

# Kriging Model with Fractional Euclidean Distance Matrices

Natalija POZNIAK<sup>1</sup>, Leonidas SAKALAUSKAS<sup>1\*</sup>, Laura SALTYTE<sup>2</sup>

<sup>1</sup>*Institute of Data Science and Digital Technologies of Vilnius University,  
Akademijos str. 4, LT04812 Vilnius, Lithuania*

<sup>2</sup>*Klaipeda University, H. Manto str. 84, Klaipeda, Lithuania  
e-mail: natalija.pozniak@gmail.com, leonidas.sakalauskas@mii.vu.lt, laura.saltyte@ku.lt*

Received: August 2018; accepted: February 2019

**Abstract.** The multidimensional data model for kriging is developed using fractional Euclidean distance matrices (FEDM). The properties of FEDM are studied by means of the kernel matrix method. It has been shown that the factorization of kernel matrix enables us to create the embedded set being a nonsingular simplex. Using the properties of FEDM the Gaussian random field (GRF) is constructed doing it without positive definite correlation functions usually applied for such a purpose. Created GRF can be considered as a multidimensional analogue of the Wiener process, for instance, line realizations of this GRF are namely Wiener processes. Next, the kriging method is developed based on FEDM. The method is rather simple and depends on parameters that are simply estimated by the maximum likelihood method. Computer simulation of the developed kriging extrapolator has shown that it outperforms the well known Shepard inverse distance extrapolator. Practical application of the developed approach to surrogate modelling of wastewater treatment is discussed. Theoretical investigation, computer simulation, and a practical example demonstrate that the proposed kriging model, using FEDM, can be efficiently applied to multidimensional data modelling and processing.

**Key words:** scattered data, fractional Euclidean distance matrices, multivariate normal distribution, homogeneous Gaussian field, maximum likelihood.

## 1. Introduction

A lot of computational models applied in the business and industry, using multidimensional data, obtained by measuring, sometimes require to fulfil expensive simulations. In these cases, 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

---

\* Corresponding author.

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 first step in kriging is to create a stochastic model that likely best describes the set of observed data. The assumption is usually made that the sample values are sufficiently homogeneous, as well as the correlation between two random points solely depends on the distance between them, but it is independent of their location. In this way, the stochastic homogeneous and isotropic model for kriging is constructed using the correlation functions that have to obey the conditions of positive definiteness according to Bochner's theorem (Abrahamsen, 1997). Following this approach, many models in spatial statistical research and applications are developed using the Whittle-Matern correlation family (Abrahamsen, 1997; Guttorp and Gneiting, 2005), Gneiting correlation class (Guttorp and Gneiting, 2005), etc. In this paper, the application of Euclidean distance matrices (EDM) to multivariate data modelling and kriging is considered, doing it without the standard way based on positive definite correlation functions.

EDM's have received increased attention due to their applications in recently active fields of research, such as molecular conformation in bioinformatics, dimensionality reduction in machine learning and statistics, semidefinite programming, wireless sensor network localization (Pham *et al.*, 2008; Ghiasi *et al.*, 2018; Koo *et al.*, 2018; Qian *et al.*, 2008). However, in recent years square EDM's are mostly studied in literature (Schoenberg, 1935; Gower, 1984; Weinberger *et al.*, 2004), although the Euclidean geometry in the real world is dealing with a square root, i.e. the fractional degree of distance squares. Moreover, fractional degrees of distance squares appear in some applications as well, however many fundamental issues about their properties remain unstudied. Therefore, this work focuses on the geometry of Fractional Euclidean Distance Matrices (FEDM) and their application in the multidimensional data analysis using kriging in the context of computer modelling and solving of practical examples. Sakalauskas (2013) considered a random field with a positive definite correlation function tending its scale parameter to zero as well as infinitely increasing the variance of field, and in this way has derived that the limiting kriging method relates with inverse FEDM's (see, also Remark 1 below). Usually EDM's are studied using the Gram matrix of the observed data set, called a kernel matrix (Gower, 1984, etc.). Using the kernel matrix method (Pozniak and Sakalauskas, 2017) has shown FEDM with a strongly fractional degree of distance squares to be nonsingular, and, consequently, the derived kriging method in can be used correctly. In this paper, the multinormal data model and the kriging method are developed taking covariances proportional to the elements of kernel matrix, and, thus, avoiding the standard way, based on positive definite correlation functions. Once constructed, the kriging model can be used for data modelling, prediction, and optimization. The objective is to discuss and demonstrate the applicability of the kriging technique, based on FEDM, and to verify the computational efficiency of the developed approach.

## 2. Geometrical Properties of FEDM

Let us assume that a data set  $X = (x_1, x_2, \dots, x_K)$  of  $K$  vectors  $x_i \in \mathbb{R}^d$ ,  $1 \leq i \leq K$ ,  $d \geq 1$  be given, at which the values of some response function  $Y = (y_1, y_2, \dots, y_K)^\top$  are used, obtained by physical measurement, computer simulation, etc. Denote the  $K \times K$  matrix of fractional degrees  $\delta$  of Euclidean distance squares among pairs of vectors by

$$A = [(\|x_i - x_j\|)^\delta]_{i,j=1}^K,$$

where  $\|x_i - x_j\| = (x_i - x_j)^\top \cdot (x_i - x_j)$ ,  $x_i \in \mathbb{R}^d$ ,  $1 \leq i, j \leq K$ ,  $0 \leq \delta \leq 1$ . Note that  $\delta = 1/2$  in a special case of usual Euclidean distances. Recently the properties of matrices of Euclidean distance squares are studied most, i.e. as  $\delta = 1$  (Schoenberg, 1935; Gower, 1984; Weinberger *et al.*, 2004). Usually the main tool for EDM study is the Gram matrix of data set  $X$ , called a kernel matrix. Thus, following this approach, the matrix

$$F = -(I - E \cdot s^\top) \cdot A \cdot (I - s \cdot E^\top) \tag{1}$$

is introduced, called hereinafter a kernel matrix, where  $I$  denotes  $K \times K$  unit matrix,  $E$  denotes  $K$ -dimensional vector-column of units,  $s \in \mathbb{R}^K$  is a certain vector-column.

**Proposition 1.** *Let  $x_i, x_j \in X$ ,  $x_i \neq x_j$ , if  $i \neq j$ ,  $1 \leq i, j \leq K$ ,  $0 \leq \delta < 1$ ,  $s^\top \cdot E = 1$ . Then, kernel matrix (1) is positively semi-definite of rank  $K - 1$ .*

The proof can be seen in Pozniak and Sakalauskas (2017), Theorem 1. Note that, if  $\delta = 1$ , then the rank of kernel matrix can be less  $K - 1$ .

**Proposition 2.** *Assume  $A$  is a FED matrix,  $s_1, s_2 \in \mathbb{R}^K$ ,  $s_1^\top \cdot E = 1$ ,  $s_2^\top \cdot E = 1$ . Then,*

$$F_1 = -(I - E \cdot s_1^\top) \cdot A \cdot (I - s_1 \cdot E^\top) = (I - E \cdot s_1^\top) \cdot F_2 \cdot (I - s_1 \cdot E^\top),$$

where

$$F_2 = -(I - E \cdot s_2^\top) \cdot A \cdot (I - s_2 \cdot E^\top).$$

The proposition implied by an easily verified equality of idempotence as

$$(I - E \cdot s_1^\top) \cdot (I - E \cdot s_2^\top) = (I - E \cdot s_1^\top).$$

Let us consider FEDM geometrical properties more.

**Proposition 3.** *Under the conditions of Proposition 1, the embedded set  $Z = (z_1, z_2, \dots, z_K)^\top$  exists such that  $z_K \in \mathbb{R}^{K-1}$ ,  $Z$  is a full rank matrix, and*

$$A = \frac{1}{2} \cdot [(z_i - z_j)^\top \cdot (z_i - z_j)]_{i,j=1}^K.$$

*Proof.* Meanwhile, according to Proposition 1, one can get  $F = \tilde{Z} \cdot \tilde{Z}^\top$  by factorization of the kernel matrix. Then, the embedded set is obtained taking first  $K - 1$  columns of set  $\tilde{Z}$ , the rank of which strictly  $K - 1$ .

Note that (1) implies

$$\text{diag}(F) = 2 \cdot A \cdot s - s^\top \cdot A \cdot s \cdot E.$$

Next, by means of the latter formula after simple manipulations one can make sure that the proposition is true, because:

$$\begin{aligned} & \frac{1}{2} \cdot [(z_i - z_j)^\top \cdot (z_i - z_j)]_{i,j=1}^K, \\ &= \frac{[z_i^\top \cdot z_i]_{i,j=1}^K + [z_j^\top \cdot z_j]_{i,j=1}^K}{2} - [z_i^\top \cdot z_j]_{i,j=1}^K \\ &= \frac{\text{diag}(F) \cdot E^\top + E \cdot \text{diag}(F)^\top}{2} - F \\ &= E \cdot s^\top \cdot A + A \cdot s \cdot E^\top - s^\top \cdot A \cdot s \cdot E \cdot E^\top + (I - E \cdot s^\top) \cdot A \cdot (I - s \cdot E^\top) \\ &= A. \end{aligned}$$

Hence, the embedded set presents itself as a nonsingular  $K - 1$ -dimensional simplex.  $\square$

**Corollary 1.** *Let us consider the factorization  $F = Z \cdot Z^\top$  of the kernel matrix  $F$ , constructed according to assumptions of Proposition 1, where  $Z$  is the embedded set. Assume  $s_c^\top \cdot E = 1$ ,  $s_c \in \mathbb{R}^K$ , and denote  $z_c = Z^\top \cdot s_c$ . Then, according to Proposition 2:*

$$\begin{aligned} F &= -(I - E \cdot s_c^\top) \cdot A \cdot (I - s_c \cdot E^\top) = (I - E \cdot s_c^\top) \cdot F \cdot (I - s_c \cdot E^\top) \\ &= (I - E \cdot s_c^\top) \cdot Z \cdot Z^\top \cdot (I - s_c \cdot E^\top) = (Z - E \cdot z_c^\top) \cdot (Z^\top - z_c \cdot E^\top). \end{aligned}$$

Hence, Corollary 1 shows us the geometrical sense of vector  $s$ , that displays the origin of coordinates to the point  $z_c$  in the embedded space. Therefore, if  $s^\top \cdot E = 1$ , it is reasonable to call  $s$  a centering vector.

The next corollary follows from (1).

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

$$\begin{aligned} F_{i,j} &= A_{i,i} + A_{j,j} - A_{i,j}, \quad 1 < i, j \leq K, \\ F_{0,0} &= 0, \quad F_{1,j} = 0, \quad F_{i,1} = 0, \quad 1 < i, j \leq K. \end{aligned}$$

**Corollary 3.** *Under the conditions of Proposition 1, the centering vector  $s = E/K$  minimizes the trace of kernel matrix.*

*Proof.* Indeed, after simple manipulations it follows:

$$\begin{aligned} \text{tr}(F) &= \text{tr}((Z - E \cdot z_c^\top) \cdot (Z^\top - z_c \cdot E^\top)) \\ &= \text{tr}((I - E \cdot s^\top) \cdot Z \cdot Z^\top \cdot (I - s \cdot E^\top)) \\ &= -\text{tr}((I - E \cdot s^\top) \cdot A \cdot (I - s \cdot E^\top)) = 2 \cdot s^\top \cdot A \cdot E - K \cdot s^\top \cdot A \cdot s. \end{aligned}$$

By differentiating the latter expression with respect to  $s$  and equating the obtained derivative to zero, one can be sure of the truth of corollary.  $\square$

**Corollary 4.** *If the conditions of Proposition 1 are valid, then  $E^\top \cdot A^{-1} \cdot E \neq 0$ , besides, the centering vector*

$$s = \frac{A^{-1} \cdot E}{E^\top \cdot A^{-1} \cdot E}$$

*displays the origin of coordinates in the centre of the circumscribed sphere of the embedded simplex.*

*Proof.* Note that the diagonal of  $F$  consists of squares of distances from vertices of the embedded simplex to the origin in the embedded space. By means of (1), one can easily verify that  $\text{diag}(F) = 2 \cdot A \cdot s - (s^\top \cdot A \cdot s) \cdot E$ . Since the distances from vertices to the centre of the circumscribed sphere should be the same, find the centering vector  $s$ , i.e. which obeys the condition  $2 \cdot A \cdot s - (s^\top \cdot A \cdot s) \cdot E = r \cdot E$ , where  $r$  is the radius of the circumscribed sphere. Thus, the latter condition implies that

$$s = \frac{r + v}{2} \cdot A^{-1} \cdot E,$$

because the inverse of matrix  $A$  exists, as proved in Pozniak and Sakalauskas (2017), Theorem 2, and, therefore, it can be used correctly. Then, using the normalization condition one can derive

$$\frac{r + v}{2} \cdot E^\top \cdot A^{-1} \cdot E = 1.$$

The latter equality means 1) that  $E^\top \cdot A^{-1} \cdot E \neq 0$ , 2) it helps us to establish:

$$\frac{r + v}{2} = \frac{1}{E^\top \cdot A^{-1} \cdot E},$$

and finally,

$$r = \frac{1}{E^\top \cdot A^{-1} \cdot E}.$$

□

Let us consider the inverse and determinant of FEDM more in detail. Denote by  $(x_0, X)$  the set obtained adding some point  $x_0 \in \mathbb{R}^d$  to the observation set  $X$ , as well as the respective vector of fractional degrees of Euclidean distance squares between  $x_0$  and  $X$  by

$$a = (\|x_1 - x_0\|^\delta, \|x_2 - x_0\|^\delta, \dots, \|x_K - x_0\|^\delta)^\top.$$

Now, define the extended kernel matrix, respective to FEDM  $\tilde{A} = \begin{bmatrix} 0 & a^\top \\ a & A \end{bmatrix}$  of the set  $(x_0, X)$ , calculated at the centering vector  $\tilde{s}^\top = (s', s^\top)$  and presented in the block matrix:

$$\tilde{F} = \begin{bmatrix} v & f^\top \\ f & F \end{bmatrix} \equiv - \begin{bmatrix} 1 - s' & -s^\top \\ -s' \cdot E & S \end{bmatrix} \cdot \begin{bmatrix} 0 & a^\top \\ a & A \end{bmatrix} \cdot \begin{bmatrix} 1 - s' & -s' \cdot E^\top \\ -s & S^\top \end{bmatrix}, \quad (2)$$

where, due to (1):

$$S = I - E \cdot s^\top, \quad (3)$$

$$F = s' \cdot E \cdot a^\top \cdot S^\top + s' \cdot S \cdot a \cdot E^\top - S \cdot A \cdot S^\top, \quad (4)$$

$$f = S \cdot A \cdot s - s' \cdot E \cdot a^\top \cdot s - S \cdot a \cdot (1 - s'), \quad (5)$$

$$v = 2 \cdot (1 - s') \cdot s^\top \cdot a - s^\top \cdot A \cdot s. \quad (6)$$

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

**Theorem 1.** *Assume FEDM*

$$A = [((x_i - x_j)^\top \cdot (x_i - x_j))^\delta]_{i,j=1}^K,$$

*the vector of fractional degrees of distances*

$$a = \{(x_i - x_0)^\top \cdot (x_i - x_0)^\delta\}_{k=1}^K$$

*as well as the centering vector  $(s', s)$  to be given, where*

$$x_i, x_j \in \mathbb{R}^d, \quad x_i \neq x_j, \quad i \neq j, \quad 0 \leq i, j \leq K, \quad K > 1,$$

$$0 \leq \delta < 1, \quad d \geq 1, \quad s \in \mathbb{R}^K, \quad s' + E^\top \cdot s = 1.$$

*Then the inverse of  $K \times K$  lower-right submatrix of kernel matrix (2) is as follows:*

$$F^{-1} = M = \frac{A^{-1} \cdot E \cdot E^\top \cdot A^{-1}}{E^\top \cdot A^{-1} \cdot E} - A^{-1} + \frac{q \cdot q^\top}{D}, \quad (7)$$

where

$$q = s + s' \cdot A^{-1} \cdot \left( a + E \cdot \frac{1 - E^T \cdot A^{-1} \cdot a}{E^T \cdot A^{-1} \cdot E} \right), \quad (8)$$

$$D = (s')^2 \cdot \left( a^T \cdot A^{-1} \cdot a - \frac{(1 - E^T \cdot A^{-1} \cdot a)^2}{E^T \cdot A^{-1} \cdot E} \right), \quad (9)$$

$$s' = 1 - E^T \cdot s. \quad (10)$$

Besides, the determinants of matrices  $F$  and  $A$  are related

$$|F| = (-1)^{K+1} \cdot |A| \cdot D \cdot E^T \cdot A^{-1} \cdot E. \quad (11)$$

**Corollary 5.** Suppose the conditions of Theorem 1 to be satisfied. Then,

$$F^{-1} - \frac{F^{-1} \cdot E \cdot E^T \cdot F^{-1}}{E^T \cdot F^{-1} \cdot E} = -A^{-1} + \frac{A^{-1} \cdot E \cdot E^T \cdot A^{-1}}{E^T \cdot A^{-1} \cdot E}.$$

The proof of Theorem 1 and Corollary 5 is given in Appendix. It is of interest that the right part of conclusion of Corollary 5 is invariant to the choice of the centering vector  $s$  and point  $x_0$ .

### 3. Creation of the Gaussian Random Field Using FEDM

Statistical approach for modelling the response function by a random field has the origin in the works of Kushner (1964), Žilinskas (1985), and Mockus (1989). In this section, a probabilistic model of response surface (scalar function) is developed, the values of which are collected by observation, physical measurements or computer simulation, etc. Due to the fact that there are no further data except the measurement performance at the experimental points, the surface which represents the objective response function can be designed as a homogeneous Gaussian random field (GRF)  $Z(x, \omega)$ , which for each point in the variable space  $x \in \mathbb{R}^d$ , is a measurable function of random event  $\omega \in (\Omega, \Sigma, P)$  in some probability space (Mockus, 1989; Jones, 2001; Sakalauskas, 2013). To model the observed data in a probabilistic way, one has to define the probability distribution of response surface values  $Z(x, \omega) = (Z(x_1, \omega), Z(x_2, \omega), \dots, Z(x_K, \omega))$  in the given set of points  $X = (x_1, x_2, \dots, x_K)$ . Consider this probability distribution with the constant mean vector:

$$EZ(x, \omega) = \mu \cdot E, \quad (12)$$

and covariance matrix

$$E(Z(x, \omega) - \mu \cdot E) \cdot (Z(x, \omega) - \mu \cdot E)^T = \beta^2 \cdot F, \quad (13)$$

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

Taking into account geometrical properties of FEDM, explored in the previous section, choose matrix  $F$  as a lower-right submatrix of kernel matrix (2), calculated using some point  $x_0$  and centering vector  $(s', s)$  such that  $s' + E^T \cdot s = 1$ ,  $s' \neq 0$ . Note that matrix  $F$  is positively defined according to Proposition 1.

Thus, the multivariate Gaussian density is written down as well:

$$p_{x_0, X}(Y) = \frac{\exp\left[-\frac{(Y-\mu \cdot E)^T \cdot F^{-1} \cdot (Y-\mu \cdot E)}{2 \cdot \beta^2}\right]}{(2\pi)^{d/2} \cdot \beta^d \cdot |F|^{1/2}}. \quad (14)$$

However, as one can easily see that this density, using denotations of the previous section and (7)–(11), can be rewritten in an equivalent manner:

$$p_{x_0, X}(Y) = \frac{\exp\left[-\frac{1}{2 \cdot \beta^2} \cdot Y^T \left(A^{-1} - \frac{A^{-1} \cdot E \cdot E^T \cdot A^{-1}}{E^T \cdot A^{-1} \cdot E}\right) \cdot Y - \frac{(Y^T \cdot q - \mu)^2}{2 \cdot \beta^2 \cdot D}\right]}{(2\pi)^{K/2} \cdot \beta^K \cdot ((-1)^{K+1} \cdot |A| \cdot D \cdot E^T \cdot A^{-1} \cdot E)^{1/2}}. \quad (15)$$

In order to describe the random field by its finite-dimensional (cumulative) distributions, latter ones have to obey the consistency conditions of symmetry and compatibility (Abrahamsen, 1997). This is done by the next theorem under the appropriate choice of centering vector.

**Theorem 2.** *The random Gaussian field  $Z(x, \omega)$  exists in some probability space  $(\Omega, \Sigma, P)$ , having the functions*

$$p_{x_0, X_K}(Y_K) = \frac{\exp\left[-\frac{1}{2 \cdot \beta^2} \cdot Y_K^T \left(A_K^{-1} - \frac{A_K^{-1} \cdot E_K \cdot E_K^T \cdot A_K^{-1}}{B_K}\right) \cdot Y_K - \frac{(Y_K^T \cdot A_K^{-1} \cdot (a_K + E_K \cdot \frac{C_K}{B_K}) - \mu)^2}{2 \cdot \beta^2 \cdot (D_K - \frac{C_K^2}{B_K})}\right]}{(2\pi)^{K/2} \cdot \beta^K \cdot ((-1)^{K+1} \cdot |A_K| \cdot (D_K \cdot a_K \cdot B_K - C_K^2))^{1/2}} \quad (16)$$

as its finite-dimensional densities of distribution of  $Y_K = (y_1, y_2, \dots, y_K)^T$ ,  $y_i = Z(x_i, \omega)$ , where  $B_K = E_K^T \cdot A_K^{-1} \cdot E_K$ ,  $C_K = 1 - E_K^T \cdot A_K^{-1} \cdot a_K$ ,  $D_K = a_K^T \cdot A_K^{-1} \cdot a_K$ ,  $X_K = (x_1, x_2, \dots, x_K)$  is a sequence of mutually disjoint points  $x_i \in \mathbb{R}^d$ , and disjoint with  $x_0 \in \mathbb{R}^d$  as well,  $1 \leq i \leq K$ ,

$$A_K = \left[ (\|x_i - x_j\|)^\delta \right]_{i,j=1}^K,$$

$$a_K = (\|x_1 - x_0\|^\delta, \|x_2 - x_0\|^\delta, \dots, \|x_K - x_0\|^\delta)^T,$$

$E_K = (1, 1, \dots, 1)^T$ ,  $E_K \in \mathbb{R}^K$ ,  $K = 2, 3, \dots$ ,  $\mu$  and  $\beta$  are parameters,  $\beta > 0$ .

*Proof.* One can notice the density (16) follows from (14), and, consequently, (15), taking as the lower-right submatrix of kernel matrix (2), calculated with the centering vector



$\tilde{s} = (1, 0, 0, \dots, 0)$ . As noted above, density (14), and (15), (16) respectively, are defined correctly due to Proposition 1. Indeed,  $\int_{Y \in \mathbb{R}^K} p_{x_0, X}(Y) dY = 1$ .

Next, the symmetry of distributions (14) and, consequently, that of (15) and (16), is shown in the standard way for multinormal distributions using the symmetry of kernel matrix (see Section 1.4.1, Abrahamsen, 1997).

Proving the compatibility, denote the covariance matrices respective to  $X_K$  and  $X_{K+1}$  by  $F_K$  and  $F_{K+1}$ , that are lower-right submatrices in the respective kernel matrices. Now, one can easily notice that, due to Corollary 2,  $F_K$  is the upper-left submatrix in the decomposition

$$F_{K+1} = \begin{pmatrix} F_K & f_K \\ f_K^\top & v_K \end{pmatrix}.$$

Thus, using the standard block matrix operations, one can make sure that:

$$p_{x_0, X_{K+1}}(Y_{K+1}) = p_{x_0, X_K}(Y_K) \exp \left[ \frac{-(Y_{K+1} - E_{x_0, X_K}(Z(x_{K+1}, \omega)|Y_{K\cdot}))^2}{2 \cdot D_{x_0, X_K}^2(Z(x_{k+1}, \omega)|Y_{K\cdot})} \right] \\ \times (2\pi \cdot D_{x_0, X_K}^2(Z(x_{k+1}, \omega)|Y_K))^{-1/2},$$

where

$$E_{x_0, X_K}(Z(x_{K+1}, \omega)|Y_{K\cdot}) = \mu + (Y_K - \mu \cdot E_K)^\top \cdot F_K^{-1} \cdot f_K$$

and

$$D_{x_0, X_K}^2(Z(x_{k+1}, \omega)|Y_{K\cdot}) = \beta^2 \cdot (v_K - f_K^\top \cdot F_K^{-1} \cdot f_K)$$

are, respectively, a posteriori mean and variance of GRF at the point  $x_{k+1}$  under known  $Y_K$  (Casella and Berger, 2002). Then the compatibility follows as well:

$$p_{x_0, X_K}(Y_K) = \int_{-\infty}^{\infty} p_{x_0, X_{K+1}}(Y_K, y) dy.$$

However, according to Kolmogorov’s consistency theorem, GRF exists having (16) as its finite-dimensional distributions (Khoshnevisan, 2002). □

EXAMPLE. Assume that the observation set is displayed on a line:  $x_k = a + t_k \cdot (b - a)$ ,  $a, b \in \mathbb{R}^d$ ,  $t_0 < t_1 < \dots < t_K$ . If  $\delta = 1/2$ , then GRF is distributed on this line as a Wiener process.

Actually, FEDM of the considered set on the line is as follows:  $A = |b - a| \cdot [|t_i - t_j|]_{i,j=1}^K$ . One can easily see that, according to Corollary 2, the lower-right submatrix of the kernel matrix, being the covariance matrix of the considered set, is the same

as that of the well-known Wiener process:

$$\begin{aligned} F &= \beta^2 \cdot |b - a| \cdot [t_i + t_j - 2t_0 - |t_i - t_j|]_{i,j=1}^K \\ &= 2\beta^2 \cdot |b - a| \cdot [\min(t_i - t_0, t_j - t_0)]_{i,j=1}^K. \end{aligned}$$

#### 4. Kriging by FEDM

The stochastic model of kriging should incorporate uncertainty about quantities in unobserved points and to quantify the uncertainty associated with the kriging estimator. Kriging gives us a way of anticipating, with some probability, a result associated with values of the parameters that have never been met before, or have been lost, to “store” the existing information (the experimental measurements), and propagate it to any situation where no measurement has been made. As it is unknown which of all function variables will be preponderant, consider them as equivalent, and thus calculate a distance between the measurement points, which now are symmetric with respect to the miscellaneous variables. Suppose that we have to predict the value of response surface  $y$  at some point  $x \in \mathbb{R}^d$ , if the set  $X = (x_1, x_2, \dots, x_K)$  of observed mutually disjoint vectors  $x_i \in \mathbb{R}^d$ ,  $1 \leq i \leq K$ ,  $K > 1$ ,  $d \geq 1$ , is fixed and the data of measurement  $Y = (y_1, y_2, \dots, y_K)^\top$  of the response surface at points of  $X$  be known. Let the data be modelled by the probabilistic Gaussian model with constant mean (12) and the covariance matrix of type (13). Then, decompose the covariance matrix  $\hat{\beta}^2 \cdot \bar{F}$  of vector  $\bar{Y} = (Y, y)$  in to the covariance matrix  $\hat{\beta}^2 \cdot F$  of the vector  $Y$ , to vector  $\hat{\beta}^2 \cdot f$  of covariances between vector  $Y$  and  $y$ , and to variance of  $y$ , denoted by  $\nu$ :

$$\hat{\beta}^2 \cdot \bar{F} = \hat{\beta}^2 \cdot \begin{pmatrix} F & f \\ f & \nu \end{pmatrix}. \quad (17)$$

According to the introduction of kriging (see Jones, 2001) it is defined by the kriging predictor

$$y(x) = Y^\top \cdot F^{-1} \cdot \left( f + E \cdot \frac{(1 - E^\top \cdot F^{-1} \cdot f)}{E^\top \cdot F^{-1} \cdot E} \right), \quad (18)$$

a posteriori variance of the predictor

$$\beta^2(x) = \hat{\beta}^2 \cdot \left( \nu - f^\top \cdot F^{-1} \cdot f + \frac{(1 - E^\top \cdot F^{-1} \cdot f)^2}{E^\top \cdot F^{-1} \cdot E} \right), \quad (19)$$

and the MLE parameter of variance

$$\hat{\beta}^2 = \frac{1}{K} \cdot \left( Y^\top \cdot F^{-1} \cdot Y - \frac{(Y^\top \cdot F^{-1} \cdot E)^2}{E^\top \cdot F^{-1} \cdot E} \right). \quad (20)$$

Let us consider the kriging method based on the probabilistic data model, derived in the previous section.

**Theorem 3.** Assume the set  $X = (x_1, x_2, \dots, x_K)$  of mutually disjoint points  $x_i, x_j \in \mathbb{R}^d$ ,  $x_i \neq x_j$ ,  $i \neq j$ ,  $1 \leq i, j \leq K$ , be given, at which the values  $Y = (y_1, y_2, \dots, y_K)^\top$  of some GRF realization are known, namely,  $y_i = Z(x_i, \omega)$ . Let

$$A = \left[ \left( (x_i - x_j)^\top \cdot (x_i - x_j) \right)^\delta \right]^{K \times K}$$

be FEDM of the vectors  $x_i \in X$ ,  $0 \leq \delta < 1$ .

Then the kriging predictor at  $x \in \mathbb{R}^d$  is as follows:

$$y(x) = Y^\top \cdot A^{-1} \cdot \left( a + E \cdot \frac{(1 - E^\top \cdot A^{-1} \cdot a)}{E^\top \cdot A^{-1} \cdot E} \right), \quad (21)$$

its variance

$$\beta^2(x) = \hat{\beta}^2 \cdot \left( a^\top \cdot A^{-1} \cdot a - \frac{(1 - E^\top \cdot A^{-1} \cdot a)^2}{E^\top \cdot A^{-1} \cdot E} \right), \quad (22)$$

where MLE of the variance parameter

$$\hat{\beta}^2 = \frac{(Y^\top \cdot A^{-1} \cdot E)^2}{E^\top \cdot A^{-1} \cdot E} - Y^\top \cdot A^{-1} \cdot Y. \quad (23)$$

Proof is given in Appendix.

Thus, kriging predictor (21) turns out to a linear extrapolator  $y(x) = Y^\top \cdot u(x)$ , with the extrapolation weights:

$$u(x) = A^{-1} \cdot \left( a + E \cdot \frac{(1 - E^\top \cdot A^{-1} \cdot a)}{E^\top \cdot A^{-1} \cdot E} \right), \quad (24)$$

and variance (22) helps us the measure of accuracy of extrapolation.

The next property of the kriging extrapolator easily follows from (24).

**Corollary 6.** The weights in (24) satisfy the condition:  $E^\top \cdot u(x) = 1$ .

Besides, it is easy to make sure that kriging predictor (13) coincides with the values of response at measured points.

**REMARK 1.** The kriging model has also been derived by Sakalauskas (2013) studying the GRF with the positive defined covariance function of the shape

$$\frac{\gamma^\delta \cdot \beta^2}{\alpha} \cdot \left( 1 - \alpha \cdot \left( \frac{r}{\gamma} \right)^\delta + O\left( \frac{r}{\gamma} \right)^{2 \cdot \delta} \right)$$

and tending the parameter  $\gamma$  to infinity.

REMARK 2. Note that the degree  $\delta$  is a perfect parameter of GRF as well, which can be estimated using the observation data. The least square estimate  $\hat{\delta}$  is estimated by the univariate minimization of variance parameter MLE (23):

$$\hat{\delta} = \arg \min_{0 \leq \delta \leq 1} \frac{1}{K} \cdot \left( \frac{(Y^T \cdot A^{-1} \cdot E)^2}{E^T \cdot A^{-1} \cdot E} - Y^T \cdot A^{-1} \cdot Y \right). \quad (25)$$

The properties of this estimate are explored by computer simulation in section 5.

### 5. Comparison of the Developed Kriging Extrapolator with the Shepard Extrapolator

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

$$y_{Shepard}(x) = Y^T \cdot u(x),$$

where

$$u(x) = \begin{cases} \frac{w(x)}{E^T \cdot w(x)}, & \text{if } |x_i - x| \neq 0 \forall i, \\ y_i, & \text{if } |x_i - x| = 0, \end{cases} \quad (26)$$

the weights are chosen in the following way:

$$w(x) = \left( \frac{1}{|x_1 - x|^\delta}, \frac{1}{|x_2 - x|^\delta}, \dots, \frac{1}{|x_K - x|^\delta} \right). \quad (27)$$

It is easy to see that the Shepard extrapolator also satisfies the condition  $E^T \cdot u(x) = 1$ .

A set of analytic test functions was chosen aiming to compare the developed kriging predictor with the Shepard extrapolator (Table 1).

At the beginning of the experiments, the behaviour of 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). The first test function is a Branin function (fourth-order polynomials) and it shows a dominant second-order trend. This function has an extremely complex and highly nonlinear behaviour. The second test function is a Linear function composed of polynomials and trigonometric functions, and it shows a strong first-order trend. The Rosenbrock function shows both the first and second-order trends.

Each test function is shown in Table 1. The nonlinear Haupt function is composed of trigonometric functions. The test function results indicate a second-order trend of both independent variables  $x$  and  $y$ .

Each test function is plotted (see Figs. 1–6).

Table 1  
Types of test functions and test function domains.

Branin

$$z(x, y) = \left(y - \frac{5x^2}{4\pi^2} + \frac{5x}{\pi} - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right)\cos x + 10, \quad (28)$$

$$x \in [-5, 10], y \in [0, 15].$$

Rosenbrook

$$z(x, y) = 100(y - x^2)^2 + (1 - x)^2, \quad (29)$$

$$x \in [-5, 5], y \in [-5, 5].$$

Rastrigin

$$z(x, y) = 20 + x^2 + y^2 - 10(\cos(2\pi x) + \cos(2\pi y)), \quad (30)$$

$$x \in [-5.12, 5.12], y \in [-5.12, 5.12].$$

Haupt

$$z(x, y) = x \sin(4x) + 1.1y \sin(2y), \quad (31)$$

$$x \in [0, 4], y \in [0, 4].$$

Himmenblau

$$z(x, y) = (x^2 + y - 11)^2 + (x + y^2 - 7)^2 + 0.1((x - 3)^2 + (y - 2)^2) \quad (32)$$

$$x \in [-5, 10], y \in [-5, 10].$$

Linear

$$z(x, y) = x^2 + x^2 \cos(x) + y \cos(y), \quad (33)$$

$$x \in [1, 2], y \in [1.5, 3].$$

Model parameters  $\beta^2$  and  $\delta$  have been estimated by the maximum likelihood method (23) and (25). The computer simulation experiment has been performed generating  $N = 200$  samples of  $K = 20$  randomly simulated points for each test function in its domain.

In the prediction problem, the value of the Euclidean fraction degree  $\delta$  could be chosen so as to get the best model results. An elegant alternative is to calculate the optimal  $\delta$  value according to (25).

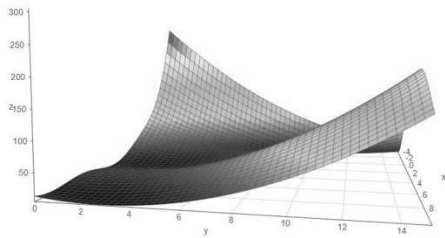


Fig. 1. Branin function.

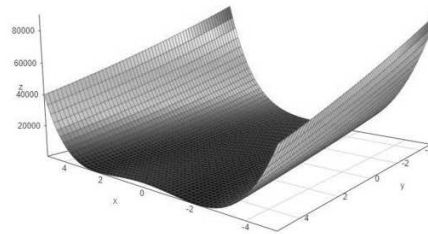


Fig. 2. Rosenbrook function.

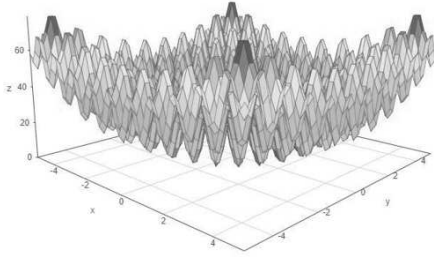


Fig. 3. Rastrigin function.

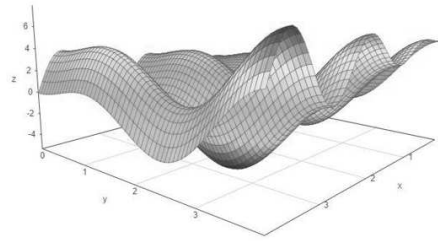


Fig. 4. Haupt function.

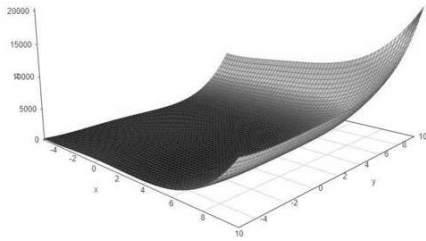


Fig. 5. Himmelblau function.

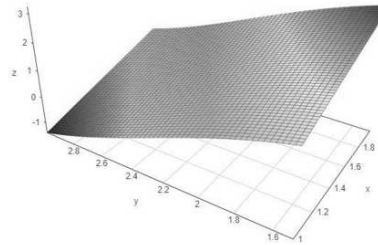


Fig. 6. Linear function.

Table 2  
Mean and standard deviation of variance parameter  $\beta^2$  and Euclidean fraction degree  $\delta$ .

Test function	Mean of $\beta^2$	Std. dev. of $\beta^2$	Mean of $\delta$	Std. dev. of $\delta$
Branin	518.398	165.630	0.800	0.012
Linear	0.406	0.049	0.769	0.025
Rosenbrook	$5.249 \cdot 10^7$	$1.859 \cdot 10^7$	0.697	0.026
Haupt	4.254	1.113	0.240	0.074
Rastrigin	87.856	26.691	0.427	0.098
Himmelblau	$4.003 \cdot 10^5$	$1.357 \cdot 10^5$	0.785	0.015

Table 3  
True error results.

Test function	$TE(y(x))$	$TE(y_{Shepard}(x))$
Branin	<b>21.919</b>	59.479
Linear	<b>0.123</b>	0.877
Rosenbrook	<b>8850</b>	13310
Haupt	<b>2.026</b>	2.47
Rastrigin	<b>12.432</b>	13.425
Himmelblau	<b>1061</b>	2149

Mean and standard deviation of variance parameter  $\beta^2$  and Euclidean fraction degree  $\delta$  are given in Table 2, histograms are presented in Figs. 7–18 ( $K = 20$ ,  $N = 200$ ).

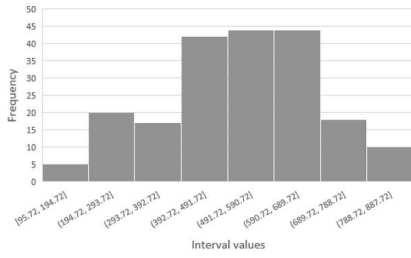


Fig. 7. Branin. Variance parameter.

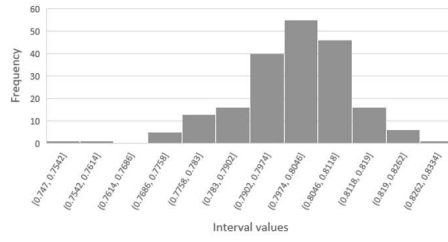


Fig. 8. Branin. Optimal Euclidean fraction degree.

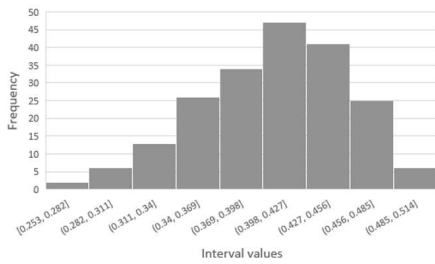


Fig. 9. Linear. Variance parameter.

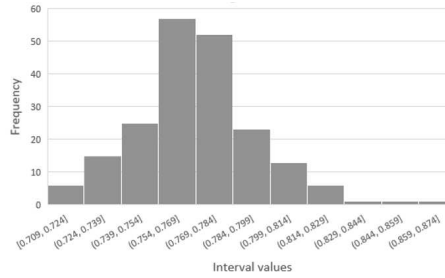


Fig. 10. Linear. Optimal Euclidean fraction degree.

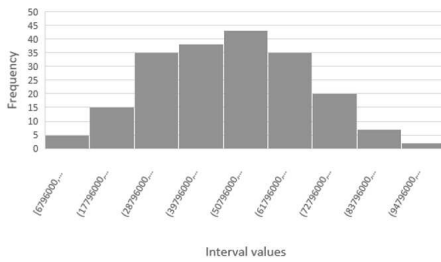


Fig. 11. Rosenbrock. Variance parameter.

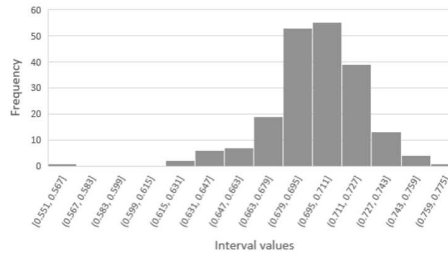


Fig. 12. Rosenbrock. Optimal Euclidean fraction degree.

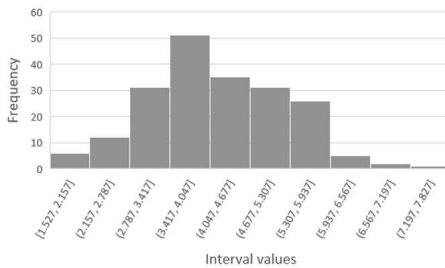


Fig. 13. Haupt. Variance parameter.

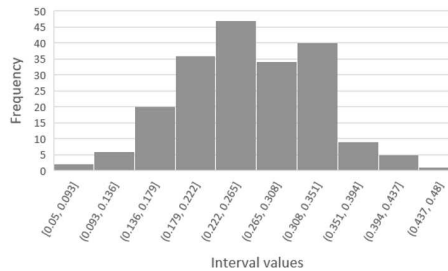


Fig. 14. Haupt. Optimal Euclidean fraction degree.

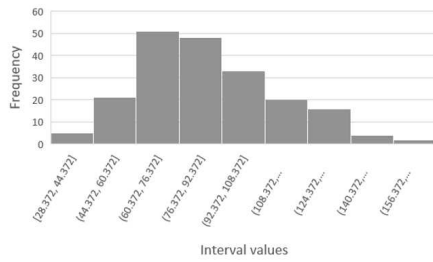


Fig. 15. Rastrigin. Variance parameter.

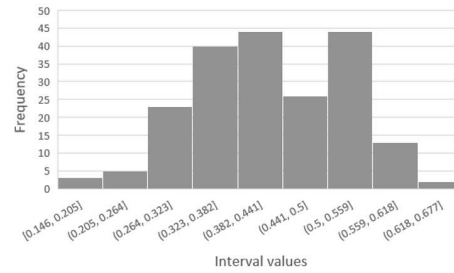


Fig. 16. Rastrigin. Optimal Euclidean fraction degree.

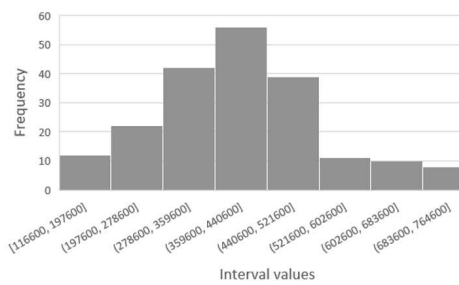


Fig. 17. Himmenblau. Variance parameter.

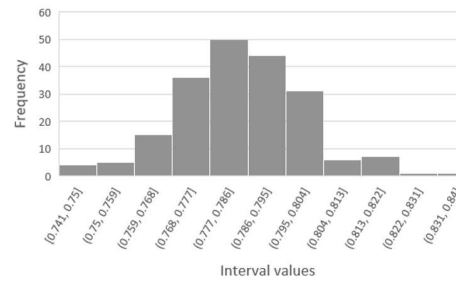


Fig. 18. Himmenblau. Optimal Euclidean fraction degree.

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

$$TE(y(x)) = \sum \sum \sqrt{(y(x) - z(x, y))^2}, \quad (34)$$

$$TE(y_{Shepard}(x)) = \sum \sum \sqrt{(y_{Shepard}(x) - z(x, y))^2}. \quad (35)$$

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  $K = 20$ ,  $N = 200$ .

Therefore, the kriging predictor, being a posterior expected value of 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.

## 6. Optimization of the Filler Effectiveness of Surface Wastewater Treatment Using Mathematical Modelling

Inadequate treatment of surface wastewater may considerably impair the quality of water. With increasing urbanization, intensification of car traffic and increasing area of impervious surfaces, the pollution of surface water and a negative impact on the aquatic en-



vironment are rising as well. One of surface wastewater treatment technologies, capable of reducing suspended solids, heavy metals and other pollutants, is surface wastewater treatment filters (Meškauskaitė and Marčiulaitienė, 2016; Meškauskaitė, 2017). 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 a vector of filler ratios. Denote different filter characteristics:  $Y_i^j, 1 \leq i \leq m, 1 \leq j \leq K$ , where  $K = 4$  is the number of experiments,  $m$  – is the number of filter characteristics, that describe capability to treat different wastes, depending on filler ratios of the filter.

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). Then the experiment matrix of filler proportions was chosen as follows:

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

After the experiment, the filtration characteristics have been presented by the measurement matrix  $Y = (Y^1, Y^2)^T$  (see Table 4).

Filters data extrapolation, based on experiments by the kriging method with distance matrices (20), (21), has been performed, where FEDM

$$A = \left[ \left( (x_{i,1} - x_{j,1})^2 + (x_{i,2} - x_{j,2})^2 + (x_{i,3} - x_{j,3})^2 + (x_{i,4} - x_{j,4})^2 \right)^{1/2} \right]_{i,j=1}^K$$

Table 4  
Filtration characteristics measurements.

Filters	Zn	Cu
	$y^1$	$y^2$
1	94.7	58.5
2	57.2	15.2
3	77.1	20.5
4	81.1	28.8

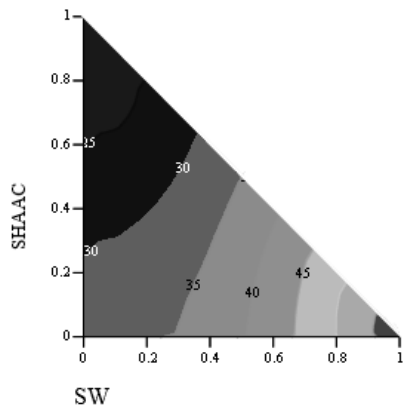


Fig. 19. Cleaning capacity in %. Copper (Cu). Quartz sand = 0.

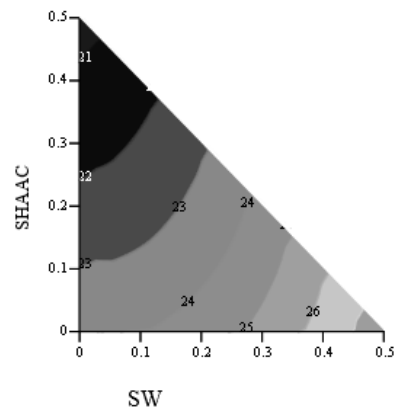


Fig. 20. Cleaning capacity in %. Copper (Cu). Quartz sand = 0.5.

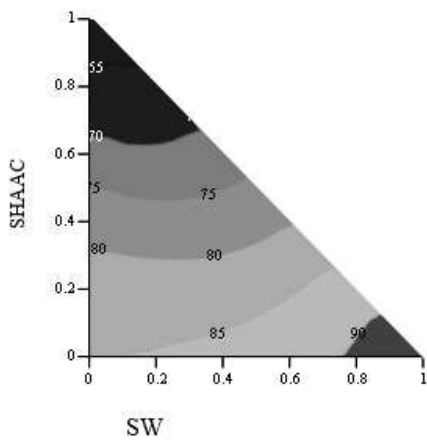


Fig. 21. Cleaning capacity in %. Zinc (Zn). Quartz sand = 0.

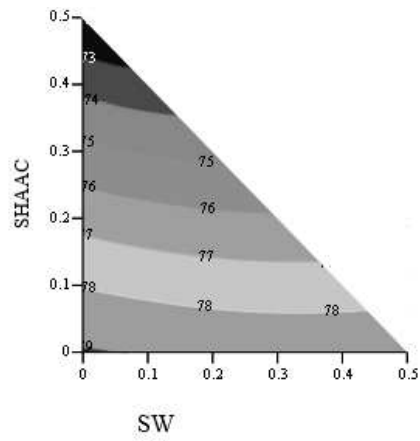


Fig. 22. Cleaning capacity in %. Zinc (Zn). Quartz sand = 0.5.

and the vector of distances from the selected extrapolation point  $x$  to the experimental points for  $i = 1, 2, 3, 4$  is

$$\tau_i(x) = ((X_{i,1} - x_1)^2 + (X_{i,2} - x_2)^2 + (X_{i,3} - x_3)^2 + (X_{i,4} - x_4)^2)^{1/2}.$$

Visualization of the cleaning efficiency (in %) for the fillers proportion (BC=1-QS-SHAAC-SW) using the kriging approach proposed, is given in Figs. 19–22.

### 7. Conclusions

Thus, the approach to multidimensional data modelling has been created using geometrical properties of FEDM, studied in the paper. It has been shown that the factorization of kernel matrix of FEDM enables us to create the embedded set being a nonsingular simplex. Using the properties of FEDM the Gaussian random field (GRF) is constructed doing it without positive definite correlation functions usually applied for such a purpose. Created GRF can be considered as a multidimensional analogue of the Wiener process, for instance, line realizations of this GRF are namely Wiener processes. Next, the multidimensional data kriging method is constructed that distinguishes by properties of homogeneity and isotropy. The developed model allows us to represent the information, obtained from any number of measurements of the objective function by a computational code or physical experiment. The resulting model is rather simple and depends on a small set of parameters (mean, variance, and parameter  $\delta$  of FEDM), that are efficiently estimated by the ML method. The results of scattered data processing by the approach, considered for analytically computed surfaces and mathematical modelling of a wastewater filter design process, illustrate the applicability of GRF with FEDM as models for scattered multidimensional data kriging.

### Appendix

*Proof of Theorem 1.* It is proved in Pozniak and Sakalauskas (2017), Theorem 2, that, under the conditions of Theorem, FEDM is nonsingular, i.e.  $|A| \neq 0$ , and thus its inverse exists, and its minors are nonzero. Taking into account (8), (10), it can be easily derived that:

$$E^T \cdot q = 1. \tag{36}$$

By virtue of the latter formula:

$$S^T \cdot q = (I - s \cdot E^T) \cdot q = q - s = s' \cdot A^{-1} \cdot \left( a + E \cdot \frac{1 - E^T \cdot A^{-1} \cdot a}{E^T \cdot A^{-1} \cdot E} \right). \tag{37}$$

Then, using (3), (4), (7)–(10), (36), (37), after simple manipulations, we construct:

$$\begin{aligned} \frac{F \cdot q \cdot q^T}{D} &= \frac{(s' \cdot E \cdot a^T \cdot S^T + s' \cdot S \cdot a \cdot E^T - S \cdot A \cdot S^T) \cdot q \cdot q^T}{D} \\ &= \frac{(s'^2 \cdot E \cdot a^T \cdot A^{-1} \cdot (a + E \cdot \frac{1 - E^T \cdot A^{-1} \cdot a}{E^T \cdot A^{-1} \cdot E}) - s' \cdot S \cdot E \cdot \frac{1 - E^T \cdot A^{-1} \cdot a}{E^T \cdot A^{-1} \cdot E}) \cdot q^T}{D} \\ &= E \cdot q^T = E \cdot s^T + s' \cdot E \cdot a^T \cdot A^{-1} + s' \cdot E \cdot E^T A^{-1} \cdot \frac{1 - E^T \cdot A^{-1} \cdot a}{E^T \cdot A^{-1} \cdot E}. \end{aligned} \tag{38}$$

On the other hand, the obvious equality

$$E^T \cdot \left( \frac{A^{-1} \cdot E \cdot E^T \cdot A^{-1}}{E^T \cdot A^{-1} \cdot E} - A^{-1} \right) = 0, \quad (39)$$

(4) and (37) yield the conclusion:

$$\begin{aligned} F \cdot \left( \frac{A^{-1} \cdot E \cdot E^T \cdot A^{-1}}{E^T \cdot A^{-1} \cdot E} - A^{-1} \right) &= (s' \cdot E \cdot a^T - (I - E \cdot s^T) \cdot A) \cdot \left( \frac{A^{-1} \cdot E \cdot E^T \cdot A^{-1}}{E^T \cdot A^{-1} \cdot E} - A^{-1} \right) \\ &= s' E \cdot \left( \frac{a^T A^{-1} \cdot E \cdot E^T \cdot A^{-1}}{E^T \cdot A^{-1} \cdot E} - a^T A^{-1} \right) - (I - E \cdot s^T) \cdot \left( \frac{E \cdot E^T \cdot A^{-1}}{E^T \cdot A^{-1} \cdot E} - I \right) \\ &= I - E \cdot s^T + \frac{s' \cdot (a^T A^{-1} \cdot E - 1) \cdot E \cdot E^T \cdot A^{-1}}{E^T \cdot A^{-1} \cdot E} - s' E \cdot a^T A^{-1}. \end{aligned} \quad (40)$$

Thus, combining (38) and (39) one can make sure that  $F \cdot M = I$ .

Using (3), (4), (10), the easily verified equality  $S^{-1} \cdot E = E/s'$  and Sylvester's determinant identity (Lay, 2005) the relation between determinants is derived:

$$\begin{aligned} |F| &= |s' \cdot E \cdot a^T \cdot S^T + s' \cdot S \cdot a \cdot E^T - S \cdot A \cdot S^T| \\ &= |S \cdot A \cdot S^T| \cdot |(E \cdot (a^T \cdot A^{-1} \cdot S^{-1}) \\ &\quad + (S \cdot a) \cdot (E^T \cdot (S^T)^{-1} \cdot A^{-1} \cdot S^{-1})) \cdot s' - I| \\ &= s'^2 \cdot |-A| \cdot |(E \cdot (a^T \cdot A^{-1} \cdot S) \\ &\quad + (S \cdot a) \cdot (E^T \cdot (S^T)^{-1} \cdot A^{-1} \cdot S^{-1})) \cdot s' - I| \\ &= (s')^2 \cdot |-A| \cdot \begin{vmatrix} a^T \cdot A^{-1} \cdot E - 1 & -a^T \cdot A^{-1} \cdot a \cdot s' \\ -\frac{E^T \cdot A^{-1} \cdot E}{s'} & a^T \cdot A^{-1} \cdot E - 1 \end{vmatrix} \\ &= -(s')^2 \cdot |-A| \cdot ((a^T \cdot A^{-1} \cdot a) \cdot (E^T \cdot A^{-1} \cdot E) - (1 - a^T \cdot A^{-1} \cdot E)^2). \end{aligned}$$

Hence, the theorem is proved.  $\square$

*Proof of Corollary 5.* Indeed, due to (7) and (36) one can easily notice that

$$F^{-1} \cdot E = \frac{q}{D},$$

and consequently:

$$E^T \cdot F^{-1} \cdot E = \frac{1}{D}.$$

*Proof of Theorem 3.* According to the block matrix inverse theorem (Lay, 2005),  $F$  is the upper left submatrix of the matrix

$$\bar{F}^{-1} - \frac{(\bar{F}_{K+1}^{-1})^\top \cdot \bar{F}_{K+1}^{-1}}{\bar{F}_{K+1,K+1}^{-1}},$$

(see (17)), bordered by zeros, namely:

$$\bar{F}^{-1} - \frac{(\bar{F}_{K+1}^{-1})^\top \cdot \bar{F}_{K+1}^{-1}}{\bar{F}_{K+1,K+1}^{-1}} = \begin{pmatrix} F & 0 \\ 0^\top & 0 \end{pmatrix}, \tag{41}$$

where the  $K + 1$ -th column of the inverse matrix  $\bar{F}^{-1}$  is denoted by  $\bar{F}_{K+1}^{-1}$ , and its lower-right element by  $\bar{F}_{K+1,K+1}^{-1}$ . Note that the  $K + 1$ -th column of matrix  $\bar{F}$  is decomposed  $\bar{F}^{K+1} = (f, v)$  as well. Now, rewrite (18) in the following manner by means of the properties of matrix operations with the elements of decompositions done:

$$\begin{aligned} & Y^\top \cdot F^{-1} \cdot (f + E \cdot \frac{(1 - E^\top \cdot F^{-1} \cdot f)}{E^\top \cdot F^{-1} \cdot E}) \\ &= \bar{Y}^\top \cdot \left( (\bar{F}^{-1} - \frac{(\bar{F}_{K+1}^{-1})^\top \cdot \bar{F}_{K+1}^{-1}}{\bar{F}_{K+1,K+1}^{-1}}) \cdot \bar{F}^{K+1} \right. \\ &\quad \left. + \bar{E} \cdot \frac{(1 - \bar{E}^\top \cdot (\bar{F}^{-1} - \frac{(\bar{F}_{K+1}^{-1})^\top \cdot \bar{F}_{K+1}^{-1}}{\bar{F}_{K+1,K+1}^{-1}}) \cdot \bar{F}^{K+1})}{\bar{E}^\top \cdot (\bar{F}^{-1} - \frac{(\bar{F}_{K+1}^{-1})^\top \cdot \bar{F}_{K+1}^{-1}}{\bar{F}_{K+1,K+1}^{-1}}) \cdot \hat{E}} \right) \\ &= \bar{Y}^\top \cdot \left( \bar{F}^{-1} \cdot \bar{F}^{K+1} + \frac{\bar{F}^{-1} \cdot \bar{E} \cdot \bar{E}^\top \cdot \bar{F}_{K+1}^{-1} - \bar{F}_{K+1}^{-1}}{\bar{F}_{K+1,K+1}^{-1} - \frac{(\bar{E}^\top \cdot \bar{F}_{K+1}^{-1})^2}{\bar{E}^\top \cdot \bar{F}^{-1} \cdot \bar{E}}} \right), \end{aligned} \tag{42}$$

where  $\bar{E}$  is a vector of  $K + 1$  units.

Next, let us consider  $\bar{F}$  being the upper-left submatrix of some kernel matrix of extended FEDM

$$\bar{A} = \begin{pmatrix} A & a \\ a^\top & 0 \end{pmatrix}.$$

Then, by virtue of Corollary 5 applied to extended matrices  $\bar{F}$  and  $\bar{A}$ , the right side of (42) is equal to

$$Y^\top \cdot \left( \bar{A}^{-1} \cdot \bar{A}^{K+1} + \frac{\bar{A}^{-1} \cdot \bar{E} \cdot \bar{E}^\top \cdot \bar{A}_{K+1}^{-1} - \bar{A}_{K+1}^{-1}}{\bar{A}_{K+1,K+1}^{-1} - \frac{(\bar{E}^\top \cdot \bar{A}_{K+1}^{-1})^2}{\bar{E}^\top \cdot \bar{A}^{-1} \cdot \bar{E}}} \right). \tag{43}$$

Repeating the considerations of transformation of (18) to (43), applied now back to (43), one can derive (21). Similarly, one can rearrange (19), then apply Corollary 5 and derive (22):

$$\begin{aligned} & \left( v - f^T \cdot F^{-1} \cdot f + \frac{(1 - E^T \cdot F^{-1} \cdot f)^2}{E^T \cdot F^{-1} \cdot E} \right) \\ &= \frac{1}{\bar{F}_{K+1, K+1}^{-1} - \frac{(\bar{E}^T \cdot \bar{F}_{K+1}^{-1})^2}{\bar{E}^T \cdot \bar{F}^{-1} \cdot \bar{E}}} = \frac{1}{\frac{(\bar{E}^T \cdot \bar{A}_{K+1}^{-1})^2}{\bar{E}^T \cdot \bar{A}^{-1} \cdot \bar{E}} - \bar{A}_{K+1, K+1}^{-1}} \\ &= a^T \cdot A^{-1} \cdot a - \frac{(1 - E^T \cdot A^{-1} \cdot a)^2}{E^T \cdot A^{-1} \cdot E}. \end{aligned}$$

Proving (23) after cumbersome but elementary manipulations we derive:

$$\begin{aligned} & F^{-1} - \frac{F^{-1} \cdot E \cdot E^T \cdot F^{-1}}{E^T \cdot F^{-1} \cdot E} \\ &= \bar{F}^{-1} - \frac{(\bar{F}_{K+1}^{-1})^T \cdot \bar{F}_{K+1}^{-1}}{\bar{F}_{K+1, K+1}^{-1}} \\ & \quad - \frac{(\bar{F}^{-1} - \frac{(\bar{F}_{K+1}^{-1})^T \cdot \bar{F}_{K+1}^{-1}}{\bar{F}_{K+1, K+1}^{-1}}) \cdot \bar{E} \cdot \bar{E}^T \cdot (\bar{F}^{-1} - \frac{(\bar{F}_{K+1}^{-1})^T \cdot \bar{F}_{K+1}^{-1}}{\bar{F}_{K+1, K+1}^{-1}})}{\bar{E}^T \cdot (F^{-1} - \frac{F^{-1} \cdot E \cdot E^T \cdot F^{-1}}{E^T \cdot F^{-1} \cdot E}) \cdot \bar{E}} \\ &= \bar{F}^{-1} - \frac{\bar{F}^{-1} \cdot \bar{E} \cdot \hat{E}^T \cdot \bar{F}^{-1}}{\bar{E}^T \cdot \bar{F}^{-1} \cdot \bar{E}} \\ & \quad - \frac{(\bar{F}_{K+1}^{-1} - \frac{\bar{E}^T \cdot \bar{F}_{K+1}^{-1}}{\bar{E}^T \cdot \bar{F}^{-1} \cdot \bar{E}} \cdot \bar{F}^{-1} \cdot \bar{E}) \cdot ((\bar{F}_{K+1}^{-1})^T - \frac{\bar{E}^T \cdot \bar{F}_{K+1}^{-1}}{\bar{E}^T \cdot \bar{F}^{-1} \cdot \bar{E}} \cdot \bar{E}^T \cdot \bar{F}^{-1})}{F^{-1}_{K, K} - \frac{(E^T \cdot \bar{F}_{K+1}^{-1})^2}{\bar{E}^T \cdot \bar{F}^{-1} \cdot \bar{E}}}. \end{aligned} \quad (44)$$

However, according to Corollary 5, matrix (44) is equal to

$$\begin{aligned} & -\bar{A}^{-1} + \frac{\bar{A}^{-1} \cdot \bar{E} \cdot \bar{E}^T \cdot \bar{A}^{-1}}{\bar{E}^T \cdot \bar{A}^{-1} \cdot \bar{E}} \\ & + \frac{(\bar{A}_{K+1}^{-1} - \frac{\bar{E}^T \cdot \bar{A}_{K+1}^{-1}}{\bar{E}^T \cdot \bar{A}^{-1} \cdot \bar{E}} \cdot \bar{A}^{-1} \cdot \bar{E}) \cdot ((\bar{A}_{K+1}^{-1})^T - \frac{\bar{E}^T \cdot \bar{A}_{K+1}^{-1}}{\bar{E}^T \cdot \bar{A}^{-1} \cdot \bar{E}} \cdot \bar{E}^T \cdot \bar{A}^{-1})}{\bar{A}_{K, K}^{-1} - \frac{(E^T \cdot \bar{A}_{K+1}^{-1})^2}{\bar{E}^T \cdot \bar{A}^{-1} \cdot \bar{E}}}. \end{aligned}$$

By analogous considerations used to get (44), one can show that the equality, similar to that in Corollary 5, takes the place of submatrices of extended kernel matrices and FEDM as well:

$$F^{-1} - \frac{F^{-1} \cdot E \cdot E^T \cdot F^{-1}}{E^T \cdot F^{-1} \cdot E} = -A^{-1} + \frac{A^{-1} \cdot E \cdot E^T \cdot A^{-1}}{E^T \cdot A^{-1} \cdot E}.$$

Thus, (23) is a direct consequence of the latter relation.  $\square$

## References

- Abrahamsen, P. (1997). *A Review of Gaussian Random Fields and Correlation Functions*. Norsk Regnesentral/Norwegian Computing Center, Oslo, Norway, 64 pp.
- Bhosekar, A., Ierapetritou, M. (2018). Advances in surrogate based modeling, feasibility analysis, and optimization: a review. *Computers & Chemical Engineering*. 108, 250–267.
- Carpio, R., Giordano, R.C., Secchi, A. (2017). Enhanced surrogate assisted global optimization algorithm based on maximizing probability of improvement. *Automatika i Telemekhanika*. 40, 2065–2070.
- Casella, G., Berger, R. (2002). *Statistical Inferences*. Wadsworth.
- Forrester, A.I.J., Keane, A.J. (2009). Recent advances in surrogate-based optimization. *Computers & Chemical Engineering*. 108, 250–267.
- Ghiasi, R., Ghasemi, M.R., Noori, M. (2018). Comparative studies of metamodeling and AI-Based techniques in damage detection of structures. *Advances in Engineering Software*. 125, 101–112.
- Gower, J.C. (1984). Multivariate analysis: ordination, multidimensional scaling and allied topics. In Lloyd, E.H. (Ed.), *Handbook of applicable mathematics, Vol. VI: Statistics*. J. Wiley and Sons, Chichester, pp. 727–781.
- Guttorp, P., Gneiting, T. (2005). *On the Whittle-Matérn correlation family*. NRCSE, Technical Report Series, NRCSE-TRS, No. 80, pp. 1–8.
- Jones, D.R. (2001). A taxonomy of global optimization methods based on response surfaces. *Journal of Global Optimization*, 21, 345–383.
- Koo, C., Hong, T., Jeong, K., Kim, J. (2018). Development of the monthly average daily solar radiation map using A-CBR, FEM, and kriging method. *Technological and Economic Development of Economy*. 24, 489–512.
- Khoshnevisan, D. (2002). *Multiparameter Processes – An Introduction to Random Fields*. Springer-Verlag.
- Krige, D.G. (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*. 52, 119–139.
- 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*. 86, 97–106.
- Kwon, H., Choi, S. (2015). A trended Kriging model with R2 indicator and application to design optimization. *Aerospace Science and Technology*. 43, 111–125.
- Lay, D.C. (2005). *Linear Algebra and Its Applications*. Addison Wesley.
- Meškauskaitė, L., Marčiulaitienė, E. (2016). Pollution of surface wastewater from petrol stations with suspended solids and petroleum products (in Lithuanian). In: *Proceedings of the 19th Conference of Lithuanian Junior Researchers*, Technika, Vilnius, Lithuania. pp. 123–129. ISSN 2029-5456.
- Meškauskaitė, L. (2017). *Stormwater Filters Experimental Studies with Construction Waste and Biocarbon*. Master Thesis. Vilnius Gediminas Technical University, Vilnius, Lithuania.
- Mockus, J. (1989). Bayesian approach to global optimization. In: *Mathematics and its Applications*. Springer Netherlands, Kluwer Academic Publishers.
- Pham, T.D., Beck, D., Brandl, M., Zhou, X. (2008). Classification of proteomic signals by block kriging error matching. *International Conference on Image and Signal Processing ICISP 2008: Image and Signal Processing*, pp. 281–288.
- Pozniak, N., Sakalauskas, L. (2017). Fractional Euclidean distance matrices extrapolator for scattered data. *Journal of Young Scientists*. 47(2), 56–61.
- Qian, P.Z.G., Wu, H., Wu, C.F.J. (2008). Gaussian process models for computer experiments with qualitative and quantitative factors. *Technometrics*. 50(3), 383–396.
- Sakalauskas, L. (2013). Locally homogeneous and isotropic Gaussian fields in kriging. *Informatika*. 24(2), 253–274.
- Schoenberg, I.J. (1935). Remarks to Maurice Fréchet's article: Sur la définition axiomatique d'une classe d'espaces vectoriels distances applicables vectoriellement sur l'espace de Hilbert. *Annals of Mathematics*. 36, 724–732.
- Shepard, D. (1968). A two dimensional interpolation function for irregularly spaced data. In: *Proceedings ACM 23rd Nat'l Conference*. pp. 517–524.
- 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*. pp. 106–113.
- 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*. 323, 120–131.
- Žilinskas, A. (1985). Axiomatic characterization of a global optimization algorithm and investigation of its search strategy. *Operations Research Letters*. 4, 35–39.

**N. Pozniak** is an external PhD student at Vilnius University Institute of Data Science and Digital Technologies.

**L. Sakalauskas**, Prof. Dr. Habil., is a principal researcher at Vilnius University Institute of Data Science and Digital Technologies. He is an associate professor at Vilnius Gediminas Technical University.

**L. Saltyte** is an associate professor at Klaipėda University. Her research interests belong to statistical data analysis and artificial intelligence.