

Ataskaita už 2022 doktorantūros metus

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

Vilniaus Universitetas
Duomenų mokslo ir skaitmeninių technologijų institutas

2022 rugsėjo 30 d.

Disertacijos tema

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		Dalyvavimas konferencijose		Publikacijos		
	Planas	Įvykdyta	Planas	Įvykdyta	Planas	Įvykdyta	Būklė
I (2019/2020)	2	2	1	1	-	-	-
II (2020/2021)	2	2	1	1	1	1	Parengta
III (2021/2022)	-	-	1	2	0	0	Under review
IV (2022/2023)	-	-	-	1	1	-	Rengiama
Iš viso:	4	4	2	5	2	1	-

Disertacijos rengimo planas

Ataskaitinio pusmečio darbo planas ir jo įvykdymas

Einamieji studijų metai (III: 2021/2022).

Dalyvavimas konferencijose		
Planas	Įvykdyta	Konferencijos tipas
<i>7th Stochastic Modeling Techniques and Data Analysis International Conference and Demographics 2022 Workshop. SMTDA 2022 and Demographics 2022. <u>Atėnai</u>, <u>Graikija</u>.</i>	<u>Neringa Urbonaitė</u> , Leonidas <u>Sakalauskas</u> ; <i>Construction of a <u>Kriqing</u> Model using Fractional Brownian Vector Fields</i>	Tarptautinė

Disertacijos rengimo planas

Publikacijos

Publikacijos			
Planas	Ivykdyta	Būklė	Publikacijos tipas
Computers & Geosciences	Neringa Urbonaite, Leonidas Sakalauskas; A Novel Fractional Kriging Method for Multi-Output Data Modelling	<u>Under review</u>	<u>turi cituojamumo rodiklį</u>
R journal	Neringa Urbonaite, Leonidas Sakalauskas; FracKrigingR	<u>Rengiama</u>	<u>turi cituojamumo rodiklį</u>

Mokslinių tyrimų ir disertacijos rengimo etapai

2.3.4 FBL'o pritaikymas daugiamačiam krigingui ir ekstrepoliavimui. Sukurtų metodų panaudojimas praktiniams uždaviniams.

Sudarytas Šepardo modelis.

Fraktalinio Krigingo, **Ko Krigingo (gstat paketas)** ir Šepardo metodai (daugiamačiai) pritaikyti sunkiųjų metalų viršutiniame dirvožemio sluoksnyje interpoliavimui.

Mokslinių tyrimų ir disertacijos rengimo etapai

2.4. Gautų duomenų analizė, apibendrinimas, išvadų parengimas:
2.4.1 Gautų rezultatų statistinė analizė;
2.4.2 Rezultatų apibendrinimas, esminių rezultatų išskyrimas;
2.4.3 Išvadų parengimas.

2022 m. spalio mėn. –

2023 m. kovo mėn.

Sukurtas R paketas FracKrigingR; Priimta į CRAN 2021-11-08; [nuoroda](#). Pakete realizuotos fraktalinio kriginio, didžiausio tikėtinumo, fraktalinio Brauno lauko generavimo, atstumų matricos skaičiavimų funkcijos. Pateikti funkcijų pavyzdžiai. Sudaryta dokumentacija.

Papildyta funkcijomis: FracField.R; FracKrig.R; FracMatrix.R; MaxLikelihood.R; RecField.R; RecMatrix.R; Vario.R; Variance.R; CovMatrix.R

Disertacijos uždaviniai

Tyrimo objektas: Fraktalinis vektorinis Brauno laukas

Tikslas: Skirtų daugiamačiams duomenims ir jų sąsajoms vertinti, generuoti, prognozuoti, taikyti, algoritmų kūrimas, remiantis Fraktalinio vektorinio Brauno lauko modeliu, kai yra stebėjimų trūkumas.

Uždaviniai:

- 1 Sukurti FvBl'o modelį;
- 2 Sudaryti FvBl'o realizacijų generavimo algoritmą;
- 3 Sudaryti FvBl'o vertinimo algoritmą taikant didžiausio tikėtimumo 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.

Fraktalinis vektorinis Brauno laukas

The set of random vectors Z_i follows a Gaussian distribution with a constant mean. Here, we exclusively utilise covariance matrices that characterise the FBvf. The block kernel matrix is expressed as follows

$$F = [F_{i,j}]_{i,j=1}^{K+1}:$$

$$F = \begin{bmatrix} F_{xx} & F_{x*}^\top \\ F_{*x} & F_{**} \end{bmatrix} \quad (1)$$

We consider the column vector of ones, $\mathbf{1}_i = 1$, and let I be the $K \times K$ unit matrix, $s_0 = 1 - s^\top \cdot \mathbf{1}$. The general expression for the FBvf covariance matrix is obtained in terms of the lower-right submatrix of the kernel matrix F , which is defined as follows:

$$F_{**} = s_0 \cdot \mathbf{1} \cdot A_{0*}^\top \cdot (I - s \cdot \mathbf{1}^\top) + s_0 \cdot (I - \mathbf{1} \cdot s^\top) \cdot A_{*0} \cdot \mathbf{1}^\top - (I - s \cdot \mathbf{1}^\top) \cdot A_{**} \cdot (I - \mathbf{1} \cdot s^\top). \quad (2)$$

Fraktalinis vektorinis Brauno laukas

for the development of spatial data models it is enough to indicate the matrix of fractional distances. Therefore, let the distances be in the form of matrix $A_{**} = [A_{i,j}]_{i,j=1}^K$, and be calculated between X as a set of vector pairs:

$$A_{**} = ((x_i - x_j) \cdot (x_i - x_j))^H. \quad (3)$$

The distances to the origin of the coordinates are then calculated using $A_{0*} = ((x_i - x_0) \cdot (x_i - x_0))^H$, $1 \leq i \leq K$. We define block matrix $A = [A_{i,j}]_{i,j=1}^{K+1}$, $A_{0*} = A_{*0}^\top$,

$$A = \begin{bmatrix} 0 & A_{0*}^\top \\ A_{*0} & A_{**} \end{bmatrix}.$$

Definition

An FBvf with a Hurst parameter, $H \in [0, 1]$, is a spatial process based on the Gaussian distribution and distances between the points.

FvBI generavimas

In this study, we indicate the importance and benefits of distance matrix A . The algorithm begins with the indicated x^1 and x^2 to obtain the values Z_1 and Z_2 , similar to the recursion method discussed above,

$$Z_1 = \Sigma_{chol} \cdot \zeta_1 \cdot 2^{0.5-H} \cdot ((x_1 - x_2)^T \cdot (x_1 - x_2))^{\frac{H}{2}},$$

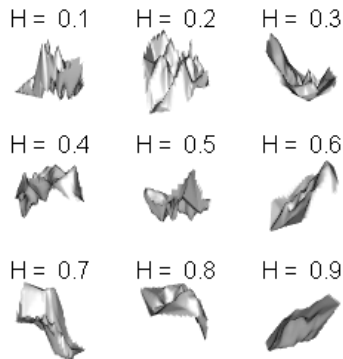
$$Z_2 = Z_1 \cdot (1 - 2^{2H-1}) + \Sigma_{chol} \cdot \zeta_2 \cdot \sqrt{1 - (2^{0.5-H} - 2^{H-0.5})^2} \cdot 2^{1-2H} \cdot ((x_1 - x_2)^T \cdot (x_1 - x_2))^{\frac{H}{2}},$$

The properties of the kernel matrix is as follows.

$$Z_i = \hat{z}_A(x_i) + \Sigma_{chol} \cdot \zeta_i \cdot \beta_A^2(x_i).$$

Choleckio metodas

Examples of fractional Brownian field surfaces were generated by applying the FBvf simulation method and choosing different values of the Hurst parameter, e.g. $H = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8,$ or 0.9 , as depicted in Pav 1.



Krigingas

Suppose that FBvf values $Z_i = \{z_{ij} = z_i(x_{ij}) | 1 \leq i \leq m, 1 \leq j \leq K\}$ are measured at the mutually distinctive points

$x_i = \{x_{ij} | 1 \leq i \leq p, 1 \leq j \leq K\}$, $A_{*0} = ((x_i - x_0) \cdot (x_i - x_0))^H$,

$F_{*x} = ((x_i - x) \cdot (x_i - x))^H$. The prediction at x_* of response vector surfaces via kriging are obtained using the conditional mean expressions:

$$\hat{z}_F(x_*) = (Z_i)^\top \cdot F_{**}^{-1} \cdot \left(f + \mathbf{1} \cdot \frac{1 - \mathbf{1}^\top \cdot F_{**}^{-1} \cdot f}{\mathbf{1}^\top \cdot F_{**}^{-1} \cdot \mathbf{1}} \right), \quad (4)$$

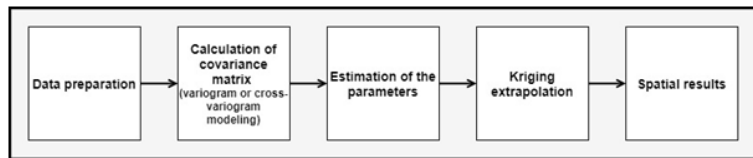
$$\hat{z}_A(x_*) = (Z_i)^\top \cdot A_{**}^{-1} \cdot \left(F_{*x} + \mathbf{1} \cdot \frac{1 - \mathbf{1}^\top \cdot A_{**}^{-1} \cdot F_{*x}}{\mathbf{1}^\top \cdot A_{**}^{-1} \cdot \mathbf{1}} \right), \quad (5)$$

and the uncertainty in estimates are captured in their conditional variance:

$$\beta_F^2(x_*) = \frac{\hat{\beta}}{K} \left(F - f^\top \cdot F_{**}^{-1} \cdot f + \frac{(1 - \mathbf{1}^\top \cdot F_{**}^{-1} \cdot f)^2}{\mathbf{1}^\top \cdot F_{**}^{-1} \cdot \mathbf{1}} \right), \quad (6)$$

$$\beta_A^2(x_*) = \frac{\hat{\beta}_A}{K} \left(F_{x*} \cdot A_{**}^{-1} \cdot F_{x*} - \frac{(1 - \mathbf{1}^\top \cdot A_{**}^{-1} \cdot F_{x*})^2}{\mathbf{1}^\top \cdot A_{**}^{-1} \cdot \mathbf{1}} \right), \quad (7)$$

Krigingas



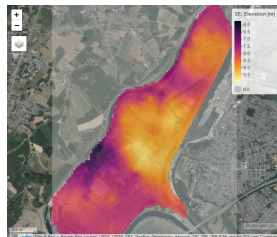
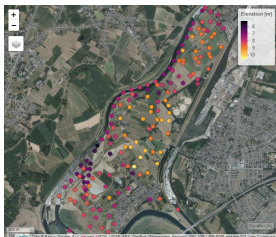
Fraktalinis krigingas (Variograma, Didžiausio tikėtimumo metodas)

Ko krigingas (Gauso variograma. gstat paketas))

Šepardo metodas (deterministinis)

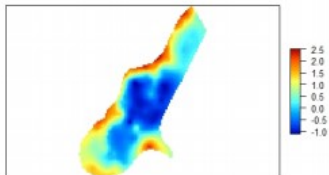
Interpolacija

- Process of predicting values at unknown locations using values at known location;
- Transforms measurements of a continuous phenomenon into continuous surface;
- Interpolation predicts within region; Extrapolation predicts outside region;

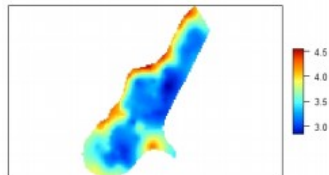


Interpolacija

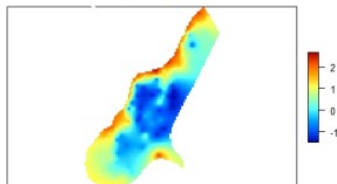
log-Cd (ppm)



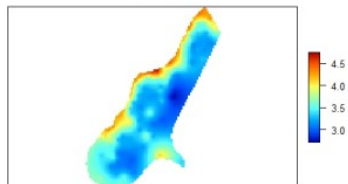
log-Cu (ppm)



log-Cd (ppm)

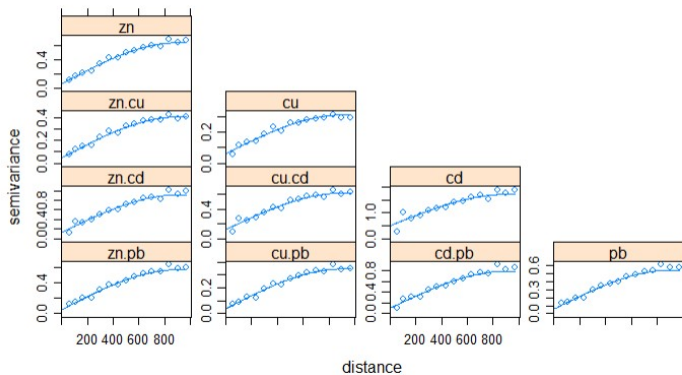


. log-Cu (ppm)



Variograma

Variograma (įstrižainė), kryžminė variograma (ne įstrižainė). Modeliavimas taikytas keturiems sunkiųjų metalų kintamiesiems.



lentelė 1: Results of the prognosis (MAE and RMSE values).

	Cd	Cu	Pb	Zn
MAE_{CO}	7.507	3.149	3.260	3.679
$RMSE_{CO}$	8.891	3.706	3.989	4.439
MAE_{FK}	6.994	2.603	3.073	3.259
$RMSE_{FK}$	8.444	3.112	3.456	3.944
MAE_{SE}	7.361	2.715	3.504	3.353
$RMSE_{SE}$	8.857	3.295	4.285	4.179

Comprehensive R Archive Network (CRAN)

- In principle, packages must pass R CMD check without warnings or significant notes to be admitted to the main CRAN package area.

FracKrigingR: Spatial Multivariate Data Modeling

Aim is to provide fractional Brownian vector field generation algorithm, Hurst para

Version: 1.0.0
Imports: [psych](#), [clusterGeneration](#), graphics, stats
Suggests: [knitr](#), [gstat](#), [sp](#), [rmarkdown](#), [raster](#)
Published: 2021-11-08
Author: Neringa Urbonaite [aut, cre], Leonidas Sakalauskas [aut]
Maintainer: Neringa Urbonaite <neringa.urbonaite@mif.vu.lt>
License: [GPL-2](#)
Copyright: Vilnius University Institute of Data Science and Digital Technol
URL: <https://github.com/NidaGreen/FracKriging>
NeedsCompilation: no
CRAN checks: [FracKrigingR results](#)

Documentation:

Reference manual: [FracKrigingR.pdf](#)

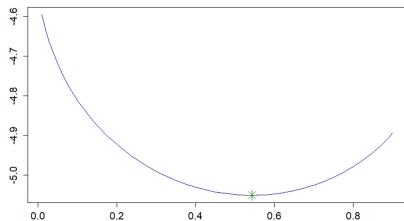
Downloads:

Package source: [FracKrigingR_1.0.0.tar.gz](#)
Windows binaries: r-devel: [FracKrigingR_1.0.0.zip](#), r-release: [FracKrigingR_1.0.0](#);
macOS binaries: r-release (arm64): [FracKrigingR_1.0.0.tgz](#), r-release (x86_64): [F](#)

Linking:

Please use the canonical form <https://CRAN.R-project.org/package=FracKrigingR> t

Comprehensive R Archive Network (CRAN)



```
library(sp)
library(gstat)
library(FracKrigingR)

data(meuse)
xy<-cbind(meuse$x,meuse$y)
X<-xy[1:50,]

Z<-as.matrix(zz1[1:50,])

K<-50
#Hurst parameter estimation
H<-MaxLikelihood(X,Z)
Xnew<-xy[51:100,]
results<- FracKrig(X,Z,Xnew,H)
```

```
results[,2], top =
max(data[,2]), bottom =
min(data[,2])
)
z3 = denormalize(
results[,3], top =
max(data[,3]), bottom =
min(data[,3])
)
RMSE<-function(z,prognosis){
rmse<-
sqrt((1/(length(z))))*sum((z-
prognosis)^2))
rmse
}
Cd<-RMSE(data[,1],z1)
Cu<-RMSE(data[,2],z2)
Pb<-RMSE(data[,3],z3)
..
```

lentelė 2: Paketo funkcijos

Funkcija	Aprašymas
FracField(K, m, H, X)	FvBl'o generavimo metodas
FracKrig(X, Z, Xnew, H)	Krigingas
FracMatrix(H, K, X)	Matrica A
CovMatrix(H, K, X, E, Z)	Matrica skirta MaxLikelihood.R
GSC(f, lowbound, upbound, tol)	Golden Section Search algoritmas
MaxLikelihood(X, Z, tol)	Didžiausio tikėtimumo metodas
RecField(K, m, H, X)	Rekursinis FvBl'o generavimas
RecMatrix(A, alpha, fa, K)	FvBl'o matrica
Variance(X, Z, Xnew, H)	Dispersija
Vario(X, Z, tol)	Variograma

Mokslinių tyrimų ir disertacijos rengimo tolimesni etapai

- Disertacijos dalių rengimas.
- Bendrųjų išvadų formulavimas.
- Paketo R papildymas.
- Publikacijos parengimas.
- Funkcijų skirtų simuliavimui analizei eksperimentų planavimui sudarymas.

Ačiū!