# Ataskaita už 2021 doktorantūros metus

Doktorantė: Neringa Urbonaitė
Vadovas: Prof. habil. dr. Leonidas Sakalauskas

Vilniaus Universitetas
Duomenų mokslo ir skaitmeninių technologijų institutas

2021 rugsėjo 30 d.

Disertacijos tema:
**Fraktalinio Brauno Lauko tyrimas ir taikymas daugiamačių duomenų modeliavime**

Vadovas:
Prof. habil. dr. Leonidas Sakalauskas

Pradžios pabaigos metai:
2019-2023

# Disertacijos rengimo planas

| Studijų metai | Egzaminai | | Konferencijos | | Publikacijos |
|---|---|---|---|---|---|
| - | Planas | Įvykdyta | Planas | Įvykdyta | Planas |
| I 2019-2020 | 2 | 2 | 1 | 1 | – |
| II 2020-2021 | **2** | **2** | 0 | **1** | 1 |
| III 2021-2022 | – | – | 1 | – | – |
| IV 2022-2023 | – | – | 1 | – | 1 |
| Iš viso | 4 | 4 | 2 | 1 | 2 |

lentelė 1: Disertacijos rengimo planas

Einamieji studijų metai (III: 2021/2022).

| Egzaminai | | Konferencijos | | Publikacijos | |
|---|---|---|---|---|---|
| Planas | ¿Ivykdyta | Planas | ¿Ivykdyta | Planas | ¿Ivykdyta |
| Fundamentalieji informatikos ir informatiko inžinerijos metodai | Išlaikyta | DAMSS'2019 | | | |
| Gilieji neuroniniai tinklai | $2021-06$ Išlaikyta | ASMDA 2021 birželio 1- 4 d. Conference of the Applied Stochastic Models and Data Analysis | Investigation of fractional Brownian fields | $1$ | $1$ |

lentelė 2: Disertacijos rengimo planas

# Mokslinių tyrimų ir disertacijos rengimo etapai

| | Darbo pavadinimas | Atlikimo terminai | Pastabos |
|---|---|---|---|
| 1. | Mokslinių tyrimų disertacijos tema apžvalga ir analizė (Lietuvoje ir užsienyje):<br>1. Atlikti daugiamačių Krigingo modelių apžvalgą.<br>2. Nustatyti (identifikuoti) mokslines problemas, kylančias uždaviniuose. | 2019 m. spalio mėn. – 2020 m. rugsėjo mėn. | Parengta mokslinė literatūros apžvalga. |
| 2. | **Mokslinio tyrimo vykdymas:**<br>2.1. Tyrimo metodikos sudarymas;<br>2.1.1Tyrimo metodikos iškeltam uždaviniui spręsti parinkimas;<br>2.1.2 Teorinio ir empirinio tyrimų suplanavimas pagal pasirinktą metodiką.<br>**2.2. Teorinis tyrimas:**<br>2.2.1 Fraktalinių Brauno laukų (FBL) savybių tyrimas;<br>2.2.2 FBL realizacijų kompiuterinio imitavimo metodo sukūrimas.<br>2.3.1 FBL pritaikymas daugiamačiam krigingui ir ekstrepoliavimui;<br>2.3.2 Atstikitinių laukų parametrų vertinimo metodų: semivariogramos ir didžiausio tikėtinumo algoritmų sukūrimas.<br>2.3.3 Kompiuterinių eksperimentų atlikimas generuojant FBL realizaciją imtis ir palyginant jų parametrų vertinimo rezultatus su kitais atsitiktiniai laukais; | 2020 m. spalio mėn. – 2021 m. balandžio mėn.<br><br>2020 m. gegužės mėn. – 2021 m. rugsėjo mėn.<br><br>2021 m. lapkričio mėn. – 2022 m. rugsėjo mėn. | Parengta FBL modelių apžvalga.<br><br>Sudarytas Fraktalinis vektorinis Brauno lauko (FvBl) modelis.<br><br>Sudarytas FvBl'o realizacijų generavimo metodas.<br><br>Sudaryti vertinimo metodai: didžiausio tikėtinumo ir variogramos.<br><br>Atliktas Monte Carlo eksperimentas.<br><br>Sudarytas krigingo metodas su FvBl'o kovariacijų matrica. |

# Disertacijos uždaviniai

**Tyrimo objektas:** daugiamačiai duomenų modeliavimo metodai, skirti vertinti, modeliuoti, generuoti fraktalinį vektorinį Brauno lauką.

**Tikslas:** Algoritmų kūrimas, skirtų daugiamačiams duomenims ir jų sąsajoms vertinti, generuoti, prognozuoti, kai yra stebėjimų trūkumas, taikant fraktalinį vektorinį Brauno lauko modelį.

**Uždaviniai:**

1. Sukurti FVBL'o modelį;
2. Sudaryti FVBL'o realizacijų generavimo algoritmą;
3. Sudaryti FVBL'o vertinimo algoritmą taikant didžiausio tikėtinumo metodą (DT) ir palyginti jį su variogramos (V) metodu;
4. Sudaryti metodą daugiamačių duomenų ekstrepoliavimui taikant FVBL'o modelį;
5. Pritaikyti sukurtą metodą praktiniams uždaviniams.

# Krigingas

**Algorithm 1:** Krigingas

**Input** : $X = (x^{(1)}, x^{(2)}, ..., x^{(K)})$, $Z = (z_{(1)}, z_{(2)}, ..., z_{(K)})$, $E_i = 1$, $1 \leq i \leq K$, $X_{new} = (x_{new}^{(1)}, x_{new}^{(2)}, ..., x_{new}^{(P)})$, $1 \leq p \leq P$, $H$

**Output:** $Y(X_{new})$

**for** $i \leftarrow 1$ **to** $K$ **do**
    **for** $j \leftarrow 1$ **to** $K$ **do**
        $A_{i,j} = ((x^i - x^j) \cdot (x^i - x^j)^T)^H$
    **end for**
**end for**
**Function** *Prognosis(A, H, K, $x_{new}$)*:
    **for** $j \leftarrow 1$ **to** $K$ **do**
        $b_i = ((x_{new} - x^i) \cdot (x_{new} - x^i)^T)^H$
    **end for**
    $y = Z^T A^{-1}(b + E\frac{(1-E^T A^{-1} b)}{E^T A^{-1} E})$;
**return**
**for** $p \leftarrow 1$ **to** $P$ **do**
    $Y(X_{new}) = Prognosis(A,\, H,\, K,\, x_{new}^{(p)})$
**end for**

# Fraktalinio vektorinio Brauno lauko modelis

Tarkime, turime duomenų rinkinį: $X = (x^{(1)}, x^{(2)}, ..., x^{(k)})$; $x_{(i)} \in \mathbb{R}^d$; $1 \leq i, j \leq k$; $\mathbf{1} = (1, 1, ..., 1)$, $A = A_{i,j}$

$$A = ((x_i - x_j)^T (x_i - x_j))^H \tag{1}$$

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

$H$ - Hursto indeksas

Taikant darbe formulę (3) nereikia žinoti koordinačių pradžios, kitaip nei tradiciniame apibrėžime pagal A. Kolmogorovą.

[3] N. Pozniak, L. Sakalauskas ir L. Saltyte, Kriging model with fractional euclidean distance matrices,“ Informatica, t. 30, pp. 367-390, 2019.

# Variograma

Vienas iš metodų ¿įvertinti parametrą $H$ yra taikant variogramą. Tai fundamentalus modelis skirtas erdvinių porcesų modeliavimui.

Geostatistikos darbuose žinoma, kad variograma iš atsitiktinės funkcijos gali būti suskaidyta ¿į atskirus taškus.

$$V_{(1)} = \frac{2}{k(k-1)}(\sum_{i=1}^{k-1}\sum_{j=1}^{k}(z_i - z_j)(z_i - z_j))$$

$$V_{(2)} = \frac{\sum_{ii=1}^{m}\sum_{jj=1}^{m}(\sum_{i=1}^{k-1}\sum_{j=1}^{k}(z_{ii,i}-z_{ii,j})(z_{jj,i}-z_{jj,j})((x_i-x_j)^T(x_i-x_j))^H)^2}{\sum_{i=1}^{k}\sum_{j=1}^{k}((x_i-x_j)^T(x_i-x_j))^{2H}}$$

$$V(H) = V_{(1)} - V_{(2)} \tag{3}$$

Iš formulės (6) optimali parametro reikšmė ¿įvertinama minimizuojant pagal $H$.

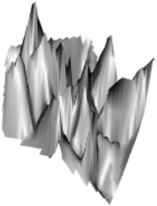
# Didžiausio tikėtinumo metodas

Taikant Didžiausio tikėtinumo metodą įvertinamos tokios parametrų reikšmės, su kuriomis gaunami rezultatai yra labiausiai tikėtini duotajam modeliui.

$$L = \frac{1}{K}(Z\ B_H(X)^{-1}Z^T - \frac{(Z\ B_H(X)^{-1}\mathbf{1})}{\mathbf{1}^T\ B_H(X)^{-1}\mathbf{1}}) \tag{4}$$

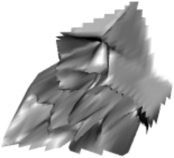
$$G(H) = \frac{1}{m}ln(L) + \frac{1}{k}ln(B_H(X)) \tag{5}$$

Tradiciškai yra taikomas variogramos metodas, o didžiausio tikėtinumo metodo taikymas erdvinių duomenų analizei yra mažai tyrinėtas.
Didžiausio tiketinumo metodą bandoma dažniausiai sudaryti tik procesams per Kolmogorovo tipo kovariacijų funkciją.

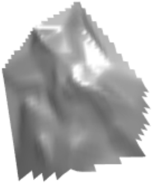


pav. 1: $H = 0.1$

pav. 2: $H = 0.9$

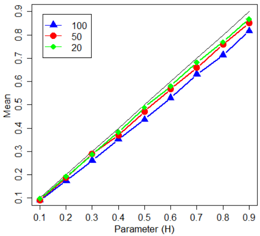


pav. 3: $H = 0.1$

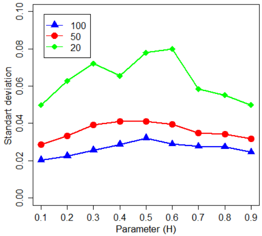


pav. 4: $H = 0.9$

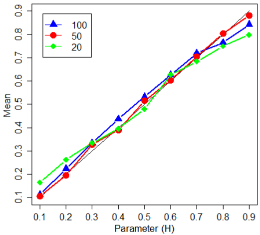
# Monte Karlo statistinis eksperimentas: DT metodas



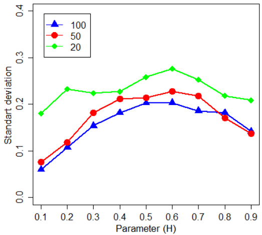

pav. 5: DT: Vidurkiai įvertintų parametrų $H$




pav. 6: DT: Standartiniai nuokrypiai parametrų $H$




pav. 7: V




pav. 8: V

| mėn. | 2017 | 2018 | 2019 | 2020 |
|---|---|---|---|---|
| 1 | 0.84 | 0.81 | 0.83 | 0.83 |
| 2 | 0.81 | 0.82 | 0.80 | 0.81 |
| 3 | 0.78 | 0.73 | 0.71 | 0.71 |
| 4 | 0.67 | 0.72 | 0.70 | 0.75 |
| 5 | 0.75 | 0.69 | 0.65 | 0.74 |
| 6 | 0.72 | 0.77 | 0.72 | 0.66 |
| 7 | 0.70 | 0.76 | 0.65 | 0.71 |
| 8 | 0.63 | 0.67 | 0.70 | 0.70 |
| 9 | 0.60 | 0.67 | 0.59 | 0.66 |
| 10 | 0.66 | 0.65 | 0.64 | 0.64 |
| 11 | 0.73 | 0.76 | 0.71 | 0.78 |
| 12 | 0.72 | 0.83 | 0.79 | 0.76 |

lentelė 3: Rezultatai, Hursto įverčiai

# Šepardo metodas

**Algorithm 2:** Shepard's Method

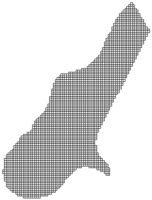
**Input** : $X = (x^{(1)}, x^{(2)}, ..., x^{(K)})$, $Z = (z^{(1)}, z^{(2)}, ..., z^{(K)})$, $E_i = 1$, $1 \leq i \leq K$, $X_{new} = (x_{new}^{(1)}, x_{new}^{(2)}, ..., x_{new}^{(P)})$, $1 \leq p \leq P$, $D$

**Output:** $Y(X_{new})$

**Function** *W*($x_{new}$, $D$ ):
    **for** $j \leftarrow 1$ **to** $K$ **do**
        $w_i = \frac{1}{((x^i - x_{new}) + (x^i - x_{new}))^D}$
    **end for**
**return**
**Function** *Prognosis*($x_{new}$, $D$ ):
    $w = W(x_{new}, D)$
    $Y = \frac{Z \cdot w}{E \cdot w}$
**return**
**for** $p \leftarrow 1$ **to** $P$ **do**
    $Y(X_{new}) = Prognosis(\ x_{new}^{(p)},\ D)$
**end for**

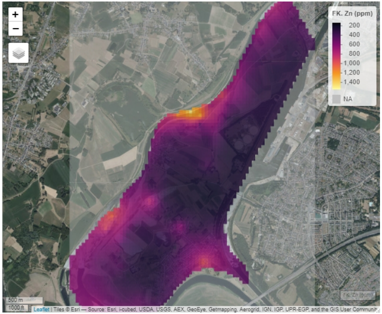
Sunkiųjų metalų (Cd, Cu, Pb, Zn) ir aukščio matavimai prie Maso upės (Olandija). $m = 5$; $K = 155$; $P = 3103$.


pav. 9: Matavimai

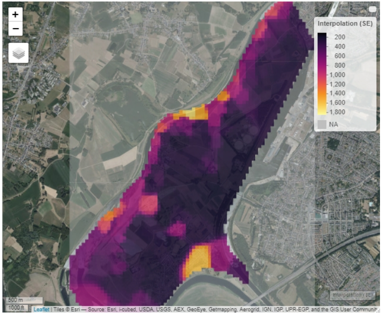


pav. 10: Gardelė

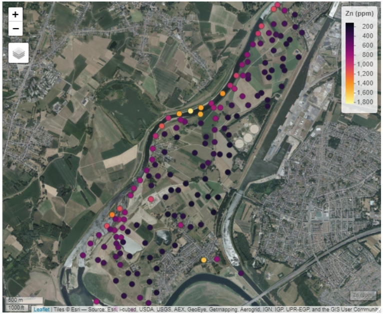
## Krigingo modelis



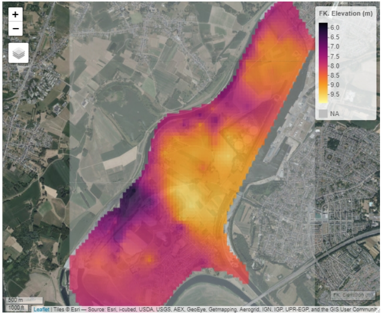

## Šepardo modelis



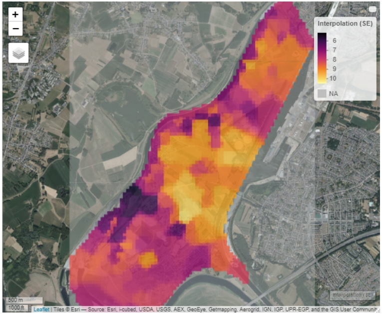

## Zn [ppm]



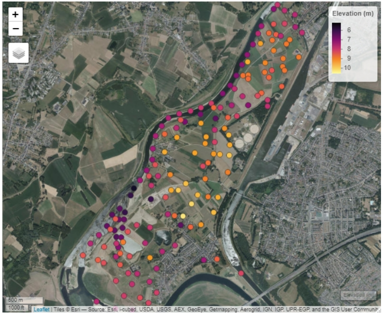
Krigingo modelis




Šepardo modelis




Aukštis [m]



# 2021/2022 m. m. darbo planas

## Disertacijos rengimo planas

1. Metodų taikymas praktiniams uždaviniams;
2. Sudalyvauti konferencijoje;
3. Parengti publikaciją;

Ačiū!