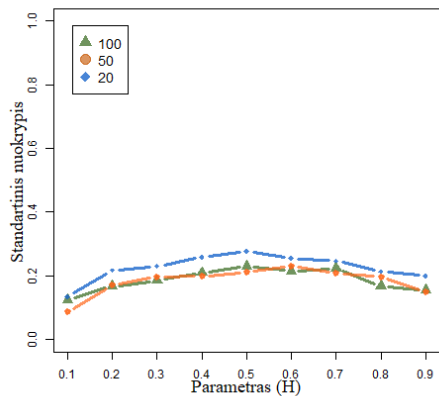
pav. 4 Didžiausio tikėtimumo metodu įvertintų parametru  $\sigma_{DT}$

Paveikslėliuose (pav. 1, 2) vazduojami taikant didžiausio tikėtimumo metodą gauti rezultatai. Kaip ir būdinga Monte Karlo metodui, ku duomenų imtis didesnė  $k$  tuo įverčių rezultatai tikslesni.

Toliau atliekamas Hursto parametro vertinimas taikant variogramos metodą. Paskaičiuojamas įvertintų parametru vidurkis ir standartinis nuokrypis. Šie rezultatai lyginami su iš anksto parinktais parametrais.



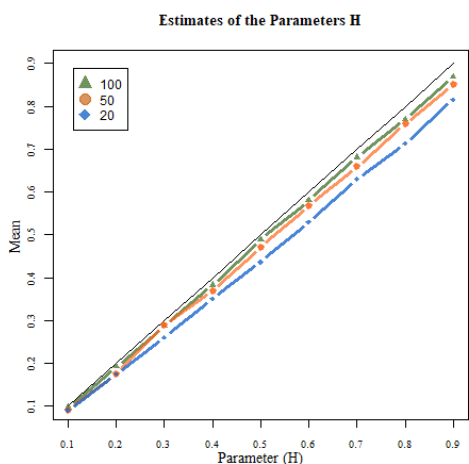
pav. 5 Variogramos (V) metodu įvertintų parametru  $\mu_V$



pav. 6 Variogramos (V) metodu įvertintų parametru  $\sigma_V$

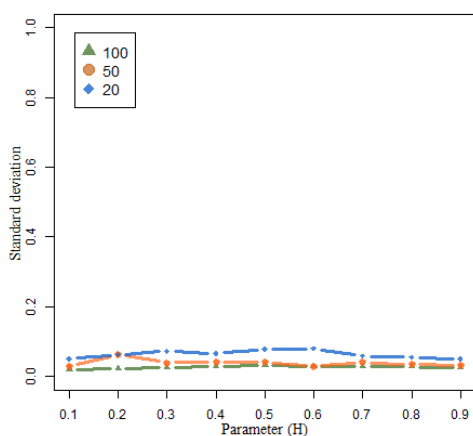
Paveikslėliuose (pav. 3, 4) vazduojami taikant variogramos metodą gauti rezultatai vidurkiai ( $\mu_V$ ) ir standartiniai nuokrypiai ( $\sigma_V$ ). Variogramos metodo vertinimo rezultatai daugiau išsibarstę apie vidurkį lyginant su didžiausio tikėtumo rezultatais. Tęsiant eksperimentus ir didinant požymių stulpelių skaičių modeliavimas nebeatliekamas su Variogramos modelių.

Rezultatai visų generuotų duomenų, kai  $p = 5$ ;  $k = 20, 50, 100$ .



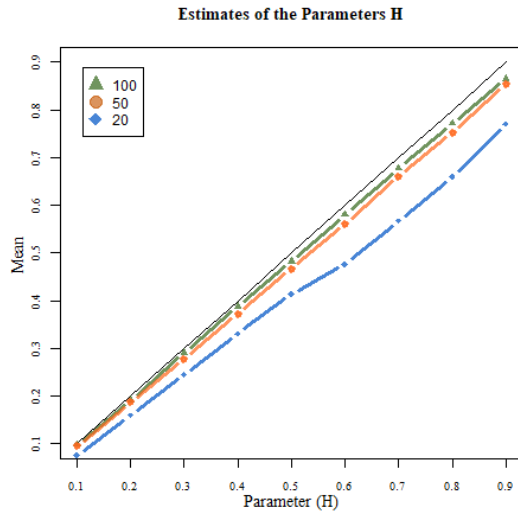
pav. 7 Didžiausio tikėtumo (DT) metodu įvertintų parametru  $\mu_{DT}$

Padidinus požymių skaičių iki  $p = 5$  didžiausio tikėtumo metodas įvertino Hursto parametrus.



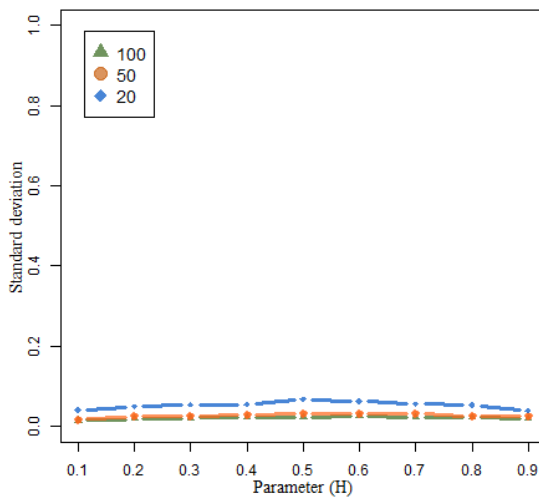
pav. 8 Didžiausio tikėtumo metodu įvertintų parametru  $\sigma_{DT}$

Rezultatai visų generuotų duomenų, kai  $p = 10$ ;  $k = 20, 50, 100$ .



pav. 9 Didžiausio tikėtumo (DT) metodu įvertintų parametru  $\mu_{DT}$

Didžiausio tikėtinumo metodas prasčiau veikė su mažesne imtimi duomenų, kaip  $k = 20$ .



pav. 10 Didžiausio tikėtinumo metodu įvertintų parametru  $\sigma_{DT}$

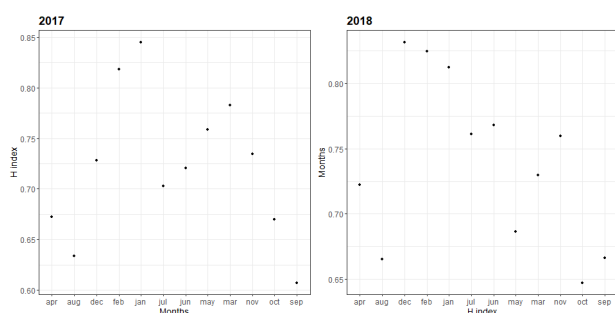
Įvertintų parametru standartinis nuokrypis išliko mažas, kaip ir su mažiau požymių reikšmių turinčiais duomenimis.

Eksperimentas parodė, tiek variogramos, tiek didžiausio tikėtinumo metodo rezultatai atitiko teorines prielaidas. Didžiausio tikėtinumo metodas veikė geriau nei variograma.

## 6 Situacijos analizė (praktinis taikymas)

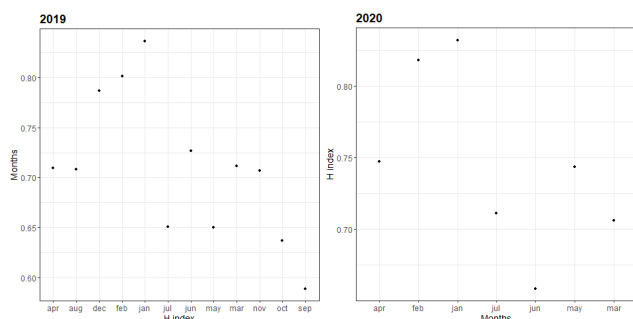
“Global Climate Monitor” tai priemonė skirta visuotiniams klimato duomenims. Tai Sevilijos universiteto klimato tyrimų grupės sukurtas projektas. Pasaulinio klimato įrankyje pateikiama nuo 1901 m. iki šių dienų klimato informacija.

Buvo pasirinkti skirtingų metų (2017, 2018, 2019, 2020), kiekvieno mėnesio Lietuvos teritorijos kritulių ir temperatūros duomenys. Lokacijas sudarė du stulpeliai žymėti Dekarto koordinatėmis. Požymius sudarė taip pat du stulpeliai, kurie atitiko temperatūrą ir kritulius. Duomenims Hursto parametrui įvertinti buvo taikytas didžiausio tikėtimumo metodas.



pav. 11 Įvertinti H parametrai, kiekvienam 2017 metų mėnesiui atskirai; pav. 12 Įvertinti H parametrai, kiekvienam 2018 metų mėnesiui atskirai

Rezultatai (pav. 11, 12, 13, 14) parodo, kad vieni mėnesiai yra labiau nuspėjami negu kiti. Visuose metuose žiemos mėnesiai pasiekia 0.8-0.9 Hursto parametą.



pav. 13 Įvertinti H parametrai, kiekvienam 2019 metų mėnesiui atskirai; pav. 14 Įvertinti H parametrai, kiekvienam 2020 metų mėnesiui atskirai

Žvelgiant į rezultatus matyti, kad temperatūra ir krituliai priklausomi vienas nuo kito ir yra determininuoti.

## 7 Išvados

Erdvinių duomenų išsibarstymą galima modeliuoti taikant kovariacinę funkciją ir įvertinant Hursto parametą. Šiomis ypatybėmis pasižymi fvBl'o modelis. Šiame darbe įvestas naujas fvBl'o apibrėžimas su nauja kovariacine funkcija. Šis žingsnis



palengvina netvarkingo duomenų išsidėstymo modeliavimą taikant fvBl'ą, nes priešingai nei taikant klasikinį fvBl'o apibrėžimą pagal Kolmogorovą, šiam modeliui nereikia nurodyti koordinacių pradžios taško.

Toliau siekiant įvertinti daugiamačių duomenų, kurie sudaro fvBl'ą, Hursto parametras, buvo sudaryti du vertinimo metodai – didžiausio tikėtimumo ir variogramos. Vertinimo metodams palyginti buvo atliktas statistinis Monte Karlo eksperimentas. Jo metu sukurtas naujas fvBl'o generavimo algoritmas kuris remiasi šiame straipsnyje įvestu nauju fvBl'o apibrėžimu. Generavimo algoritmas buvo naudotas Monte Karlo eksperimento metu. Šio eksperimento realizavimui imitavimo metodas generavo duomenis su iš anksto parenkamu Hursto parametru. Tai leido generuoti patogią aplinką sudarytiems vertinimo metodams – didžiausio tikėtimumo ir variogramos – ir palyginti juos. Šie metodai įvertinus Hursto parametras leido lengvai nuspėti duomenų koreliaciją, išsidėstymą. Monte Karlo statistiniu eksperimentu buvo parodyta, kad didžiausio tikėtimumo metodas veikė geriau nei variogramos - rezultatų išsibarstymas apie vidurkį buvo mažesnis.

Siekiant pagrįsti metodo pritaikomumą, metodu modeliuoti buvo skirtingų metų temperatūros ir kritulių Lietuvos domenys. Pastebėta, kad pagal atitinkamą kritulių kiekį galima nuspėti temperatūrą ir atvirkščiai. Daugelyje mėnesių Hursto parametro įvertis buvo aukštesnis už 0.5.

Naujo fvBl apibrėžimo įvedimas; Hursto parametro vertinimo metodų - didžiausio tikėtimumo ir variogramos - pritaikymas daugiamačiams vektoriniams duomenims leidžia paprasčiau įvertinti išsidėstymą daugiamačių erdvinių duomenų.

## 8 Literatūra

- [1] N. Cressie, „The origins of kriging,“ *Mathematical Geology*, t. 22, pp. 239-252, 1990.
- [2] M. Fernandez-Martinez, M. Sanchez-Granero, J. T. Segovia ir I. Roman-Sanchez, „An accurate algorithm to calculate the Hurst exponent of self-similar processes,“ *Physics Letters A*, t. 378, nr. 32-33, pp. 2355-2362, 2014.
- [3] M. M. Garcia, A. C. M. Pereira, J. L. Acebal ir A. R. B. de Magalhaes, „Forecast model for financial time series: An approach based on harmonic oscillators,“ *Physica A: Statistical Mechanics and its Applications*, t. 549, pp. 124-365, 2020.
- [4] B. Mandelbrot, „Une classe de processus homothétiques a soi; Application a la loi climatologique de H. E. Hurst,“ *Comptes Rendus de L'Academie des Sciences de Paris*, t. 260, pp. 3274-3277, 1965.

- [5] Q. Rufu, L. Liangzhao ir K. Cuiping, „A GIS-based software for forecasting pollutant drift on coastal water surfaces using fractional Brownian motion,“ *Environmental Modelling and Software*, t. 92, pp. 252-260, 2017.
- [6] S. Chang, „Fractional Brownian motion in biomedical signal processing, physiology, and modern physics,“ *5th International Conference on Bioinformatics and Biomedical Engineering*, pp. 1-4, 2011.
- [7] P. Tafti, R. Delgado-Gonzalo, A. Stalder ir M. Unser, „Fractal modelling and analysis of flow-field images,“ *IEEE International Symposium on Biomedical Imaging: from nano to macro*, pp. 49-52, 2010.
- [8] F. Richard, „Tests of isotropy for rough textures of trended images,“ *Statistica Sinica*, t. 26, nr. 3, pp. 1279-1304, 2014.
- [9] P. Hailang, Z. Weicheng, J. Wen, P. Wang, J. S. Yang ir X. Zhang, „Roughness change analysis of sea surface from visible images by fractals,“ *IEEE Access*, t. 8, pp. 78519-78529, 2020.
- [10] N. Zhang ir D. W. Apley, „Fractional Brownian fields for response surface,“ *Journal of Quality Technology*, t. 46, nr. 4, pp. 285-301, 2014.
- [11] M. L. Stein, „Fast and exact simulation of fractional Brownian surfaces,“ *Journal of Computational and Graphical Statistics*, t. 11, nr. 3, pp. 587-599, 2002.
- [12] X. Emery, „Conditioning simulations of Gaussian random fields by ordinary kriging,“ *Mathematical Geology*, t. 39, nr. 6, pp. 607-623, 2007.
- [13] T. Gneiting, Z. Sasvari ir M. Schlathaer, „Analogies and correspondences between variograms and covariance functions,“ *Advances in Applied Probability*, t. 33, nr. 3, pp. 617-630, 2001.
- [14] N. Pozniak, L. Sakalauskas ir L. Saltyte, „Kriging model with fractional euclidean distance matrices,“ *Informatica*, t. 30, pp. 367-390, 2019.
- [15] N. Pozniak ir L. Sakalauskas, „Fractional euclidean distance matrices extrapolator for scattered data,“ *Jaunuju Mokslininku Darbai*, pp. 56-61, 2017.
- [16] A. Kolmogoroff, „Wiensche spiralen und einige andere interessante Kurven im Hilbertschen Raum,“ *C. R. (Doklady) Acad. Sci. URSS*, t. 26, nr. 2, pp. 115-118, 1940.
- [17] B. Mandelbrot ir V. Ness, „Fractional Brownian motions, fractional noises and applications,“ *SIAM Review*, t. 10, nr. 4, pp. 422-437, 1968.
- [18] P. Levy, *Processus stochastiques et mouvement brownien*, Paris: Gauthier-Villars & Cie, 1965.
- [19] Z. Gelbaum, „Fractional Brownian fields over manifolds,“ *Transactions of the American Mathematical Society*, t. 366, nr. 9, pp. 4781-4814, 2013.
- [20] I. Molchanov ir K. Ralchenko, „A generalisation of the fractional Brownian field based on non-Euclidean norms,“ *Journal of Mathematical Analysis and Applications*, t. 430, nr. 1, pp. 262-278, 2015.
- [21] B. B. Mandelbrot ir J. R. Wallis, „Robustness of the rescaled range R/S in the measurement of noncyclic long run statistical dependence,“ *Water Resources Research*, t. 5, p. 967, 1969.
- [22] M. S. Raimundo ir J. Okamoto, „Application of Hurst exponent (H) and the R/S analysis in the classification of FOREX securities,“ *International Journal of Modeling and Optimization*, t. 8, nr. 2, pp. 116-124, 2018.

- [23] L. Hai ir D. Hong, „Application of wavelet-based maximum likelihood estimator in measuring market risk for fossil fuel,“ *Sustainability*, t. 11, nr. 10, pp. 125-134, 2019.
- [24] P. Fieguth ir A. Willsky, „Fractal estimation using models on multiscale trees,“ *IEEE Transactions on Signal Processing*, t. 44, nr. 5, pp. 1297-1300, 1996.
- [25] M. Li, P. Zhang ir J. Leng, „Improving autocorrelation regressio for the Hurst parameter estimation of long-range dependent time series based on golden section search method,“ *Physica A*, t. 445, pp. 189-199, 2016.
- [26] C. Stolojescu ir A. Isar, „A comparison of some Hurst parameter estimators,“ *13th International Conference on Optimization of Electrical and Electronic Equipment (OPTIM)*, pp. 1152-1157, 2012.
- [27] P. Inacio, B. Lakic, M. Freire, M. Pereira ir P. Monteiro, „The design and evaluation of the simple self-similar sequences generator,“ *Information Sciences*, t. 179, nr. 23, pp. 4029-4045, 2009.
- [28] C. Graham ir D. Talay, *Stochastic Simulation and Monte Carlo Methods*, Berlin: Springer, 2013.
- [29] L. Sakalauskas, „Towards implementable nonlinear stochastic,“ *Coping with Uncertainty*, t. 581, pp. 257-279, 2006.
- [30] F. Yarlikaya-Ozkurt, C. Vardar-Acar, Y. Yolcu-Okur ir G.-W. Weber, „Estimation of the Hurst parameter for fractional Brownian motion using the CMARS method,“ *Journal of Computational and Applied*, t. 259, nr. Part B, pp. 843-850, 2014.
- [31] A. Brouste, J. Istas ir S. Lambert-Lacroix, „Conditional fractional Gaussian fields,“ *The R Journal V*, t. 8, 2016.
- [32] I. S. Reed, P. C. Lee, T. K. Truong, Spectral representation of Fractional Brownian Motion in n dimensions and its properties, *Transactions on Information Theory*, nr. 5, 1995