VILNIUS UNIVERSITY

Natalija
POZNIAK

Study and Application of Euclidean Matrices to Surrogate Modelling

SUMMARY OF DOCTORAL DISSERTATION

Natural Sciences,
Informatics N 009

VILNIUS 2019

This dissertation was written between 2014 and 2018 at Vilnius University.

Academic supervisor – Prof. Habil. Dr. Leonidas Sakalauskas (Vilnius University, Natural Sciences, Informatics – N 009).

This doctoral dissertation will be defended in a public meeting of the Dissertation Defence Panel:

Chairman – Prof. Dr. Olga Kurasova (Vilnius University, Natural Sciences, Informatics – N 009).
Members:
Prof. Dr. Nitin Agarwal (University of Arkansas at Little Rock, U.S., Natural Sciences, Informatics – N 009),
Prof. Dr. Romualdas Kliukas (Vilnius Gediminas Technical University, Technological Sciences, Informatics Engineering – T 007),
Prof. Dr. Rytis Maskeliūnas (Kaunas University of Technology, Technological Sciences, Informatics Engineering – T 007),
Prof. Dr. Julius Žilinskas (Vilnius University, Natural Sciences, Informatics – N 009).

The dissertation shall be defended at a public meeting of the Dissertation Defence Panel at 12:00 a. m. on the 17th September 2019 in Auditorium 203 of the Institute of Data Science and Digital Technologies of Vilnius University.
Address: Akademijos g. 4, LT-08412 Vilnius, Lithuania
The summary of the doctoral dissertation was distributed on the 17th of August, 2019.
The text of this dissertation can be accessed at the library of Vilnius University, as well as on the website of Vilnius University:
www.vu.lt/lt/naujienos/ivykiu-kalendorius

VILNIAUS UNIVERSITETAS

Natalija
POZNIAK

Euklido matricų tyrimas ir taikymas surogatiniam modeliavimui

DAKTARO DISERTACIJOS SANTRAUKA

Gamtos mokslai,
Informatika N 009

VILNIUS 2019

Disertacija rengta 2014– 2018 metais Vilniaus universitete.

Mokslinis vadovas – prof. habil. dr. Leonidas Sakalauskas (Vilniaus universitetas, gamtos mokslai, informatika – N 009).

Gynimo taryba:
Pirmininkė – prof. dr. Olga Kurasova (Vilniaus universitetas, gamtos mokslai, informatika – N 009).

Nariai:
prof. dr. Nitin Agarval (Little Rock Arkanzaso Universitetas, JAV, gamtos mokslai, informatika – N 009),
prof. dr. Romualdas Kliukas (Vilniaus Gedimino technikos universitetas, technologijos mokslai, informatikos inžinerija – T 007),
prof. dr. Rytis Maskeliūnas (Kauno technologijos universitetas, technologijos mokslai, informatikos inžinerija – T 007),
prof. dr. Julius Žilinskas (Vilniaus universitetas, gamtos mokslai, informatika – N 009).

Disertacija ginama viešame Gynimo tarybos posėdyje 2019 m. rugsėjo mėn. 17 d. 12 val. Vilniaus universiteto Duomenų mokslo ir skaitmeninių technologijų instituto 203 auditorijoje.
Adresas: Akademijos g. 4, LT-08412, Vilnius, Lietuva,
Disertacijos santrauka išsiuntinėta 2019 m. rugpjūčio mėn. 17 d.
Disertaciją galima peržiūrėti Vilniaus univeristeto bibliotekoje ir VU interneto svetainėje adresu:
https://www.vu.lt/naujienos/ivykiu-kalendorius

1. INTRODUCTION

1.1. Scope and relevance

Data science uses methods, processes, algorithms to extract knowledge and insights from structured and unstructured data. It involves various fields of mathematics, statistics, and computer science. Data science seeks an actionable and logically based model for prediction, extrapolation or interpolation uses. That makes data science different from traditional analytics and close to data mining. Therefore, data science allows generating effective analytical methods in areas such as medicine and social sciences that do not have specific data models in those areas.

A surrogate model is an engineering method used when an outcome of interest cannot be easily directly measured. Thus, a model of the outcome is used instead. Quite frequently, computer simulation of real systems can take many minutes, hours or even days to complete. As a result, routine tasks such as design optimization, design space exploration, sensitivity analysis and what-if analysis become impossible since they require thousands or even millions of simulation evaluations. One way to solve this problem is constructing approximation models, known as surrogate models, response surface models, metamodels or emulators that simulate a modelling object in a simplified way, creating and applying.

1.2. Relevance of the problem

The data extrapolation method of kriging was first developed by Georges Matheron (1963), based on the Master's thesis of Krige (1951). The kriging method is proposed to increase the efficiency of solving modelling, prediction and optimization problems (Jones, 2001; Forrester and Keane, 2009; Xiao et al., 2018). Since the kriging

method was developed as a consequence of its accurate predictions (Krige, 1951), it became one of the most promising numerical approaches in the various fields of engineering design, spatial statistics, experimental design and so on (Kwon and Choi, 2015; Bhosekar and Ierapetritou, 2018; Carpio et al., 2017).

In general, kriging is a method of interpolation for which the interpolated values are modelled by a multivariate Gaussian process, governed by prior covariances.

The basic idea of kriging is to predict the value of a response function at a given point by computing a weighted average of the known values of this function in the neighbourhood of the point. Thus, kriging gives a way of anticipating, with some probability, a result associated with values of the parameters that have never been met before using the existing information (e.g. the experimental measurements).

The need for surrogate models arises when reconstructing missed data, extrapolating data, or planning optimal experiments with a response function that requires a variety of resources, or when the results of experiments or calculations are only partially available, or predicting data from observed processes.

Most engineering planning problems require experiments to construct objective functions as variables functions. For example, the effect of laser radiation on optical materials is a result of computer-based calculations of the effects observation data, as well as the efficiency of wastewater treatment plants, depends on the proportions or quality of the filter, which is analyzed based on several filters created and realized during the experiments, etc. In both cases, data from a computational experiment or a real experiment can be used to create a real object surrogate model. Measuring points depending on many parameters may have a random or deterministic nature. The results depend on many parameters, but the selection and effect of these parameters is often unknown.

In addition, the amount of data collected is small because each additional experiment requires time and expenses. Sometimes, an

experiment can be dangerous, or the proper conditions needed to conduct that experiment can be very rare. Real challenges associated with these problems can occur very often. For example, the study of the safety of a nuclear reactor in critical states is related to the study of the operation of a reactor in a safe state (Levenson, 1981). Spacecraft flight trajectory studies that require trajectory prediction based on observed data are quite dangerous and expensive. The goal of chemical engineering to create new materials is often related to the design of experiments based on data obtained in laboratories after many experiments etc. Such data structures are studied by traditional deterministic or statistical methods. However, deterministic data analysis by traditional methods of interpolation or extrapolation requires various additional assumptions and does not take into account the uncertainty associated with data reconstruction (Shepard, 1968; Shumaker, 1976). Therefore, it is relevant to study the application of Gaussian random field (GRF) models to the analysis of experimental data. The properties of GRF depend on the covariances describing the dependencies between the points where computer or physical experiments were performed. The statistical models of GRF, when covariances are described by Euclidean distances between objects in fractional degrees have not yet been well studied, therefore their investigation and application to the analysis of experimental data for solving extrapolation, optimization or experiment design problems is a relevant problem.

1.3. Research Object

The research object of research is data science models and methods for analysis of structured multidimensional data. The focus is on surrogate modeling of multidimensional data based on properties of distances between measurement data matrices.

1.4. Aim and Objectives of the Research

The aim of this work is to study the properties of Euclidean distances with fractional degree rates, to create a surrogate kriging model based on them and apply it to extrapolate multidimensional data, to design series of experiments and to solve multi-extremal problems. In order to achieve these goals, the following problems are solved:

1. Study of fractional Euclidean distance matrices (FEDM) properties after using kernel matrices.
2. Creation of a multidimensional data model with homogeneous and isotropic Gaussian fields.
3. Development and application of a surrogate kriging algorithm based on the properties of fractional Euclidean distance matrices for extrapolation.
4. Application of the created multidimensional data model to design a series of experiments.
5. Application of the created multidimensional data model to optimize multi-extremal functions.

1.5. Scientific Novelty

The novelty of the research is following:

1. The properties of fractional Euclidean distance matrices studied through the properties of the kernel matrix and it is shown that most often the fractional Euclidean distance matrix is nonsingular.
2. Using FEDM properties, a GRF multidimensional data model created.
3. A multidimensional data kriging algorithm adapted for extrapolation developed.

4. Algorithm for extremal experiment planning numerically implemented by the Monte Carlo method.
5. The Bayesian algorithm of multi-extremal problem solving based on the properties of a fractional Euclidean distance matrix developed.

1.6. Practical Significance of the Results

The kriging method developed in the dissertation can be applied to extrapolate scattered data. The developed experiment planning method can be effectively used to plan a series of extremal experiments. The Bayesian optimization algorithm developed in the dissertation can be applied for solving multi-extremal problems where the computation of the objective function requires high costs.

The following practical results have been obtained:

1. An efficient algorithm for extrapolation of multidimensional data developed.
2. An efficient algorithm for solving multi-extremal problems, where the calculation of the objective involves high costs, developed.
3. Planning algorithm of extremal series of experiments developed, which is adapted for optimal planning of wastewater filters.

1.7. Defending propositions

Defending propositions of the thesis are:

1. The basic properties of fractional Euclidean distance matrices can be expressed using the properties of the kernel matrix.
2. The developed algorithm allows an efficient solution to scattered data extrapolation and kriging problems.

3. The developed Bayesian optimization algorithm allows an efficient solution to multi-extremal problems where the calculation of the objective function involves high costs.
4. Developed experiment planning methods allow effective planning of extremal experiment series.

1.8. Approbation and Publications of the Research

The results of the dissertation were presented at 5 international and 2 national scientific conferences. The main results of the dissertation were published in 3 periodical scientific publications: 1 of them is published in the journal, which is included in *ISI Web of Science* data base with own citation index, 2 of them are published in the periodical journals, referenced at *CEEOL*, *Index Copernicus,* and other data bases.

1.9. Structure of the Dissertation

The paper consists of five main parts – chapters, results and conclusions and references.

The first chapter discusses the aim of the dissertation, objectives, methods and the list of approbation and publications of the dissertation results.

In the second chapter, the relevance of selected topic is discussed and research on FEDM properties is presented. In the third chapter based on proven FEDM properties and a multidimensional data model is created using homogeneous isotropic Gaussian fields, which is adapted for kriging and extrapolation algorithms. In the fourth chapter, the method of planning a series of extremal experiments, which is studied by computer modelling and applied to solve the practical task of wastewater filter design is described. In the fifth chapter the Bayesian method for optimization of multi-extremal functions is developed by applying Gaussian fields, which are

described by fractional Euclidean distance matrices. The properties of the constructed optimization algorithm are investigated by computer simulation.

2. STUDY OF GEOMETRIC PROPERTIES OF DISTANCE MATRICES WITH FRACTIONAL DEGREE RATES

This chapter describes the data structures considered in the dissertation, their representation by Euclidean distance matrices with fractional degree rates, and the multidimensional geometry properties described by such matrices.

2.1. Multidimensional data

Multidimensional data is constantly being collected and studied in various scientific fields. Recently, the analysis of multidimensional data has become an important problem in applied mathematics, which is being actively solved in almost all areas of research. Collaborative data science explores methods and applications of multidimensional data, for example in econometrics, financial and economic analysis, biology and medicine, and processing of observed results.

2.2. Fractional Euclidean distance matrix

Square Euclidean distance matrices have been studied in the literature (Schoenberg, 1935; Gower, 1984; Weinberger, 2004), but the Euclidean distance, which represents the distance between two vectors, is calculated by subtracting the square root of these squares.

Let us assume that a data set of K d- dimensional vectors is

$$X = (x_1, x_2, \dots, x_K), \quad (1)$$

where $x_i \in \Re^d, 1 \le i \le K$. Let us denote the *K*x*K* matrix of fractional degrees δ of Euclidean distance squares among pairs of vectors by

$$A = \left[\left(\left|x_i - x_j\right|\right)^{\delta}\right]_1^K. \quad (2)$$

where $0 \le \delta \le 1$, $\left|x_i - x_j\right| = \left(x_i - x_j\right)^T \cdot \left(x_i - x_j\right)$.

Let us call Euclidean distance matrices with fractional degree rates $0 \le \delta < 1$, fractional Euclidean distance matrices (FEDM). Matrices of squared Euclidean distances are mainly studied in the literature, where $\delta = 1$. However, an important case is the standart Euclidean distances, where $\delta = \frac{1}{2}$.

The properties of FEDM are studied using the kernel matrix (see (Schoenberg, 1935; Gower, 1984; Weinberger, 2004)

$$F = -(I - E \cdot s^T) \cdot A \cdot (I - s \cdot E^T), \quad (3)$$

where $s \in \Re^K$, $s^T \cdot E = 1$, I denotes $K \times K$ unit matrix, E denotes K -dimensional vector-column of units. Kernel matrices, when $s = \frac{E}{K}$ and $s = (0,0,\ldots,1)$ are the most studied.

2.3. Positive definiteness of fractional Euclidean distance matrices

The main property of FEDM is that the kernel matrix is positively defined and its rank is $K - 1$, if the points in set (1) are different. In general, the squared Euclidean distance matrix degree rate may be less than $K - 1$, when $\delta = 1$ (see Schoenberg, 1935; Gower, 2004). This property of FEDM is considered in the following theorem:

Theorem 1. Then $A = \left[\left(\left(x_i - x_j\right)^T \cdot \left(x_i - x_j\right)\right)^{\delta}\right]_1^K$ of vectors $x_i \in \Re^p$, $x_i \neq x_j$, $i \neq j$, $1 \leq i,j \leq K$, $s^T \cdot E = 1$, $0 \leq \delta < 1$ FEDM, then, kernel matrix F is positively semi-definite of rank $K - 1$.

2.4. Geometrical properties of fractional Euclidean distance matrices

This chapter formulates and proves that:

Proposition 1. There is an embedded set $Z = (z_1, z_2, \ldots, z_K)^T$, $z_K \in \Re^{K-1}$, of rank $K - 1$, such that $A = \frac{1}{2} \cdot \left[\left(z_i - z_j\right)^T \cdot \left(z_i - z_j\right)\right]_1^K$.

Corollary 1. Let us consider kernel matrix F, constructed according to assumptions of Proposition 1. Let us assume $s_c \cdot E = 1$, $s_c \in \Re^K$ and $z_c = Z^T \cdot s_c$, where $F = Z \cdot Z^T$. Then

$$F = -(I - E \cdot {s_c}^T) \cdot A \cdot (I - s_c \cdot E^T)$$
$$= (I - E \cdot {s_c}^T) \cdot F \cdot (I - s_c \cdot E^T) =$$
$$= (I - E \cdot {s_c}^T) \cdot Z \cdot Z^T \cdot (I - s_c \cdot E^T) = (Z - E \cdot {z_c}^T) \cdot$$
$$(Z^T - z_c \cdot E^T). \blacksquare$$

Hence, Corollary 1 shows us the geometrical sense of s.

Corollary 2. If under the conditions of Proposition 1, the origin in the embedded space is displaced at the point z_1, namely, $s = (1,0,0,\ldots,0)$, then the elements of the kernel matrix are

$$F_{i,j} = A_{i,i} + A_{j,j} - A_{i,j},\ 1 < i,j \leq K,$$
$$F_{0,0} = 0,\ F_{1,j} = 0,\ F_{i,1} = 0,\ 1 < i,j \leq K. \blacksquare$$

For example, kernel matrix with $s = (0,0,\ldots,1)$ means that in the embedded space the coordinate start is at z_K.

Corollary 2. Under the conditions of Theorem 1, the centring vector $s = \frac{E}{K}$ minimizes the trace of kernel matrix.

Corollary 3. If the conditions of Theorem 1 are valid, then the centering vector $s = \frac{A^{-1} \cdot E}{E^T \cdot A^{-1} \cdot E}$ displays the origin of coordinates in the centre of the circumscribed sphere of the embedded set.

2.5. Experimental blocks with fractional Euclidean distance matrices

This chapter presents the inverse of FEDM more in detail. Let us denote a point $x \in \Re^d$. The fractional degrees of Euclidean distance squares between the X and x vectors

$$a = \left(|x_1 - x|^\delta,\ |x_2 - x|^\delta, \ldots, |x_K - x|^\delta\right)^T.$$

Now, define the extended kernel matrix, of the set $X \cup x$, calculated at the centering vectors $(s^T,\ s')$ and presented in the block matrix:

$$\tilde{F} = \begin{bmatrix} F & f \\ f^T & v \end{bmatrix} \equiv$$

$$\equiv -\begin{bmatrix} S & -s' \cdot E \\ -s^T & 1 - s' \end{bmatrix} \cdot \begin{bmatrix} A & a \\ a^T & 0 \end{bmatrix} \cdot \begin{bmatrix} S^T & -s \\ -s' \cdot E^T & 1 - s' \end{bmatrix},$$

where

$$S = I - E \cdot s^T,$$
$$F = s' \cdot E \cdot a^T \cdot S^T + s' \cdot S \cdot a \cdot E^T - S \cdot A \cdot S^T,$$
$$f = S \cdot A \cdot s - s' \cdot E \cdot a^T \cdot s - S \cdot a \cdot (1 - s'),$$
$$v = 2 \cdot (1 - s') \cdot s^T \cdot a - s^T \cdot A \cdot s.$$

Then next theorem relates the inverse and determinant of FEDM with that of the kernel matrix.

Theorem 2. Let us assume $A = \left[\left(\left(x_i - x_j\right)^T \cdot \left(x_i - x_j\right)\right)^{\delta}\right]_1^K$, $a = \left(|x_1 - x|^{\delta},\ |x_2 - x|^{\delta} \ldots, |x_K - x|^{\delta}\right)^T$, where $x_i,\ x_j,\ x \in \Re^d$, $x_i \neq x_j, i \neq j, x_i \neq x, 1 \leq i, j \leq K, 0 \leq \delta < 1, d \geq 1, E^T \cdot s \neq 1$. Then the inverse F^{-1} of the kernel matrix is

$$F^{-1} = M = -A^{-1} + \frac{A^{-1} \cdot E \cdot E^T \cdot A^{-1}}{E^T \cdot A^{-1} \cdot E} - \frac{q \cdot q^T}{D}, \text{ where}$$

$$q = s + s' \cdot A^{-1} \cdot \left(a + E \cdot \frac{1 - E^T \cdot A^{-1} \cdot a}{E^T \cdot A^{-1} \cdot E}\right),$$

$$D = s'^2 \cdot \left(a^T \cdot A^{-1} \cdot a - \frac{\left(1 - E^T \cdot A^{-1} \cdot a\right)^2}{E^T \cdot A^{-1} \cdot E}\right),$$

$$s' = 1 - E^T \cdot s.$$

3. KRIGING SURROGATE MODEL

In this chapter multidimensional data model using homogeneous Gaussian fields whose covariance matrix is described by the FEDM, is proposed and studied. Based on this model, a kriging algorithm is developed and adapted to extrapolate scattered multidimensional data. The properties of the developed kriging algorithm are studied by computer simulation, and the developed extrapolation algorithm is applied to the analysis of experimental data of wastewater filters. The purpose of surrogate modelling is to create a surrogate model that is as close as possible to the object being modelled, using the results of a small number of experiments. It should be noted that the interpretation of experimental measurements or computerized calculations has a probabilistic nature since the reconstructed data is a certain approximation and, secondly, measurement and calculation errors must be taken into account. In addition, deterministic methods are relatively ineffective in solving problems with a large number of variables, the latter occurring most frequently in real-world situations.

3.1. Random Gaussian field model described by FEDM

Random Gaussian fields are widely applied to probabilistic modelling. For example, modelling with homogeneous Gaussian fields (multidimensional stationary processes) has found many applications in bioinformatics, engineering, and physics, where data is obtained experimentally or using specific computer codes (Adler 1981; Bogush, 1986; Goff, 1994; Chamon, 1996; Lopez-Caballero 2010; Zaichik, 2010).

To model the observed data in a probabilistic way, one has to define the probability distribution of response surface values $Z(x,\omega) = (Z(x_1,\omega), Z(x_2,\omega), .., Z(x_K,\omega))$ in the given set of points $X = (x_1, x_2, \dots, x_K)$. Consider this probability distribution with the constant mean vector

$$EZ(x,\omega) = \mu \cdot E, \tag{4}$$

and the covariance matrix is directly proportional to the corresponding submatrix of the kernel matrix

$$E(Z(x,\omega) - \mu \cdot E) \cdot (Z(x,\omega) - \mu \cdot E)^T = \beta^2 \cdot F, \tag{5}$$

where μ and β are parameters, $\beta > 0$, and F is a positively defined matrix.

Let us say, measurements or computer simulations values for a objective function were obtained:

$$Y = (y_1, y_2, \dots, y_K)^T \tag{6}$$

Taking into account geometrical properties of FEDM, explored in the previous section, we can write the probability distribution function of the GRF values.

Theorem 3. The random Gaussian field $Z(x,\omega)$ exists in some probability space (Ω,Σ,P), , having the functions

$$p_{x_0,X_K}(Y_K) = \frac{e^{-\frac{1}{2\cdot\beta^2}\cdot Y_K^T\left(A_K^{-1}-\frac{A_K^{-1}\cdot E_K\cdot E_K^T\cdot A_K^{-1}}{E_K^T\cdot A_K^{-1}\cdot E_K}\right)\cdot Y_K-\frac{\left(Y_K^T\cdot A_K^{-1}\cdot\left(a_K+E_K\cdot\frac{1-E_K^T\cdot A_K^{-1}\cdot a_K}{E_K^T\cdot A_K^{-1}\cdot E_K}\right)-\mu\right)^2}{2\cdot\beta^2\cdot\left(a_K^T\cdot A_K^{-1}\cdot a_K-\frac{\left(1-E^T\cdot A_K^{-1}\cdot a_K\right)^2}{E_K^T\cdot A_K^{-1}\cdot E_K}\right)}}}{(2\pi)^{\frac{K}{2}}\cdot\beta^K\cdot\left((-1)^{K+1}\cdot|A_K|\cdot\left(a_K^T\cdot A_K^{-1}\cdot a_K\cdot E_K^T\cdot A^{-1}\cdot E_K-\left(1-E_K^T\cdot A_K^{-1}\cdot a_K\right)^2\right)\right)^{\frac{1}{2}}}, \quad (7)$$

as its finite-dimensional densities of distribution of $Y_K = (y_1, y_2, \ldots, y_K)^T$, $y_i = Z(x_i,\omega)$, $X_K = (x_1, x_2, \ldots, x_K)$ is a sequence of muttually disjoint points , $x_i \in \Re^d$, and disjoint with $x_0 \in \Re^d$, as well, $1 \le i \le K$, $A_K = \left[\left(\|x_i - x_j\|\right)^\delta\right]_{i,j=1}^K$, $a_K = \left(\|x_1 - x_0\|^\delta,\ \|x_2 - x_0\|^\delta \ldots, \|x_K - x_0\|^\delta\right)^T$, $E_K = (1,\ 1\ \ldots,1)^T$, $E_K \in \Re^K$, $K = 2,3,\ldots$, μ and β are parameters, $\beta > 0$.

Example. Let us assume that the observation set is displayed on a line: $x_k = a + t_k \cdot (b-a),\ a,\ b \in \Re^d,\ t_0 < t_1 < \ldots, < t_K$. If $\delta = \frac{1}{2}$, then GRF is distributed on this line as a Wiener process.

Let us write the logarithmic probability function:

$$L(Y,\mu,d^2,\ \delta) = \frac{1}{2}ln(|F|^{\frac{1}{2}}) + d\cdot ln(\beta) + \frac{(Y-\mu\cdot E)^T\cdot F^{-1}\cdot(Y-\mu\cdot E)}{2\cdot\beta^2}. \quad (8)$$

Theorem 4. Maximum likelihood parameters (MLE) μ and β^2 are as follows:

$$\hat{\mu} = \frac{Y^T\cdot F^{-1}\cdot E}{E^T\cdot F^{-1}\cdot E} = \quad (9)$$

$$= \frac{Y^T \cdot A^{-1} \cdot E}{E^T \cdot A^{-1} \cdot E} + Y^T \cdot \left(s - A^{-1} \cdot \left(\frac{E \cdot (E^T \cdot A^{-1} \cdot a \cdot \sigma^T + 1 - s')}{E^T \cdot A^{-1} \cdot E} - a \cdot s' \right) \right)$$

$$\hat{\beta}^2 = \frac{1}{K} \cdot \left(Y^T \cdot F^{-1} \cdot Y - \frac{\left(Y^T \cdot F^{-1} \cdot E\right)^2}{E^T \cdot F^{-1} \cdot E} \right) = \frac{1}{K} \cdot \left(\frac{(Y^T \cdot A^{-1} \cdot E)^2}{E^T \cdot A^{-1} \cdot E} - Y^T \cdot A^{-1} \cdot Y \right). \quad (10)$$

Remark 1. The least square estimate $\hat{\delta}$ is estimated by the univariate minimization of variance parameter MLE:

$$\hat{\delta} = arg \min_{0 \le \delta \le 1} \frac{1}{K} \cdot \left(\frac{(Y^T \cdot A^{-1} \cdot E)^2}{E^T \cdot A^{-1} \cdot E} - Y^T \cdot A^{-1} \cdot Y \right). \quad (11)$$

3.2. Conditional mean and conditional variance

For the best estimate of GRF for certain subsets of points $\hat{X}$, let us apply the conditional GRF mean under the condition that the values $Y = (y_1, y_2, \ldots, y_N)^T$ were observed, since such a model ensures least squares estimates.

Theorem 5. Let us assume $X = (x_1, x_2, \ldots, x_K)$ of mutually disjoint points, $x_i,\ x_j \in \Re^d$, $x_i \neq x_j$, $i \neq j$, $1 \le i, j \le K$, be given, at which the values $Y = (y_1, y_2, \ldots, y_K)^T$ of some GRF realization are known, namely, $y_i = Z(x_i, \omega)$. Let $A = \left[\left(\left(x_i - x_j\right)^T \cdot \left(x_i - x_j\right) \right)^{\delta} \right]_1^K$

be FEDM of the vector $x_i \in X$, $0 \le \delta < 1$.

Then the kriging predictor at $x \in \Re^d$ is as follows:

$$y(x) = Y^T \cdot A^{-1} \cdot \left(a + E \cdot \frac{(1 - E^T \cdot A^{-1} \cdot a)}{E^T \cdot A^{-1} \cdot E} \right), \quad (12)$$

its variance

$$\beta^2(x) = \hat{\beta}^2 \cdot \left(a^T \cdot A^{-1} \cdot a - \frac{(1 - E^T \cdot A^{-1} \cdot a)^2}{E^T \cdot A^{-1} \cdot E} \right) \quad (13)$$

where a, v, FEDM are blocks in representation $\tilde{F}$ and $\hat{\beta}^2$ MLE of the variance parameter (10).

3.2. The kriging method

Thus, kriging predictor (12) turns out to a linear extrapolator $y(x) = Y^T \cdot u(x)$, with the extrapolation weights

$$u(x) = A^{-1} \cdot \left(a + E \cdot \frac{(1 - E^T \cdot A^{-1} \cdot a)}{E^T \cdot A^{-1} \cdot E} \right), \quad (14)$$

and variance (13) helps us the measure of accuracy of extrapolation. The next property of the kriging extrapolator easily follows from (14).

Corollary 3. The weights in (14) satisfy the condition: $E^T \cdot u(x) = 1$.

3.3. Extrapolation of scattered multidimensional data

A Shepard extrapolator was chosen to evaluate the accuracy of the kriging extrapolator developed and to prove the efficiency of the extrapolation. The results of these extrapolator calculations is shown in Table 3.

The well-known approach for data interpolation is presented by the Shepard method (Shepard, 1968)

$$z(x) = \begin{cases} \frac{Y^T \cdot w(x)}{E^T \cdot w(x)}, & if \quad |x_i - x| \neq 0 \quad \forall i \\ y_i, & if \quad |x_i - x| = 0, \end{cases}, \quad (15)$$

the weights are chosen in the following way

$$w(x) = \left(\frac{1}{|x_1 - x|^{\delta}}, \frac{1}{|x_2 - x|^{\delta}}, \ldots, \frac{1}{|x_K - x|^{\delta}}\right). \quad (16)$$

3.4. Computer simulations

At the beginning of the experiments, the response of the system is generally unknown, so different analytic functions should be designed. In total six types of functions are considered (Kwon and Choi, 2015). Each test function is shown in Table 1.

Table 1. Types of test functions and test function domains.

Branin $z_{TF}(x,y) = \left(y - \frac{5x^2}{4\pi^2} + \frac{5x}{\pi} - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right) cos\, x + 10$	$x \in [-5\,;10]$ $y \in [0\,;15]$	(17)
Rosenbrook $z_{TF}(x,y) = 100(y - x^2)^2 + (1 - x)^2$	$x \in [-5\,;5]$ $y \in [-5\,;5]$	(18)
Rastrigin $z_{TF}(x,y) = 20 + x^2 + y^2 - 10(cos(2\pi x) + cos(2\pi y))$	$x \in [-5.12\,;5.12]$ $y \in [-5.12\,;5.12]$	(19)
Haupt $z_{TF}(x,y) = x\, sin(4x) + 1.1y\, sin(2y)$	$x \in [0\,;4]$ $y \in [0\,;4]$	(20)
Himmenblau $z_{TF}(x,y) = (x^2 + y - 11)^2 + (x + y^2 - 7)^2 + 0.1((x-3)^2 + (y-2)^2)$	$x \in [-5\,;10]$ $y \in [-5\,;10]$	(21)
Kwon $z_{TF}(x,y) = x^2 + x^2\cos(x) + y\, cos(y)$	$x \in [1\,;2]$ $y \in [1.5\,;3]$	(22)

Mean and standard deviation of variance parameter $\hat{\beta}^2$ and Euclidean fraction degree δ are given in Table 2. A histogram depicting an estimate of the variance of the Branin function and a histogram depicting the mean of the parameter δ are given and are presented in Fig. 1. (*K = 20, N = 200*).

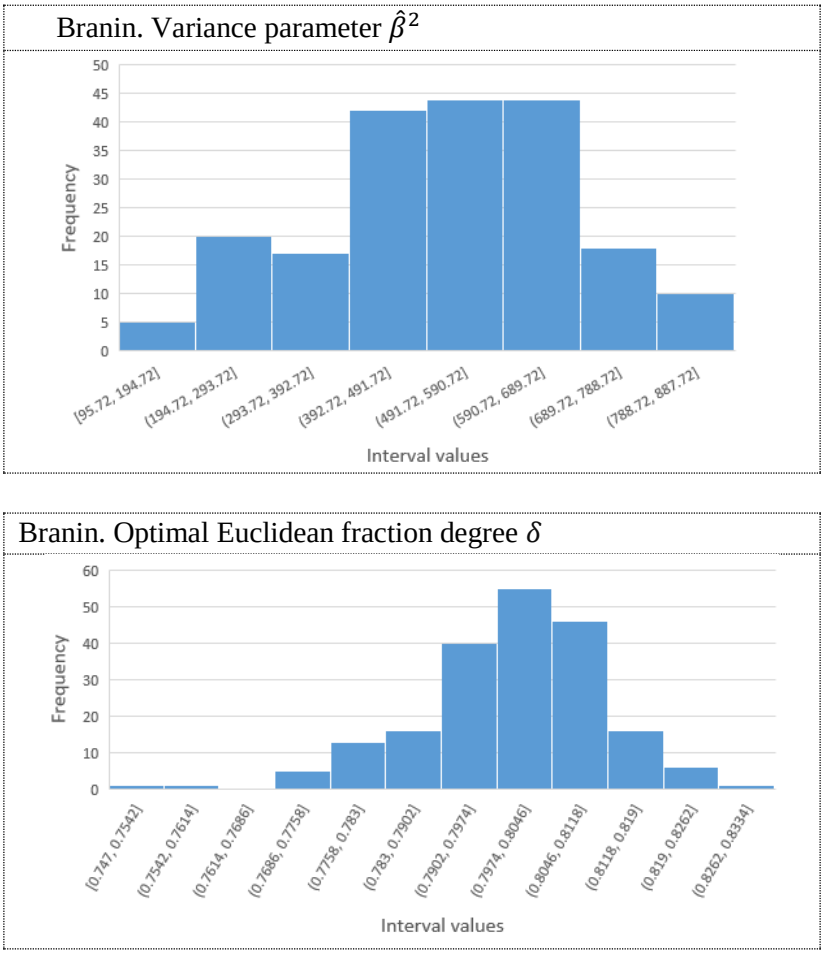

Fig 1. Branin. Variance parameter and optimal Euclidean fraction degree

In the prediction problem, the value of the Euclidean fraction degree δ could be chosen so as to get the best model results. An elegant alternative isto calculate the optimal δ value according to (11).

Table 2. Mean and standard deviation of variance parameter $\hat{\beta}^2$and Euclidean fraction degree δ.

Test function	Mean of $\hat{\beta}^2$	Std. dev. of $\hat{\beta}^2$	Mean of δ	Std. Dev. of δ
Branin	518.39	165.63	0.80	0.01
Kwon	0.41	0.05	0.77	0.03
Rosenbrook	$5.25 \cdot 10^7$	$1.86 \cdot 10^7$	0.67	0.03
Haupt	4.25	1.11	0.24	0.07
Rastrigin	87.86	26.69	0.43	0.09
Himmenblau	$4.00 \cdot 10^5$	$1.36 \cdot 10^5$	0.79	0.02

To compare the accuracy of the model the True Error Criterion is introduced (Kwon and Choi, 2015). It is defined as follows

$$TE(z(x,y)) = \sum\sum\sqrt{(z(x,y) - z_{TEST\ FUNCTION}(x,y))^2}, \tag{23}$$

$$TE\left(z_{Shepard}(x,y)\right) = \sum\sum\sqrt{\left(z_{Shepard}(x,y) - z_{TEST\ FUNCTION}(x,y)\right)^2} \tag{24}$$

The results are summarized in Table 3, and the best estimation of the true error are marked in bold for each test function. True error comparisons results are as follows (*K = 20, N = 200)*:

Table 3. True error results.

Test function	$TE(z(x,y))$	$TE(z_{Shepard}(x,y))$
Branin	**21.92**	59.48
Kwon	**0.12**	0.88
Rosenbrook	**8850**	13310
Haupt	**2.03**	2.47
Rastrigin	**12.43**	13.43
Himmenblau	**1061**	2149

Therefore, the kriging predictor, being a posterior expected value of the Gaussian random field, presents itself as an efficient extrapolator of scattered data, and, in turn, the variance of kriging predictor is an efficient measure of prediction or extrapolation error.

3.5. Modelling of effectiveness of eurface wastewater treatment filter filler

The effectiveness of filters, filled with construction waste and biocarbon, was analysed using the kriging method with distance matrices. The developed method allows modelling filter characteristics with different filler ratios, based on the previous experimental studies of filters.

Let $x = (x_1, x_2, x_3, x_4)^T$ be filler ratios vector, where x_1– quartz sand (QS), x_2– shredded autoclaved aerated concrete (SHAAC), x_3– stone wool (SW), x_4– biochar (BC). Let us denote the vector of results of filter characteristics measurement: Y_i^j, $1 \le i \le m$, $1 \le j \le K$, where $K = 4$ – number of experiments, and m – number of filter characteristics. Assume these filter characteristics be describing capability to treat different wastes.

Filter fillers: 1 – Quartz sand; 2 – Shredded autoclaved aerated concrete (66.7%) and stone wool (33.3%); 3 – Shredded autoclaved aerated concrete (33.3%) and biocarbon (66.7%); 4 – Shredded autoclaved aerated concrete (33.3%), biocarbon (33.3%) and stone wool (33.3%).

These filters with different fillers are designed for treatment of the main pollutants of surface wastewater: zinc (*Zn*), copper (*Cu*) and etc.. Then the experiment matrix of filler proportions was chosen as follows:

$$X = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.667 & 0 & 0.333 \\ 0 & 0 & 0.333 & 0.667 \\ 0 & 0.333 & 0.333 & 0.333 \end{pmatrix}$$

After the experiment, the filtration characteristics have been presented by the measurement matrix (see Table 4).

Table 4. Filtration characteristics measurements (%).

Filter	*PB*	*CD*	*ZN*	*CU*	*TC*	*TN*	*SS*	*BOD5*
	Y^1	Y^2	Y^3	Y^4	Y^5	Y^6	Y^7	Y^8
1	78.0	55.1	94.7	58.5	27.3	41.1	97.2	42.1
2	73.7	22.2	57.2	15.2	38.3	15.6	94.1	43.4
3	75.0	43.9	77.1	20.5	50.7	17.3	94.1	58.9
4	77.3	47.3	81.1	28.8	43.2	19.9	93.4	60.4

Visualization of the copper (*CU*) cleaning efficiency (in %) for the fillers proportion (*BC=1–QS– SHAAC–SW*) using the kriging approach proposed, is given in Fig. 2.

Two cases were studied: the filter was half filled with quartz sand (QS = 0.5) and without quartz sand fill (QS = 0).

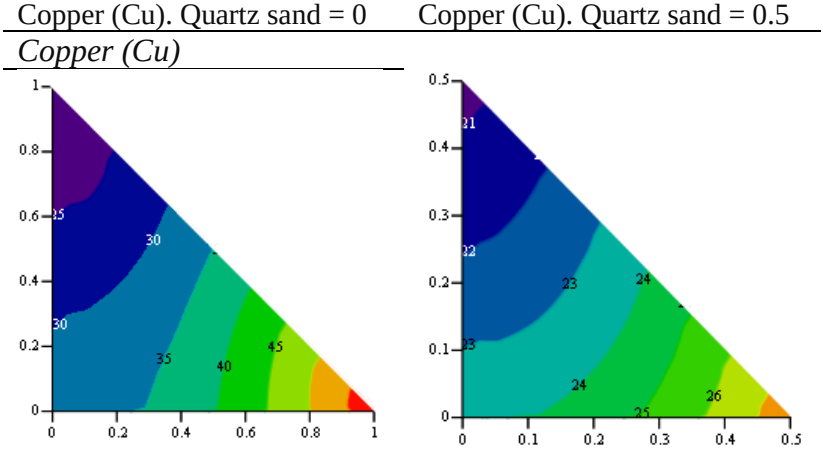

Fig 2. Cleaning capacity in %.

The developed kriging extrapolator was adapted to maximize characteristics using the Monte Carlo method (1000 trials).

Table 5. Maximization results.

Filter fillers		*CU*	*ZN*	*CD*	*PB*	*TC*	*TN*	*SS*
		Y^1	Y^2	Y^3	Y^4	Y^5	Y^6	Y^7
Quartz sand	x_1	0	0.03	0.06	0.03	0.45	0.03	0
Shredded autoclaved aerated concrete	x_2	0	0.02	0.09	0.04	0.5	0.02	0.03
Stone wool	x_3	0.05	0.03	0.03	0.02	0.04	0.02	0.03
Biocarbon	x_4	0.95	0,92	0.82	0.91	0.01	0.93	0.94
The result of extrapolation	**%**	55.95	92.48	51.88	77.73	48.01	39.18	96.93

4. SOLVING MULTI-EXTREMAL PROBLEMS

This chapter explores the Bayesian optimization of multi-extremal functions in which a random Gaussian field with a covariance matrix, expressed through the FEDM kernel matrix, is selected in the model of the function to be optimized.

4.1. Formulation of the optimization problem

Let us denote the continuous objective function, to be minimized on some compact feasible set. The Bayes optimization method ussually is build assuming that the values of the objective function are the realization of some random field. Let us denote that the objective function $f(x) = f(x, \omega)$ depends on a random scenario following from some probability space: $\omega \in (\Omega, \Sigma, P)$. Thus, the optimization problem to be solved is minimization

$$f(x, \omega) \to \min_{x \in D},$$

where $D \in \Re^N$ is a feasible set. Let us assume, that the objective function is continuous and the feasible set is compact.
The utility function $\Psi(f, b)$ must be selected for the evaluation of the quality of the optimization process. In general, this function depends on the results of the optimization steps:

$$\Psi(f, b) = \Psi(Y^K; X^K) = \Psi\left(Y^{k-1}, y^k; X^{k-1}, x^k\right). \qquad (26)$$

Let the utility function satisfy the following conditions:

A. The expected value of the utility function is defined from below.
B. Function $\boldsymbol{\Psi}\left(\boldsymbol{Y^{k-1}}, \boldsymbol{y^k}; \boldsymbol{X^{k-1}}, \boldsymbol{x^k}\right)$ is concave with respect to the variables $\boldsymbol{y^k}$.

C. Function $\boldsymbol{\Psi}(\boldsymbol{f}, \boldsymbol{b})$ satisfies the Lipschitz condition.
D. Function $\boldsymbol{\Psi}(\boldsymbol{f}, \boldsymbol{b})$ is monotonically increasing.

4.2. Recursive Equation System

Let us prove that the Bayesian method satisfies the system of recursive equations.

Theorem 6. Let the utility function satisfies the conditions A-D. Then Bayesian optimization method satisfies the system of recursive equations

$$U_{K-1}(Y^{K-1}) = \min_{x^K \in D} E_{x^K}(\Psi(Y^{K-1}, y^K)|Y^{K-1}), \tag{27}$$

$$U_k(Y^k) = \min_{x^{k+1} \in D} E_{x^{k+1}}(U_{k+1}(Y^k, y^{k+1})|Y^k), \tag{28}$$

$$k = 1,2,..,K-1$$

4.3. One-Step Bayesian optimization methods

The one-step Bayesian optimization point for each iteration must satisfy the equation:

$$R(\hat{x}^{k+1}) = \min_{x^{k+1}} \int_{-\infty}^{Y_{min}^k} \left(z - Y_{min}^k\right) \cdot \frac{e^{-\frac{\left(z - y(x^{k+1})\right)^2}{2 \cdot \hat{\beta}^2 \cdot \beta^2(x^{k+1})}}}{\sqrt{2\pi} \cdot \hat{\beta} \cdot \beta(x^{k+1})} dz$$

Note that the function minimized here can be written analytically.

4.4. Bayesian optimization algorithm

By applying the Bayesian optimization algorithm, the main computation time is wasted in finding the FEDM inverse matrix. In

this work a recursive algorithm for calculating conditional mean and conditional variance is developed, which allows us to shorten this time.

The algorithm:

1. Selection of initial points.
2. FEDM of initial points inversion.
3. Calculating the variance of GRF.
4. Optimization of the expected utility function by the Monte Carlo method.
5. Optimization of the expected utility function by the local landing method.
6. FEDM inversion using block matrices inversion formulas.
7. If the stoping condition is satisfied, the Bayesian method is terminated, otherwise step 4 is repeated.

4.5. Computer modelling

The effectiveness of the developed Bayes algorithms was analysed by computer modeling using the Monte Carlo method. The test function

$$G(x,y) = \sqrt{(x-6)^2 + (y-6)^2}, 0 \le x, y \le 10, \quad (29)$$

Branin (17) and Rastrigin (19) functions were investigated. Figures 3-5 depicted the average convergence curve given by the Bayesian method (BM) for optimizing test functions. The convergence curves presented for comparison are obtained using the simple Monte Carlo (MKM) method and the Simulated Annealing method (MAM).

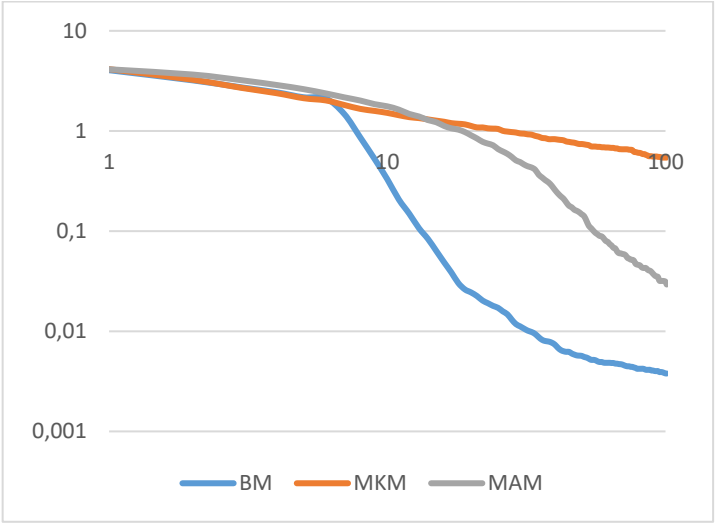

Fig 3. Test function (29) result

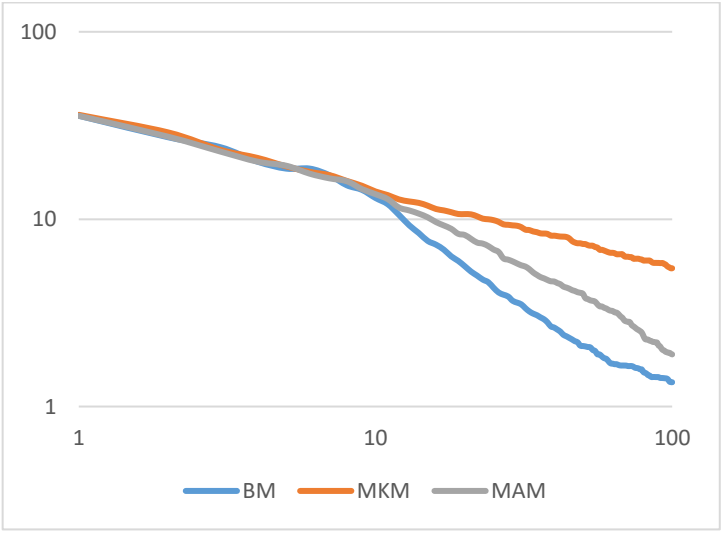

Fig 4. Rastrigin test function (19) result

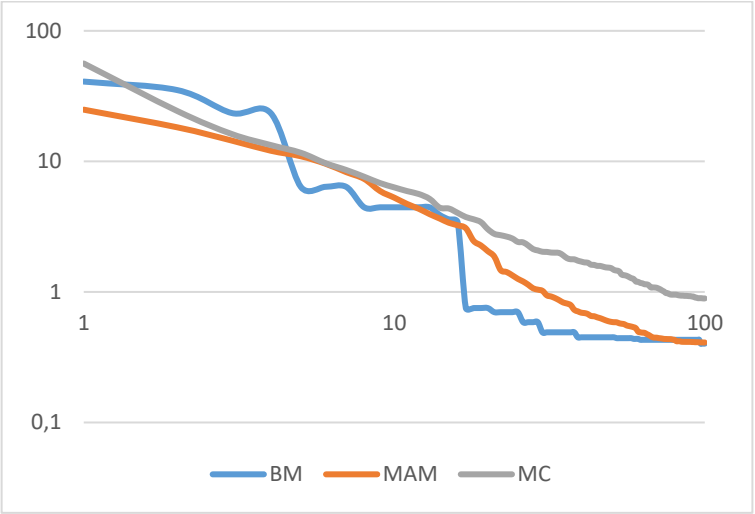

Fig 5. Branin test function (17) result

The computer simulation results show that the Bayesian method is more effective than the known methods for solving multidimensional problems.

5. OPTIMAL PLANNING OF EXTREMAL EXPERIMENTS

Let us consider the Bayesian experiment design method. The method is based on the use of observations / data obtained during the experiment.

5.1. Application of random Gaussian fields to experimental design

It is assumed that the result of the experiment after performing the experiment described by the parameter set x is equal to $f(x, \omega)$,

where $f(x,\omega)$ is the realization of GRF described in the second chapter.

Let us say that the random set $\tilde{X} = \left(X, \hat{X}\right)$ is given, of which the subset X consists of N vectors, where experiments and response surface values are obtained, by physical experiments, computer simulations, etc., and the subset $\hat{X} = (x_{N+1}, x_{N+2}, \ldots, x_K)$, consists of vectors where the values of the response modelled by the random field must be predicted, where $x_i \in \Re^d$, $1 \le i \le K$, $d \ge 1$. Let us denote the set of response function values $Y = (y_1, y_2, \ldots, y_N)^T$, $0 < N < K$, calculated for set X points.

Theorem 7. Let us assume that the data subset $X = (x_1, x_2, \ldots, x_N)$ consists of N vectors, where the values of the GRF realizations are $Y = (y_1, y_2, \ldots, y_N)^T$, and a subset $\hat{X} = (x_{N+1}, x_{N+2}, \ldots, x_K)$ consists of vectors, where GRF values are to be predicted, where $x_i \in \Re^d$, $N < K$, $1 \le i \le K$, $d \ge 1$. Let $\tilde{A} = \left[\left(\left(x_i - x_j\right)^T \cdot \left(x_i - x_j\right)\right)^\delta\right]$ FEDM vectors $x_i \in \Re^d$, $x_i \ne x_j$, $i \ne j$, $1 \le i,j \le K$, $0 \le \delta < 1$, where $\tilde{E}^T \cdot \tilde{s} = 1$ and $\sigma \ne 0$. Then, the conditional mean vector and the conditional covariance matrix of the GRF values for the subset $\hat{X}$ are written as follows:

$$y(\hat{X}) = Y^T \cdot A^{-1} \cdot \left(a + E \cdot \frac{(\varepsilon^T - E^T \cdot A^{-1} \cdot a)}{E^T \cdot A^{-1} \cdot E}\right) \qquad (30)$$

$$\beta(\hat{X}) = \hat{\beta}^2 \cdot \left(a^T \cdot A^{-1} \cdot a - \eta - \frac{(\varepsilon - a^T \cdot A^{-1} \cdot E) \cdot (\varepsilon^T - E^T \cdot A^{-1} \cdot a)}{E^T \cdot A^{-1} \cdot E}\right) \qquad (31)$$

where ε, , a, η, are the respective blocks of FEDM and kernel matrix decomposition, $\hat{\beta}^2$ is MLE of parameter β.

The following expressions allow us to write down the meaning of the expected utility function:

$$U(\hat{X}) = \int_{\Re^L} max(\hat{y}_{max}, \hat{y}_{N+1}, .., \hat{y}_K)\, p\left(\hat{Y}, y(\hat{X}), \beta(\hat{X})\right) d\hat{Y} = \qquad (32)$$

$$= y_{max} + \int_{max(\hat{y}_{N+1}, .., \hat{y}_K) \geq y_{max}} (max(\hat{y}_{N+1}, .., \hat{y}_K) - y_{max})\, p\left(\hat{Y}, y(\hat{X}), \beta(\hat{X})\right) d\hat{Y}.$$

A series of new experiments should be planned to maximize the expected increment of the objective response function. Thus, the plan for an optimal experiment series is a solution of the equation:

$$\hat{X}_{opt} = arg\max_X \int_{max(y_{N+1}, .., y_K) \geq y_{max}} (max(y_{N+1}, .., y_K) - y_{max}) \cdot \ln\left(p\left(Y, y(X), \beta(X)\right)\right) p\left(Y, y(\hat{X}_{opt}), \beta(\hat{X}_{opt})\right) dY. \qquad (33)$$

Because the integrals obtained in expression (33) are multidimensional, the Monte Carlo method was applied to design the series of experiments according to (33).

Let $(Y_1, Y_2, .., Y_M)$ be a sample of a multidimensional Gaussian distribution $N\left(y(X_1), \beta(X_1)\right)$ of size M. Then one can then denote the Monte Carlo estimates as follows:

$$E\left(H(Y, y_{max})\right) = \frac{1}{M} \cdot \sum_{j=1}^{M} H\left(Y_j, y_{max}\right)$$

$$E(H(Y, y_{max}) \cdot Y) = \frac{1}{M} \cdot \sum_{j=1}^{M} H\left(Y_j, y_{max}\right) \cdot Y_j \qquad (34)$$

$$E(H(Y, y_{max}) \cdot Y \cdot Y^T) = \frac{1}{M} \cdot \sum_{j=1}^{M} H\left(Y_j, y_{max}\right) \cdot Y_j \cdot Y_j^T$$

The estimates are inserted into expression (33), that is maximized by $\hat{X}$ for fixed estimates (34). Let us denote this

intermediate optimization result as $\hat{X}_{i+1}$. Now, the new sample is recalculated using $\hat{X}_{i+1}$ for computing conditional expectation, covariances, while the maximization is repeated. The process is terminated, when the difference of results of two consecutive steps become negligible.

5.2. Computer modelling

The algorithm for experimental series planning has been explored by the Monte Carlo method. The set from N=10 has been uniformly randomly simulated in the feasibility set and respective values of the test function have been computed.
The test function

$$G(x,y) = \sqrt{(x-5)^2 + (y-5)^2},\ 0 \le x,y \le 10 \qquad (35)$$

was investigated. The accuracy of the calculations was 10^{-3}.
Matrix of generated initial points and function values:

$$X = \begin{pmatrix} 0.0126842 & 9.8850847 \\ 1.9332302 & 1.1907973 \\ 5.850061 & 0.0892266 \\ 3.503081 & 5.3166416 \\ 8.2283773 & 6.0176413 \\ 1.74129 & 1.6624948 \\ 7.1049541 & 4.5078589 \\ 3.0398605 & 0.570559 \\ 0.9141127 & 7.8331917 \\ 1.4731341 & 5.1987616 \end{pmatrix}$$

The initial values of the planned series have been chosen randomly:

$$\hat{X} = \begin{pmatrix} 1.545929 & 3{,}6984743 \\ 6.0756318 & 7.616284 \\ 0.6163341 & 7.7764243 \\ 1.3140974 & 4.8630338 \\ 4.8852246 & 7.4508873 \end{pmatrix}$$

After 80 itterations results were obtained and depicted in Fig. 6-8.

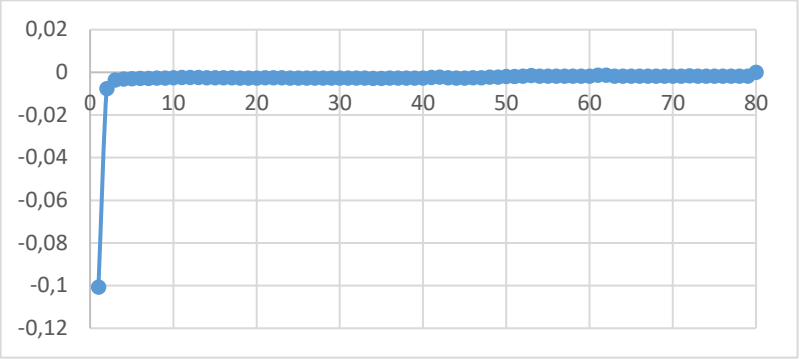

Fig 6. The difference between the values of the probability function (32) before and after one optimization step

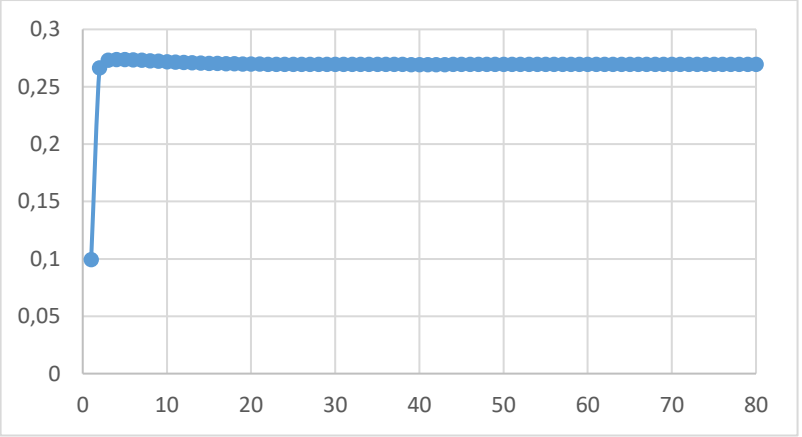

Fig 7. Expected increase in utility function (32) after each optimization step

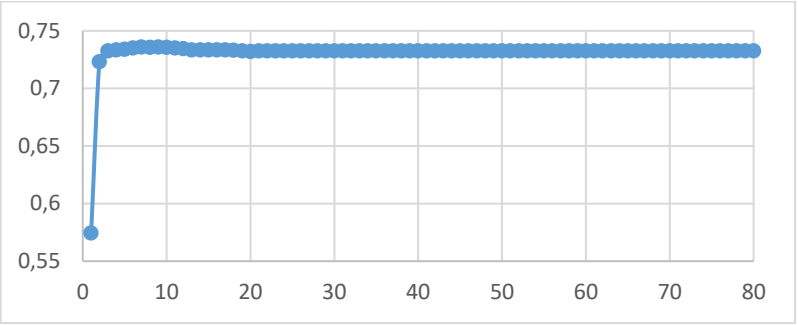

Fig 8. Probability of change in utility function (32)

The results of the objective function after 49 steps:

1.4374187	5.1898026
4.9101319	5.5414241
1.7202134	6.7837714
1.4733495	5.1987896
4.4044542	5.1045358

We compare the results to the initial values of the objective function:

$G(x), x \in \hat{X}$	$G(x), x \in X$
3.565067	6.986573
0.547449	4.892954
3.731863	4.991803
3.535661	**1.533754**
0.608276	3.387226
	4.667248
	2.156409
	4.844223
	4.97363
	3.535661

The values of the test function at the points of the proposed series were **0.5474** and is less than the initial value equal to **1.5338**.

5.3. Modelling of effectiveness of surface wastewater treatment filter filler

The created method is applied for optimal design of waste water filters. The results of the initial studies were used to plan the experiments on the wastewater treatment filters. Thus, the experiment matrix of filler proportions is as follows:

$$X = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.667 & 0 & 0.333 \\ 0 & 0 & 0.333 & 0.667 \\ 0 & 0.333 & 0.333 & 0.333 \end{pmatrix}$$

The filtration characteristics measurement of element *TC*, at this experiment matrix were:

$$z_{TC} = \begin{pmatrix} 27.26 \\ 38.33 \\ 50.67 \\ 43.18 \end{pmatrix}$$

The new plan for proportions for filter fillers by the developed method has been created:

Quartz sand	Shredded autoclaved aerated concrete	Biochar	Stone wool
0	1	0	0
0	0	0	1
0,334	0	0	0,666
0	0	0,16	0,84

RESULTS AND CONCLUSIONS

Data structures described by fractional Euclidean distance matrices (FEDMs) are quite common in applications. The statistical models of RGF, when covariates are described by Euclidean distances between objects in fractional degrees, have not yet been well studied, so their study and application to the analysis of experimental data for solving extrapolation, optimization or experiment design problems is a topical issue.

The following results were obtained in the dissertation:

1. It is proved that if a multidimensional data set consists of different vectors, the FEDM kernel matrix of this data set is negatively defined and its rank is one unit lower than the FEDM order.
2. Expressions for inverse FEDM matrix and determinant via kernel matrix were constructed.
3. Developed and investigated multidimensional data model with random Gaussian fields whose covariance matrix is described by FEDM kernel matrix.
4. Based on this model, a kriging algorithm is developed to extrapolate the scattered multidimensional data, which depends on several parameters whose estimates are obtained by the maximum likelihood method.
5. Computer simulation results showed that the developed extrapolation method is superior to the Shepard extrapolator.
6. Bayesian method for solving Bayesian multi-extremal problems using the GRF model was developed for response function, showing that this method satisfies the system of recursive equations.
7. A simplified method for solving these recursive equations is proposed and an algorithm for its implementation is developed.

8. Applying the developed GRF model, a method of planning a series of extremal experiments was developed by applying the Monte Carlo approximation method for planning a series of extremal experiments.
9. The developed method of extremal experiments was investigated by computer simulations and applied to plan a series of experiments of wastewater treatment filter.

The obtained theoretical and experimental results allow us to make the following conclusions:

1. The kriging method developed depends only on the FEDM and does not depend on the centering vector s and the beginning of the coordinates.
2. The FEDM degree rate $\boldsymbol{\delta}$ is a method parameter which can be estimated using the least squares method.
3. Computer simulations have shown that the developed kriging extrapolator is more efficient than the average standard error criterion over the Shepard extrapolator, often applied to scattered data analysis.
4. The results of computer simulations showed that the developed multi-extremal optimization method is more efficient than other known global optimization methods (Simulated Annealing) according to the convergence.

LIST OF PUBLICATIONS ON THE TOPIC OF DISSERTATION

1. *Pozniak, N., Sakalauskas, L. (). New Bayesian method for multiextremal optimization, Computational Methods and Techniques, KU, (accepted).*
2. Pozniak N., Sakalauskas L., Saltyte L. (2019). Kriging Model with Fractional Euclidean Distance Matrices. Informatica (Clarivate Analytics, Scopus), vol. 30, nr. 2, p. 367-390, http://dx.doi.org/10.15388/Informatica.2019.210.
3. Pozniak, N., Sakalauskas, L. (2019). The method for the optimal experiment design. Proceedings of International Scientific Conference Contemporary Issues in Business, Management and Economics Engineering, VGTU, ISBN 978-609-476-161-4, https://doi.org/10.3846/cibmee.2019.012.
4. Pozniak N., Sakalauskas L., (2017). Fractional Euclidean Distance Matrices Extrapolator for Scattered Data. Jaunųjų mokslininkų darbai. (Index Copernicus, CEEOL), nr. 2 (47), p. 56–61, (ISSN 1648-8776), https://doi.org/10.21277/jmd.v47i2.156.

REFERENCES

[1] Abrahamsen P., 1997, A Review of Gaussian random fields and correlation functions, *Norsk Regnesentral/Norwegian Computing Center*, p. 64.

[2] Adler R. J., 1981, *The Geometry of Random Fields*, John-Wiley & Sons, Chichester.

[3] Alfakih A. Y., Anjos M. F., Piccialli V., Wolkowicz H., 2011, Euclidean Distance Matrices, Semidefinite Programming, and Sensor Network Localization (a survey), *Portugaliae Mathematica*, vol. 68, p. 53–102.

[4] Balaji R., Bapat R. B., 2006, On Euclidean distance matrices, *Linear Algebra and its Applications*, vol. 424, p.. 108–117.

[5] Bartkute V., Sakalauskas L., 2009, Statistical inferences for termination of Markov type random search algorithms. *Journal of Optimization Theory and Applications*, vol. 141, no. 3, p. 475–493.

[6] Beauzamy B., 2004, *Methodes probabilistes pour l'etude des phenomenes reels*, Ouvrage edite par la Societe de Calcul Mathematique, http://www.scmsa.eu/RMM/HS_e.pdf

[7] Ben-Israel A., Greville T.N.E, 2003, *Generalized Inverses: Theory and Applications (2nd edition)*, Springer-Verlag, New York.

[8] Bhosekar A., Ierapetritou M., 2018, Advances in surrogate based modeling, feasibility analysis, and optimization: A review, *Computers & Chemical Engineering*, vol. 108, p. 250–267.

[9] Biehl M., Ghio A., Schleif F. M., 2015, Developments in computational intelligence and machine learning, *Neurocomputing*, vol. 169, p. 185–186.

[10] Bisgaard M. E., Sasvari Z., 2000, *Characteristic functions and moment sequences: Positive definiteness in probability*. Nova Science Publishers.

[11] Blum A., 2005, Random Projection, Margins, Kernels, and Feature-Selection, Subspace, *Latent Structure and Feature Selection*, p. 52–68.
[12] Bogomolny E., Bohigas O., Schmit C., 2008, Distance matrices and isometric embeddings, eprint arXiv:0710.2063.
[13] Bogush, A.J., Elkins, R.E., 1986, Gaussian field expansions for large aperture antennas, *IEEE Transactions on Antennas and Propagation*, AP-34(2).
[14] Bomba A., Safonyk A., 2013, Mathematical modeling of aerobic wastewater treatment in porous medium, *Zeszyty Naukowe Wyższej Szkoły Informatyki*, vol. 12, nr. 1, p. 21–29.
[15] Carpio R., Giordano R. C., Secchi A., 2017, Enhanced Surrogate Assisted Global Optimization Algorithm Based on Maximizing Probability of Improvement, *Computer Aided Chemical Engineering*, vol. 40, p. 2065–2070.
[16] Cavoretto R., De Rossi A., 2010, Fast and accurate interpolation of large scattered data sets on the sphere, *Journal of Computational and Applied Mathematics*,vol. 234, p. 1505–1521.
[17] Casella, G., Berger, R., 2002, Statistical Inferences. Wadsworth.
[18] Chamon C. C., Mudry C., Wen X. G., 1996, Localization in Two Dimensions, Gaussian field theories, and multifractality. *Phys. Rev. Lett*., vol. 77, p. 4194–4197.
[19] Forrester A. I. J., Keane A. J., 2009, Recent advances in surrogate-based optimization, *Progress in Aerospace Sciences*, vol. 45, Issues 1–3, p. 50–79.
[20] Franke R., 2015, Scattered Data Interpolation: Tests of Some Methods, *Mathematics of computation*, vol. 38, nr. 157, p. 181–200.
[21] Genton M. G., 2000, The Correlation Structure of Matheron's Classical Variogram Estimator Under Elliptically Contoured Distributions, *Mathematical Geology*, vol. 32, p. 127–137.

[22] Gillespie D. T.,1996, Exact numerical simulation of the Ornstein–Uhlenbeck process and its integral, *Physical Revue E*, vol. 54, p. 2084–2091.
[23] Gneiting T., 2002, Compactly Supported Correlation Functions, *Journal of Multivariate Analysis*, No.83, p. 493–508.
[24] Goff, J. A., Holliger, K., Levander, A., 1994, Model fields: A new method for characterization of random seismic velocity heterogeneity. *Geophys. Res. Letters*, vol. 21, p. 493–496.
[25] Goff, J. A., Jordan, T. H., 1988, Stochastic modeling of seafloor morphology: Inversion of sea beam data for second-order statistics, *J. Geophys. Res.*, vol. 93, p. 13589–13608.
[26] Golub G. H., Van Loan C. F., 1996, *Matrix Computations, 3rd. ed.*, Baltimore, MD: Johns Hopkins, p. 51.
[27] Gower J. C., 1982, Euclidean distance geometry, *Math. Scientist*., 7, l–14.
[28] Gower J. C., 1984, Multivariate analysis: Ordination, multidimensional scaling and allied topics, *Handbook of Applicable Mathematics: Vol. 6. Statistics (E. H. Lloyd, Ed.)*, p. 727–781, Wiley, Chichester.
[29] Gower J. C., 1983, Euclidean matrices and their Euclidean approximation, *in Third International Conference on Data Analysis and lnfmatics (E. Diday, Ed.), INRIA, Versailles*, p. 1–19.
[30] Hartle J. B., Hawking S.W., Hertog T., 2013, Vector Fields in Holographic Cosmology, p. 17, arXiv:1305.7190v1 [hep-th], DOI: 10.1007/JHEP11(2013)201.
[31] Hinrichs A., Vybiral J., 2011, On positive positive-definite functions and Bochner's Theorem, *Journal of Complexity*, Vol. 27, P. 264–272.
[32] Jones D. R., 2001, A Taxonomy of Global Optimization Methods Based on Response Surfaces, *Journal of Global Optimization, Kluwer Academic Publishers, Netherlands*, vol. 21., p. 345–383.

[33] Hawking S., 2014, Singularities and the geometry of spacetime, *The European Physical Journal H*, vol. 39, p. 413–503.
[34] Horn R. A., Johnson C. R., 2013, *Matrix Analysis 2nd ed.*, ISBN 978-0-521-54823-6, Cambridge University Press.
[35] Khoshnevisan D., 2002, *Multiparameter Processes An Introduction to Random Fields*, Springer-Verlag New York.
[36] Krige, D. G., 1951, A statistical approach to some mine valuations and allied problems at the Witwatersrand, Master's thesis of the University of Witwatersrand, Australia.
[37] Krislock N., Wolkowicz H., 2012, Euclidean Distance Matrices and Applications, in: Anjos, M. F., Lasserre, J. B., *Handbook on Semidefinite, Conic and Polynomial Optimization*, Vol. 166 of International Series in Operations Research & Management Science, p. 879–914.
[38] Kushner H., 1964, A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise, *Journal of Basic Engineering*, vol. 86, p. 97–106.
[39] Kwon H., Choi S., 2015, A trended Kriging model with R2 indicator and application to design optimization, *Aerospace Science and Technology*, vol. 43, p. 111–125.
[40] Lay D. C., 2005, *Linear Algebra and Its Applications*, Addison Wesley.
[41] Levenson M., Rahn F., 1981, Realistic estimates of the consequences of nuclear accidents, *Nuclear Technology*, vol. 53, p. 99–110.
[42] Li J., Lu B. L., 2009, Anadaptive image Euclidean distance, *Pattern Recognition*, vol. 42, p. 349–357.
[43] Menafoglio A., Petris G., 2015, Kriging for Hilbert-space valued random fields: The operatorial point of view, *Journal of Multivariate Analysis*, vol. 146, p. 84–94.
[44] Micchelli C. A., 1986, Interpolation of Scattered Data: Distance Matrices and Conditionally Positive Definite Functions, *Constructive Approximation*, vol. 2, p. 11–22.

[45] Mockus J., 1989, Bayesian approach to global optimization, *Kluwer*.
[46] Mockus J., Eddy W., Reklaitis G., 1997, Bayesian Heuristic Approach to Discrete and Global Optimization, *Nonconvex Optimization and Its Applications*.
[47] Morgenroth E., Arvin E., Vanrolleghem P., 2002, The use of mathematical models in teaching wastewater treatment engineering, *Water Science & Technology*, p. 229–233.
[48] Murphy R. F., 2015, Building cell models and simulations from microscope images, *Methods*, vol. 96, p. 33–39.
[49] North G. R., Wang J., Genton M. G, 2011, Correlation Models for Temperature Fields, *Journal of Climate*, vol. 24, p. 5850–5862.
[50] Novak M., Horvat P., 2012, Mathematical modelling and optimisation of a waste water treatment plant by combined oxygen electrode and biological waste water treatment model, *Applied Mathematical Modelling*, vol. 36, p. 3813–3825.
[51] Niec J., Spychala M., Zawadzki P., 2016, New approach to modelling of sand filter clogging by septic tank effluent, *Journal of Ecological Engineering*, vol. 17 (2), p. 97–107.
[52] Pardalos P. M., Rosen J. B., 1987, Constrained Global Optimization: Algorithms and Applications, *Springer-Verlag Berlin, Heidelberg*, ISBN: 0-387-18095-8.
[53] Pozniak N., Sakalauskas L., Saltyte L., 2019, Kriging Model with Fractional Euclidean Distance Matrices, *Informatica*, vol. 30, nr. 2, p. 367-390.
[54] Pozniak, N., Sakalauskas, L., 2017, Fractional Euclidean distance matrices extrapolator for scattered data. Journal of Young Scientists. 47(2), 56–61.
[55] Schoenberg I. J., 1935, Remarks to Maurice Frechet's article: Sur la definition axiomatique d'une classe d'espaces vectoriels distancies applicables vectoriellement sur l'espace de Hilbert, *Ann. Math*., vol. 36, p. 724–732.

[56] Sakalauskas L., 2013, Locally Homogeneous and Isotropic Gaussian Fields in Kriging, Informatica, vol. 24, No. 2, p. 253–274.

[57] Saric M., Ek C. H., Kragic D., 2011, Dimensionality Reduction via Euclidean Distance Embeddings, School of Computer Science and Communication , TRITA-CSC-CV 2011:2 CVAP320.

[58] Secchi A. R., Cardozo N. S. M., Neto E. A., Finkler T. F., 2006, An Algorithm for Automatic Selection and Estimation of Model Parametres, *International Symposium on Advanced Control of Chemical Processes Gramado, Brazil, IFAC Proceedings*, vol. 2 (2), p. 789–794.

[59] Shapiro A., 2013, Sample Average Approximation, p. 1350–1355. 10.1007/978-1-4419-1153-7_1154.

[60] Shepard D., 1968, A Two Dimensional Interpolation Function for Irregularly Spaced Data, *Proc. ACM 23rd Nat'l Conf.*, p. 517–524.

[61] Sherman J., Morrison W. J., 1949, Adjustment of an Inverse Matrix Corresponding to Changes in the Elements of a Given Column or a Given Row of the Original Matrix (abstract), *Annals of Mathematical Statistics*, 20: 621. doi:10.1214/aoms/1177729959.

[62] Schoenberg I. J., 1935, Remarks to Maurice Frechet's article: Sur la definition axiomatique d'une classe d'espaces vectoriels distancies applicables vectoriellement sur l'espace de Hilbert," *Ann. Math*., vol. 36, p. 724–732.

[63] Weinberger K. Q, Sha F., Saul L. K., 2004, Learning a Kernel Matrix for Nonlinear Dimensionality Reduction, *Proceedings of the Twenty-first International Conference on Machine Learning* (ICML, 04), p. 106–113.

[64] Xia Q., Shi T., 2018, A cascadic multilevel optimization algorithm for the design of composite structures with curvilinear

fiber based on Shepard interpolation, *Composite Structures*, vol. 188, p. 209–219.

[65] Xiao M., Zhang G., Breitkopf P., Villon P., Zhang W., 2018, Extended Co-Kriging interpolation method based on multi-fidelity data, *Applied Mathematics and Computation*, col. 323, p. 120–131.

[66] Zaichik L., Alipchenkov V., Avetissian A., 2011, A Statistical Model for Predicting the Heat Transfer of Solid Particles in Turbulent Flows, *Flow Turbulence Combust*, vol.86, p.497–518.

[67] Zilinskas A., 1985, Axiomatic characterization of a global optimization algorithm and investigation of its search strategy, *Operations Research Letters*, vol. 4, Issue 1, p. 35–39.

ABOUT THE AUTHOR

Natalija Pozniak was born on June 1th, 1987 in Vilnius. In 2006, she graduated Vilnius Žirmūnų Gymnasium. In 2010, she received a bachelor's degree and professional qualification of a teacher in mathematics at Vilnius University. In 2012, received Master of Mathematics and professional qualification of a teacher at Vilniaus University. From 2010 till 2015 worked at travel agency. From 2015, is working as the digital studies manager and lecturer of mathematics and informatics at Vilniaus kolegija/University of Applied Sciences (VIKO). From 2014 till 2018 was a doctoral student at the VU Institute of Data Science and Digital Technologies.

SANTRAUKA

Duomenų mokslas tiria metodus, procesus ir algoritmus, skirtus aptikti žinioms ir daryti įžvalgoms iš struktūrizuotų ir nestruktūrizuotų duomenų. Jis apima įvairias matematikos, statistikos, kompiuterių mokslo sritis. Duomenų mokslas siekia surasti veiksmingą, logiškai pagrįstą modelį, skirtą prognozavimo, ekstrapoliavimo arba interpoliavimo tikslams. Tuo duomenų mokslas skiriasi nuo tradicinės analitikos ir yra artimas duomenų gavybai. Todėl duomenų mokslas leidžia generuoti veiksmingus analizės metodus, tokiose srityse, kaip medicina, socialiniai mokslai, kuriose nėra specifinių tų sričių duomenų modelių.

Surogatinis modeliavimas yra inžinerinis metodas, skirtas nagrinėti procesams, kurie gali būti stebimi tiesiogiai tik iš dalies. Tad nestebima proceso dalis yra modeliuojama surogatiniu modeliu, sukurtu pagal proceso stebėjimo rezultatus. Gana dažnai realių sistemų kompiuterinis imitavimas gali ilgai užtrukti arba reikalauti naudoti brangią ir sudėtingą įrangą. Todėl pasikartojančių uždavinių, kaip projekto optimizavimas, projekto erdvės tyrimas, jautrumo analizė ir scenarijų KAS..., JEIGU, analizė gali būti neįmanoma – jiems gali prireikti tūkstančių ar net milijono imitavimo ciklų. Vienas iš šios problemos sprendimo būdų yra aproksimuojančių modelių, žinomų kaip surogatiniai modeliai, atsako paviršiaus modeliai, metamodeliai arba emuliatoriai, imituojantys modeliuojamą objektą supaprastintu būdu, kūrimas ir taikymas.

1.1. Problemos aktualumas

Krigingo duomenų ekstrapoliavimo metodas buvo aprašytas Krige'o (1951) ir Mattherono (1963). Krigingo surogatiniai modeliai dažnai taikomi ekstrapoliavimo ir optimizavimo uždaviniams spręsti (Jones (2001), Forrester (2009), Manyu Xiao (2018) ir kt.). Krigingas yra vienas iš žinomų duomenų ekstrapoliavimo metodų, naudojamų inžinerijos ir kitose srityse (Bhosekar (2017), Carpio (2017)).

Surogatinių modelių poreikis atsiranda rekonstruojant praleistus duomenis, duomenų ekstrapoliacijoje ar planuojant optimalius eksperimentus su atsako funkcija, kuriai vertinti reikia įvairių resursų, arba kai atliktų eksperimentų ar skaičiavimų rezultatai prieinami tik iš dalies, arba prognozuojant stebėtų procesų duomenis.

Dauguma inžinerinių planavimo uždavinių reikalaujama atlikti eksperimentus, kad būtų sukonstruotos tikslo funkcijos, kaip kintamųjų funkcijos. Pavyzdžiui, tiriant lazerinės spinduliuotės poveikį optinėms medžiagoms yra atliekami kompiuteriniai to poveikio skaičiavimų rezultatai. Taip pat nuotekų valymo įrenginių efektyvumo tyrime, priklausomai nuo filtro užpildų proporcijų ar kokybės, kuri analizuojamas remiantis keliais per eksperimentus sukurtais ir realizuotais filtrais. Abiem atvejais, t. y. kompiuterinio skaičiuojamojo eksperimento arba realaus eksperimento gauti duomenys, gali būti naudojami realaus objekto surogatiniam modeliui kurti.

Matavimų taškai, priklausantys nuo daugelio parametrų, gali turėti atsitiktinį arba deterministinį pobūdį. Rezultatai priklauso nuo daugelio parametrų, tačiau šių parametrų parinkimas ir poveikis dažnai yra nežinomas. Be to, surinktų duomenų apimtys būna mažos, kadangi kiekvienas papildomas eksperimentas reikalauja laiko ir sąnaudų. Kartais eksperimentas gali būti pavojingas arba tinkamų sąlygų, reikalingų tam eksperimentui atlikti, gali pasitaikyti labai retai. O realūs uždaviniai, susiję su šiomis problemomis, gali pasitaikyti labai dažnai. Pavyzdžiui, branduolinio reaktoriaus kritinių būsenų saugumo požiūriu tyrimas esti susijęs su saugios būsenos reaktoriaus darbo tyrimu (Levenson (1981)).

Raketų arba kosminių aparatų judėjimo trajektorijų tyrimai, kuriais reikalaujama prognozuoti trajektorijas pagal stebėtus duomenis, yra gana pavojingi ir brangūs. Chemijos inžinerijos tikslas – sukurti naujas medžiagas, dažnai yra susijęs su eksperimentų planavimu pagal duomenis, kurie gauti laboratorijose po daugybės bandymų ir t. t. Tokių duomenų struktūros tiriamos tradiciniais

deterministiniais arba statistiniais metodais. Tačiau deterministinė duomenų analizė tradiciniais interpoliacijos arba ekstrapoliacijos metodais reikalauja įvairių papildomų prielaidų ir neatsižvelgia į neapibrėžtį, susijusią su duomenų rekonstravimu (Shepard (1968), Shumaker (1976)). Todėl yra aktualu išnagrinėti atsitiktinių Gauso laukų (AGL) modelių taikymą eksperimentinių duomenų analizei. AGL savybės priklauso nuo kovariacijų, aprašančių priklausomybes tarp taškų, kuriuose atlikti kompiuteriniai ar fizikiniai eksperimentai. Statistiniai AGL modeliai, kai kovariacijos yra aprašomos Euklido atstumų tarp objektų trupmeniniais laipsnių rodikliais, dar nebuvo pakankamai gerai išnagrinėti, todėl jų tyrimas bei taikymas eksperimentinių duomenų analizei, sprendžiant ekstrapoliavimo, optimizavimo ar eksperimentų planavimo uždavinius, yra aktuali problema.

1.2. Tyrimų objektas

Disertacijos tyrimų objektas yra duomenų mokslo modeliai ir metodai, skirti struktūrizuotų daugiamačių duomenų analizei. Pagrindinis dėmesys skiriamas daugiamačių duomenų surogatiniam modeliavimui pagal atstumų tarp matavimo duomenų matricų savybes.

1.3. Tyrimų tikslas ir uždaviniai

Darbo tikslas – ištirti Euklido atstumų su trupmeniniais laipsnių rodikliais savybes ir remiantis jomis sukurti surogatinį krigingo modelį bei pritaikyti jį daugiamačiams duomenims ekstrapoliuoti, eksperimentų serijoms planuoti bei daugiaekstremaliniams uždaviniams spręsti.

Siekiant šio tikslo yra sprendžiami tokie uždaviniai:

1) Trupmeninių Euklido atstumų matricų savybių tyrimas įvedus branduolio matricas.

2) Daugiamačių duomenų modelio homogeniniais ir izotropiniais Gauso laukais sudarymas.
3) Surogatinio krigingo algoritmo, paremto trupmeninių Euklido atstumų matricų savybėmis, sudarymas ir pritaikymas ekstrapoliavimui.
4) Sudaryto daugiamačio duomenų modelio pritaikymas eksperimentų serijoms planuoti.
5) Sudaryto daugiamačio duomenų modelio pritaikymas daugiaekstremalinėms funkcijoms optimizuoti.

1.4. Tyrimų metodai

Disertacijoje suformuluoti uždaviniai sprendžiami taikant daugiamatės statistikos ir kompiuterinio modeliavimo metodus. Geometrinėms trupmeninių Euklido atstumų matricų savybėms tirti yra pritaikyta veiksmų su blokinėmis matricomis teorija. Sudaryto daugiamačio duomenų modelio atsitiktiniais Gauso laukais parametrams vertinti pritaikytas didžiausio tikėtinumo metodas. Skaitmeniniam sukurto eksperimentų planavimo metodui realizuoti pritaikytas Monte Karlo metodas.

1.5. Mokslinis naujumas

Disertacijoje gauti šie rezultatai:

1) Trupmeninių Euklido atstumų matricų savybės išnagrinėtos per branduolio matricos savybes ir parodyta, kad dažniausiai trupmeninių Euklido atstumų matrica yra neišsigimusi.
2) Pasinaudojus TEAM savybėmis sukurtas daugiamačių duomenų modelis atsitiktiniais Gauso laukais.
3) Sudarytas daugiamačių duomenų krigingo algoritmas, pritaikytas ekstrapoliavimui.
4) Sudarytas algoritmas ekstremaliniams eksperimentams planuoti, skaitmeniškai realizuojamas Monte Karlo metodu.

5) Sudarytas daugiaekstremalinių uždavinių sprendimo Bajeso algoritmas pagal trupmeninių Euklido atstumų matricos savybes.

1.6. Praktinė darbo reikšmė

Disertacijoje sudarytas krigingo metodas gali būti pritaikytas išbarstytiesiems daugiamačiams duomenims (angl. *Scattered Data*) ekstrapoliuoti. Sukurtas eksperimentų planavimo metodas gali būti efektyviai panaudotas ekstremalinių eksperimentų serijoms planuoti. Disertacijoje sudarytas Bajeso optimizavimo algoritmas gali būti pritaikytas daugiaekstremaliniams uždaviniams spręsti, kai tikslo funkcijos apskaičiavimas reikalauja didelių sąnaudų.
Disertacijoje gauti šie praktiniai rezultatai:

1) Sudarytas efektyvus daugiamačių duomenų ekstrapoliavimo algoritmas.
2) Sukurtas efektyvus daugiaekstremalinių uždavinių sprendimo algoritmas, kai tikslo apskaičiavimas susijęs su didelėmis sąnaudomis.
3) Sudarytas ekstremalinių eksperimentų serijų planavimo algoritmas, kuris pritaikytas nuotekų filtrų optimaliam planavimui.

1.7. Ginamieji teiginiai

1) Trupmeninių Euklido atstumų matricų pagrindines savybes galima išreikšti per branduolio matricos savybes.
2) Sudarytas algoritmas leidžia efektyviai spręsti išbarstytųjų duomenų ekstrapoliavimo ir krigingo uždavinius.
3) Sudarytas Bajeso optimizavimo algoritmas leidžia efektyviai spręsti daugiaekstremalinius uždavinius, kai tikslo funkcijos apskaičiavimas susijęs su didelėmis sąnaudomis.
4) Sudaryti eksperimentų planavimo metodai leidžia efektyviai planuoti ekstremalinių eksperimentų serijas.

1.8. Darbo rezultatų aprobavimas

Disertacijos rezultatai pristatyti 7-se mokslinėse konferencijose bei pateikti 3 mokslinėse publikacijose.

BENDROSIOS IŠVADOS

Duomenų struktūros, aprašomos trupmeninių Euklido atstumų matricomis (TEAM), taikymuose pasitaiko gana dažnai. Statistiniai AGL modeliai, kai kovariacijos yra aprašomos Euklido atstumų tarp objektų trupmeniniais laipsnių rodikliais, dar nebuvo pakankamai gerai išnagrinėti, todėl jų tyrimas bei taikymas eksperimentinių duomenų analizei, sprendžiant ekstrapoliavimo, optimizavimo ar eksperimentų planavimo uždavinius, yra aktuali problema.

Disertacijoje gauti tokie rezultatai:

1) įrodyta, kad jeigu daugiamačių duomenų aibę sudaro skirtingi vektoriai, šios duomenų aibės TEAM branduolio matrica yra neneigiamai apibrėžta, o jos rangas vienetu mažesnis už TEAM eilę;
2) sudarytos išraiškos TEAM atvirkštinei matricai bei determinantui išreikšti per branduolio matricą;
3) sukurtas ir ištirtas daugiamačių duomenų modelis atsitiktiniais Gauso laukais, kurių kovariacijų matrica aprašoma TEAM branduolio matrica;
4) remiantis šiuo modeliu, sukurtas krigingo algoritmas, pritaikytas išbarstytiesiems daugiamačiams duomenims ekstrapoliuoti, priklausantis nuo kelių parametrų, kurių įverčiai gaunami didžiausio tikėtinumo metodu;
5) kompiuterinio modeliavimo rezultatai parodė, kad sukurtas ekstrapoliavimo metodas yra pranašesnis už Šepardo ekstrapoliatorių;
6) sudarytas Bajeso daugiaekstremalinių uždavinių sprendimo metodas, taikant atsako funkcijai modeliuoti sukurtą AGL modelį, parodant, kad šis metodas tenkina rekursyvinių lygčių sistemą;
7) pasiūlytas supaprastintas šių rekursyvinių lygčių sprendimo metodas ir sudarytas jį realizuojantis algoritmas;
8) pritaikius sukurtą AGL modelį, sudarytas ekstremalinių eksperimentų serijų planavimo metodas, pritaikant ekstremalinių

eksperimentų serijai planuoti Monte Karlo imties vidurkio aproksimacijos metodą;

9) sukurtas ekstremalinių eksperimentų metodas ištirtas kompiuterinio modeliavimo būdu ir pritaikytas nuotekų valymo filtro eksperimentų serijoms planuoti.

Gauti teoriniai ir kompiuterinio modeliavimo rezultatai leidžia daryti sekančias išvadas.

1. Sukurtas krigingo metodas priklauso tik nuo TEAM, ir nepriklauso nuo centravimo vektoriaus $\boldsymbol{s}$ ir koordinačių pradžios.

2. TEAM laipsnio rodiklis $\boldsymbol{\delta}$ yra metodo parametras, kuriam įvertinti galima taikyti mažiausių kvadratų metodą.

3. Kompiuterinio modeliavimo rezultatai parodė, kad sukurtas krigingo ekstrapoliatorius yra efektyvesnis pagal vidutinės standartinės paklaidos kriterių už Šepardo ekstrapoliatorių, dažnai taikomą išbarstytų duomenų analizei.

4. Kompiuterinio modeliavimo rezultatai parodė, kad sukurtas daugiaekstremalinio optimizavimo metodas yra efektyvesnis už kitus žinomus globalaus optimizavimo metodus (Simulated Annealing).

AUTORĖS PUBLIKACIJŲ SĄRAŠAS DISERTACIJOS TEMA

1. *Pozniak, N., Sakalauskas, L. (). New Bayesian method for multiextremal optimization, Computational Methods and Techniques, KU, (under review).*
2. Pozniak N., Sakalauskas L., Saltyte L. (2019). Kriging Model with Fractional Euclidean Distance Matrices. Informatica (Clarivate Analytics, Scopus), vol. 30, nr. 2, p. 367-390, http://dx.doi.org/10.15388/Informatica.2019.210.
3. Pozniak, N., Sakalauskas, L. (2019). The method for the optimal experiment design. Proceedings of International Scientific Conference Contemporary Issues in Business, Management and Economics Engineering, VGTU, ISBN 978-609-476-161-4 https://doi.org/10.3846/cibmee.2019.012.
4. Pozniak N., Sakalauskas L., (2017). Fractional Euclidean Distance Matrices Extrapolator for Scattered Data. Jaunųjų mokslininkų darbai. (Index Copernicus, CEEOL), nr. 2 (47), p. 56–61, (ISSN 1648-8776), https://doi.org/10.21277/jmd.v47i2.156.

TRUMPAI APIE AUTORĘ

Natalija Pozniak Vilniaus universitete įgijo matematikos bakalauro (2006 m.) ir matematikos magistro (2012 m.) laipsnius. Nuo 2015 metų iki šiol dirba ir dėsto Vilniaus kolegijoje. 2014-2018 metais studijavo Vilniaus universiteto Matematikos ir informatikos instituto doktorantūroje (Gamtos mokslai, informatika).

NOTES

Natalija Pozniak

STUDY AND APPLICATION OF EUCLIDEAN MATRICES TO SURROGATE MODELLING

Summary of Doctoral Dissertation
Natural Sciences
Informatics (N 009)
Editor Zuzana Šiušaitė

Natalija Pozniak

EUKLIDO MATRICŲ TYRIMAS IR TAIKYMAS SUROGATINIAM MODELIAVIMUI

Daktaro disertacijos santrauka
Gamtos mokslai
Informatika (N 009)
Redaktorė Jorūnė Rimeisytė - Nekrašienė

Vilniaus universiteto leidykla
Universiteto g. 1, LT-01513 Vilnius
El. p. info@leidykla.vu.lt,
www.leidykla.vu.lt
Tiražas 45 egz.