

Hossam Talaat Elshambaky*

Enhancing the predictability of least-squares collocation through the integration with least-squares-support vector machine

<https://doi.org/10.1515/jag-2018-0017>

Received May 15, 2018; accepted August 8, 2018

Abstract: Least-squares collocation (LSC) is a crucial mathematical tool for solving many geodetic problems. It has the capability to adjust, filter, and predict unknown quantities that affect many geodetic applications. Hence, this study aims to enhance the predictability property of LSC through applying soft computing techniques in the stage of describing the covariance function. Soft computing techniques include the support vector machine (SVM), least-squares-support vector machine (LS-SVM), and artificial neural network (ANN). A real geodetic case study is used to predict a national geoid from the EGM2008 global geoid model in Egypt. A comparison study between parametric and soft computing techniques was performed to assess the LSC predictability accuracy. We found that the predictability accuracy increased when using soft computing techniques in the range of 10.2%–27.7 % and 8.2%–29.8 % based on the mean square error and the mean error terms, respectively, compared with the parametric models. The LS-SVM achieved the highest accuracy among the soft computing techniques. In addition, we found that the integration between the LS-SVM with LSC exhibits an accuracy of 20 % and 25 % higher than using LS-SVM independently as a predicting tool, based on the mean square error and mean error terms, respectively. Consequently, the LS-SVM integrated with LSC is recommended for enhanced predictability in geodetic applications.

Keywords: Least-Squares Collocation, Least-Squares Support Vector Machine, Multilayer Feed Forward Network, Radial Basis Neural Network

1 Introduction

Least-squares collocation (LSC) is the most general form of the adjustment process. It includes least-squares adjust-

ment, filtering, and prediction steps within a combined algorithm [16]. It is one of the several methods for solving the problem of spatial signal interpolation and prediction from discrete noisy measurements [42]. The primary concept of LSC is the principle of unbiased minimum error variance prediction when the underlying function is considered as a second-order stochastic process with a known, or estimated, covariance function [15]. The history of LSC is based on ideas developed by H. Moritz from 1970 to 1973 for optimal gravity field interpolation, prediction, filtering, and parameter estimation. The method was further developed by Krarup for solving partial differential equations, e. g., the Laplace equation, using heterogeneous data both at the boundary and in space. Subsequently, most geodetic measurements, parameters, and signals were managed with LSC, simultaneously. The theoretical background can be found in the literature [40, 47, 48]. LSC has widely been used to treat many geodetic and geophysics applications: geoid determination, prediction of vertical deflections, gravity anomaly prediction, computation of spherical harmonic coefficients, prediction and modeling of atmospheric fields, denoising and optimal separation of geodetic and geophysical signals, establishing error correction for network-based mobile positioning, spatial-temporal prediction of crustal deformation, prediction of land uplift phenomena, coordinate transformation, height datum transformation, and photogrammetry [6, 11, 12, 16, 20, 27, 42, 43, 46, 66, 67, 70, 72, 74]. LSC has many advantages; it is robust, flexible to combine with heterogeneous data, and easy to use [6, 20, 67]. In addition, it is considered as an optimal method of interpolation based on the empirical covariance function, especially with inhomogeneous and sparse data [43]. Moreover, it can adjust, interpolate, and predict parameters, signals, and noise in a unified model [47]. Nevertheless, LSC has its own disadvantages: (i) the distortion of predicted signals from its real values, due to the inconsistency between the signals and the observation uncertainties, (ii) it is only suitable for smooth and stable stochastic fields, (iii) the inability to reproduce the spatial variability adopted by the covariance function owing to the smoothing effect, and (iv) it is time consuming because of the number of equations that need to be solved [20, 42, 72]. Owing to previous

*Corresponding author: Hossam Talaat Elshambaky, Civil Dept., MISR Higher Institute for Engineering and Technology in Mansoura, Elmanasoura, Egypt, e-mail: hossam4000talat@gmail.com

deficiencies, many efforts have been expended to elaborate LSC. For instance: estimating covariance parameters using the restricted maximum likelihood approach; enhancing LSC estimation ability by eliminating its inherent smoothing effect; preserving most of its local prediction accuracy especially at high levels of signal-noise ratio; using an adaptive factor to relate the variance components of the signals and observation noise to secure the balance state through the prediction process; applying nonstationary covariance function to model the empirical values especially when the spatial relationship of the measured quantities changes significantly owing to the change in topography; and using an iterative procedure without a pre-assumed trend function to adapt LSC with pre-assumption parameters describing the variance in signals, measurements, and trend function [11, 12, 17, 38, 42, 72]. The correct choice of covariance function is critical for geoid height transformation, and covariance analysis is indispensable before geoid computation. Further, the success of LSC calculations depends primarily on the pre-chosen covariance function [67] because a correct covariance function leads to optimal predicted results. Although the covariance function is not well defined, it can yield acceptable results that are reasonable, and does deviate significantly from the reality [70]. However, it is known that the covariance function should describe the dependence structure sufficiently, and this is difficult as it probably would not have a simple analytical form. By fitting a suitable expression to the empirical covariance data, simplicity and optimality can be achieved [47]. We agree with the statement, “the primary problem is to presume a behavior of the covariance function that in turn is based on some ideal conditions producing nice properties of mathematical convergence [42].” Therefore, we attempt not to depend on the parametric expressions to describe the covariance function; instead, we attempt to use soft computing techniques such as support vector machine (SVM), least-squares-support vector machine (LS-SVM), and artificial neural network (ANN) to describe the previous dependence structure between measurements and signals to increase the accuracy of LSC’s predictability. It is difficult to separate between the SVM and LS-SVM, because the LS-SVM is derived from the SVM; both are soft computing techniques. The SVM was introduced by Vapnik within the area of statistical learning theory [61, 62, 68]. It is an extension of the support vector classifier [36]. The primary concept for both methods is to handle the nonlinear regression-fitting problem by transferring the observations from their primal space to the feature higher space. This is achieved by using a nonlinear mapping function to guarantee the linear regression relationship between

the inputs and outputs. In a real situation, to obtain a predetermined nonlinear mapping function is difficult. Instead, the effect of this function can be obtained implicitly through applying Mercer’s theorem, where the principle of positive definite kernel is applied. Subsequently, they can be trained with a learning algorithm from optimization theories to yield an optimum solution [63, 71]. More details about the difference between the two methods will be shown later. The final technique used was the ANN, and it can be used as a universal approximator function. This function (with a suitable number of neurons and hidden layers) can approximate any integrable function from one finite-dimensional space to another [34]. The ANN in this study is used as a fitting tool to fit the empirical values to establish the covariance function, as will be shown later. Two types of ANNs are implemented: the multilayer feed forward neural network (MFFNN), and the radial basis neural network (RBNN). Soft computing techniques have similar advantages; they can mitigate the problems with few observations, nonlinearity, high dimensions, and local minimum [71]. In addition to the previous advantages, the LS-SVM has fewer parameters and a faster training process [64]. They are kernel-based estimation techniques and have shown to be powerful nonlinear regression methods [22]. Further, they do not require strict rules in choosing the regularization parameters or the kernel functions. The unsuitable choice of previous parameters can lead to the over-fitting problem; using high-dimensional kernels results in reduced training speed and is time consuming. It is difficult to understand the learned functions, or in other words, the estimated weights. They exhibit unexpected behaviors. Further, they depend on trial and error to yield the optimum structural design. Some techniques suffer from the local minima issue such as the ANN [9, 32, 33, 34, 61, 62, 68]. Recently, soft computing techniques have been used successfully in some geodetic applications such as datum transformation, GIS applications, building geoid surface models, and in the transformation between global and national geoid models [1, 8, 18, 19, 21, 29, 30, 31, 35, 41, 45, 49, 52, 59, 60, 64, 65, 69, 71, 75, 76, 77]. In this study, we compare the soft computing techniques used to build the covariance functions with other well-known classical covariance functions such as the Hirvonen, Gaussian, and second and third Markov functions applied to the transformation process of geoid heights. The transformation process is used to transform the geoid heights generated by the global geopotential model to the national geometrical geoid, and the reverse is also true. We have chosen this application because it represents a crucial geodetic problem to developing countries such as Egypt. It is known that

the transformation process is a fast and cheap facility compared to establishing new leveling networks, particularly in new development areas. This study is divided into seven parts: (i) description of the steps to build the LSC model, (ii) explanation of the methods used to build the analytical covariance function (ACF), (iii) summary of the available data used in a real case study, (iv) narration of the research methodology, (v) analysis of the study results, (vi) comparison of the results obtained using the LS-SVM as a direct predicting tool against the integration with the LSC solution, and (vii) presentation of the conclusions and recommendations.

2 Steps required to build the LSC model

In this section, the steps used to build the LSC model will be described. The LSC algorithm can be subdivided into five steps. The first step is to prepare the raw data. In this step, it is noteworthy that the preconditions of this study are restricted to the two-dimensional zero-mean stationary-isotropic model [39]. To guarantee the stationary characteristic, the measured data at common points should be reduced to a trend surface; this can be performed by subtracting the trend surface value from the measured data, and subsequently use the reduced measurements [56]. Further, to gain the zero mean in the reduced measured data at common points, the mean value of the reduced measurements is subtracted from every individual reduced measured data to obtain the signals at common points [11]. In the second step, the spatial dependence structure between signals at common points should be calculated according to their spatial distribution. For gridded measurements, the covariances C_y , C_x , in both directions: north–south and east–west, can be calculated according to Eq. (1). The system of equations above represents a grid composed of squares of width a , and the region is assumed to be a rectangular of sides Ma and Na , M and N being integers. The value of signals P at the grid points form a rectangular matrix P_{ij} , where $i = 1, 2, \dots, M$, and $j = 1, 2, \dots, N$. Term C_r represents the covariance of the weighted mean that forms the desired estimates [47].

$$\left. \begin{aligned} C_y(ka) &= \frac{1}{(M-k)N} \sum_{j=1}^N \left(\sum_{i=k+1}^M P_{i-k,j} P_{i,j} \right) \\ C_x(ka) &= \frac{1}{(N-k)M} \sum_{j=1}^N \left(\sum_{i=k+1}^M P_{i,j-k} P_{i,j} \right) \\ C_r(ka) &= \frac{(M-k)NC_y(ka) + M(N-k)C_x(ka)}{(M-k)N + (N-k)M} \end{aligned} \right\} \quad (1)$$

For the scattered measurements, the covariances can be calculated as in Eq. (2),

$$C_r = \frac{1}{N_r} \sum P_i P_j \quad (2)$$

where C_r is the covariance between points i and j at a Euclidian distance less than or equal to the interval distance r , N_r is the total number of pair points, and $(P_i P_j)$ is the product of the signals at common point locations i and j . Both Eqs. (1) and (2) represent the empirical equation values at a specified interval distance r [48, 58]. At this stage, the empirical data can be represented by a histogram showing the probability density of the correlated reduced measured data within the predetermined interval distance r . Next, in the third step, the empirical values will be fitted to a chosen ACF, which can be determined from its parameters. These parameters are the correlation distance (D), and the variance of the noise (C_0). The parameters can be estimated by least-squares curve fitting. It is used to estimate the optimum curve to describe the relationship between the covariance, and the distance at every pair of common points. From the adjustment process, the optimum parameters can be calculated. Those fitted curves may be constant, sinusoidal, Gaussian, exponential, exponential cosine, as well as exponential sine and cosine, or any other form suitable to represent the empirical data [46]. In the fourth step, from the fitted curve or the covariance function, the variance–covariance matrix (VCM) can be generated for both signals at the common and computational points. The computational points are the points at which the predicted signals required are to be estimated. The variance–covariance matrix contains both the autocovariance and cross-covariance matrices. The autocovariance matrix has the diagonal elements of variances for the signals of the common points, whereas the cross-covariance matrix has the covariances between the signals of the common and computational points. The form of the VCM can be expressed as in Eq. (3).

$$\begin{bmatrix} C_{qq} & C_{qp} \\ C_{qp}^T & C_{LL} \end{bmatrix} \quad (3)$$

Here, C_{qq} is the auto-covariance matrix at computational points, C_{qp} is the cross-covariance matrix between both common and computational points, and C_{LL} is the autocovariance matrix of the measurements. Term C_{LL} can be calculated as in Eq. (4).

$$C_{LL} = C_{pp} + C_{nn} \quad (4)$$

Here, C_{pp} is the autocovariance matrix for the common points' signals; C_{nn} is the autocovariance matrix for the

measurement noise. It is noteworthy that the noise is assumed to be uncorrelated, and therefore would be considered as a single value added to the signal variances of the common points [46]. Additionally, if no noise is assumed in the measurements (or it is small enough to be neglected), subsequently, C_{LL} can be reduced to the autocovariance matrix C_{pp} for the signals of the common points only [10]. At the final step, the objective is to obtain the unknown parameters where the trend surface can be adjusted, filter the measurements, and estimate the predicted quantity at the computational points. Consequently, the system of equations that relates the measurements, unknown parameters, signals, and noise, should be rearranged in the general form of LSC as in Eq. (5),

$$L = Ax + Bs + n \quad (5)$$

where L is the measurement vector, and n is the randomly distributed uncorrelated residual vector related to the measurements. Because systematic errors do not exist, it can be called noise. Further, s is a vector of all signal quantities to be estimated, and because of systematic errors, it can be called a signal. It can be represented as $\begin{bmatrix} P \\ Q \end{bmatrix}$,

where P , Q represent the signals at both the common and computational points, respectively. B consists of the $[I \ 0]$ matrix that contains the identity and zero matrices, respectively. Ax is the trend or deterministic part, where A is the design matrix, and x is the vector of the unknown parameters [16, 74]. In the solution of the previous system, if only the prediction of the signals at the computational points is intended, Eq. (6) can be used [59]:

$$Q = C_{qp}(C_{pp})^{-1}P \quad (6)$$

The previous equation is a special case where we assumed the trend surface is known. Therefore, we will not adjust the parameters associated with the trend surface; otherwise, Eq. (6) can be modified to Eq. (7).

$$Q = C_{qp}(C_{pp})^{-1}(L - A\hat{x}) \quad (7)$$

The modified part of the equation $(L - A\hat{x})$ represents the residuals adjusted by the estimated parameters \hat{x} . In this case, the unknown parameters can be estimated according to Eq. (8) [44]:

$$\hat{x} = (A^T C_{pp}^{-1} A)^{-1} A^T C_{pp}^{-1} L \quad (8)$$

The accuracy for every element in the LSC adjustment process can be found in detail in the literature [10, 46]. From the previous steps, we noticed that the interval distance r affects the behavior of the covariance function that

represents the core of the LSC technique. This distance is critical in forming the characteristics of the behavior, but it has neither strict rules to follow nor appears as a parameter in the ACF expression. Nevertheless, the parameters that describe the behavior of the covariance function are based on this interval distance [16]. This distance is always chosen by experience or arbitrarily, as presented in many studies [6, 11, 12, 37, 48, 58, 70, 72, 74]. In addition, some prior criteria could be considered when selecting this distance. It should be neither too short to avoid spikes at the data points (the improvement can cover the region) nor too long to avoid discontinuity in the combined solution [26]. The second issue is the presumed spatial correlation between the measurement signals, as depicted in Eqs. (1) and (2). This relation is based only on the linear kernel, when many different kernels can be used to describe this spatial relation, as will be shown in the next sections herein. This study aims to investigate the impact of both previous factors on the accuracy of the LSC predictability.

3 Methods used to build the ACF

In this section, we will apply two distinct groups to build the ACFs; parametric and soft computing techniques groups. The first group (parametric) represents the most well-known classical ACFs. Those functions are used previously in the geodesy to determine the geoid heights. The models depend primarily on the Euclidian distance between the common and computational points. The name and formulas describing the models are illustrated in Table 1 [11, 39, 44, 47]. The variations in the models can be attributed to the local data structure and the effect of the global features (trend surface) [58]. Term C_0 represents the variance value at the zero correlation distance, D is the optimum correlation distance, and d is any distance at which the covariance value $C(d)$ is required.

The second group (soft computing techniques) contains two categories: the SVM and the LS-SVM. The second

Table 1: Different parametric models of analytical covariance functions.

Name	Formula
Cauchy (Hirvonen)	$C(d) = C_0 / (1 + (d/D)^2)$
Gaussian1	$C(d) = C_0 e^{-(d^2/D^2)}$
Gaussian 2	$C(d) = C_0 e^{-d^2/D^2}$
Second Markov	$C(d) = C_0 (1 + d/D) e^{-(d/D)}$
Third Markov	$C(d) = C_0 (1 + d/D + d^2/3D^2) e^{-(d/D)}$

category includes the ANN. This category is divided to the MFFNN and the RBNN. To understand the operating mechanism of these categories, it is first crucial to understand their mathematical derivations.

3.1 Mathematical structure of SVM and LS-SVM

The primary differences between the two methods are as follows: in the first method, the constraints are the inequality that contains the insensitive loss function, and the model solves a quadratic programming problem in dual space to obtain a sparse solution. In the second method, the constraints turn into equations with a least-squares quadratic loss function, and the model is solved as a linear system in dual space under a least-squares cost function [64, 73]. The derivation of the mathematical formulation for the LS-SVM is the same as that of the SVM with slight differences. Both include the same general steps. First, the formulation of the problem is performed in the primal weight space as a constrained optimization problem. Next, the Lagrange formulation is used. Subsequently, the optimality conditions can be constructed. Finally, the problem is solved in the dual space of Lagrange multipliers; the solution of the Lagrange multipliers is called the support vectors [63]. When a nonlinear decision boundary is required, the word “machine” is added to the term “support vector,” as the solution is generated automatically [36]. The observations or measurement data can be represented as $\{x_i, y_i\}_{i=1}^N$; the dataset contains the input vector x_i and observation value y_i , and the total number of input vectors x_i is equal to the total number of observations N . The input vector $x_i \in \mathbb{R}^n$ implies that every input vector has the number of components n ; in other words, it has a primal space equal to n . The individual output $y_i \in \mathbb{R}$, or it has a one-dimensional space. The relationship will be enforced to be linear regression, as in Eq. (9).

$$\begin{aligned} y_i - w^T \varphi(x_i) - b &\leq \epsilon + \xi_i \\ -y_i + w^T \varphi(x_i) + b &\leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* &\geq 0 \end{aligned} \quad (9)$$

Here, $\varphi(x_i)$ is the unknown nonlinear mapping function; the role of this function, as mentioned previously, is to transfer the input vector x_i from its primal space n to another higher space n_h or feature space to facilitate the linearity ($\varphi(x_i) : \mathbb{R}^n \rightarrow \mathbb{R}^{n_h}$). Further, w^T represents a transpose weight vector associated with the input vector (x_i), and b represents the bias of the linear model, where ($b \in \mathbb{R}$). Term ϵ is the Vapnik-insensitive loss function at

which accuracy is required for the approximation. (ξ_i, ξ_i^*) represent the slack variables. When using the LS-SVM, Eq. (9) can be reduced to Eq. (10), where the equality constraints hold.

$$y_i = w^T \varphi(x_i) + b + e_i \quad (10)$$

Here, e_i represents the residual error. The previous equation represents one constraint, and the system of constraints can be built based on all measurements. The issue now is the optimization problem based on the optimality condition as in Eq. (11) for the SVM, and as in Eq. (12) for the LS-SVM.

$$\min(w, b, \xi_i, \xi_i^*) = \frac{1}{2} w^T w + \frac{1}{2} c \sum_{i=1}^N (\xi_i + \xi_i^*) \quad (11)$$

Here, c is a constant > 0 and determines the amount up to which deviations from the desired insensitivity ϵ accuracy is tolerated.

$$\min(w, b, e) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^N e_i^2 \quad (12)$$

Term γ is a regularization parameter, which will be calculated later. To achieve optimality, the previous optimization equations should be solved in the dual space of Lagrange multipliers, as depicted in Eqs. (13) and (14), for both the SVM and LS-SVM, respectively.

$$\begin{aligned} \mathcal{L}(w, b, \alpha, \alpha^*, \eta, \eta^*) &= \frac{1}{2} w^T w + \frac{1}{2} c \sum_{i=1}^N (\xi_i + \xi_i^*) \\ &\quad - \sum_{i=1}^N \alpha_i (\epsilon + \xi_i - y_i + w^T \varphi(x_i) + b) \\ &\quad - \sum_{i=1}^N \alpha_i^* (\epsilon + \xi_i^* + y_i - w^T \varphi(x_i) - b) - \sum_{i=1}^N (\eta_i \xi_i + \eta_i^* \xi_i^*) \end{aligned} \quad (13)$$

$$\begin{aligned} \mathcal{L}(w, b, e, \alpha) &= \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^N e_i^2 - \sum_{i=1}^N \alpha_i (w^T \varphi(x_i) + b + e_i - y_i) \end{aligned} \quad (14)$$

Here, $(\alpha_i, \alpha_i^*, \eta_i, \eta_i^*) \in \mathbb{R}$ represent the Lagrange multipliers > 0 . To solve the previous equations, partial differentiations of the unknowns ($w, b, \alpha, \alpha^*, \eta, \eta^*, e$) are performed and equated with zero. For the SVM, the most popular method used to solve the system is the sequential minimal optimization. This algorithm depends on the decomposition method, whose details can be found in the literature [24, 25, 53]. In the LS-SVM, after equating the derivatives to zero, Eq. (15) is produced.

$$w = \sum_{i=1}^N \alpha_i \varphi(x_i); \quad \sum_{i=1}^N \alpha_i = 0; \quad \alpha_i = y_i e_i; \\ w^T \varphi(x_i) + b + e_i - y_i = 0 \quad (15)$$

By eliminating both (w, e) and solving for (b, α) , the following solution is obtained:

$$\begin{bmatrix} 0 & 1_v^T \\ 1_v & \Omega + I/Y \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \quad (16)$$

where $1_v = [1, \dots, 1]$ is a vector of dimension $(N, 1)$, the output y is a vector of measurements $[y_1, \dots, y_N]^T$. $\alpha = [\alpha_1, \dots, \alpha_N]^T$, and Ω is a matrix representing the effect of the nonlinear mapping function by applying Mercer's theorem for kernel, as it is calculated implicitly instead of obtaining the mapping function explicitly, as in Eq. (17).

$$\Omega_{ij} = K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j) \quad (17)$$

Here, $K(x_i, x_j)$ represents the positive definite kernel function; this function can assume many forms. Finally, after estimating the unknowns in both systems of the SVM and LS-SVM, any point in the primal space can be calculated from both models, as in Eq. (18) [14, 22, 36, 63, 73]. In our case, the covariance value $C(d)$ at distance d will be estimated.

$$C(d) = \sum_{i=1}^N \alpha_i K(x_i, x_j) + b \quad (18)$$

As mentioned previously, different kernel functions will be used. These kernel functions are used to describe the spatial correlation relationship. In this study, three types of kernel functions are used in the nonlinear mapping process: the linear, polynomial, and radial base functions. Their mathematical expressions are displayed in Table 2 [14, 63]. The cross-validation method is used to determine both the hyperparameter selection regularization (γ) and the variance Gaussian kernel (σ) [63].

3.2 Mathematical structure of ANN

The mathematical formulation of the MFFNN can be divided into three steps. The first step is to relate the ini-

tial output S_k^l with the input signal's vector x_i within the hidden layer, as depicted in Eq. (19). Here, w_i^l represents the connection weight vector from the previous layer of neurons, b_0 is the bias weight vector that corresponds to an additional independent input, n is the number of input signal sources to a definite destination neuron, l is the number of hidden layers, and k is the number of destination neuron ANNs. Further, the input vector x_i represents the distance values based on the predefined interval distance r .

$$S_k^l = \sum_{i=1}^n x_i w_i^l + b_0^l \quad (19)$$

In the second step, the initial output S_k^l will be changed to the output a_k^l through a pre-chosen transfer function $f(S_k^l)$ according to Eq. (20), where a hyperbolic tangent sigmoid transfer function is used.

$$a_k^l = f(S_k^l) = \frac{e^{S_k^l} - e^{-S_k^l}}{e^{S_k^l} + e^{-S_k^l}} \quad (20)$$

In the final step, the empirical values' vector $C(d)$, which contains the covariances as depicted in Eqs. (1) or (2), will be related to the transferred data a_k^l through Eq. (21)

$$C(d) = \sum_{i=1}^n a_k^l w_i^l + b_0^l \quad (21)$$

The objective of the ANN is to estimate the value of all unknowns (w_i^l, b_0^l). Hence, the ANN will minimize the error between the estimated empirical values and the input empirical values by applying the training process. The training process depends primarily on updating the initial values of the unknowns by propagating the effect of the error back to the unknowns [32]. The error minimization process uses the Levenberg–Marquardt algorithm, which adaptively varies the unknown parameters (weights and biases) between the gradient descent update and the Gauss–Newton update until the algorithm reaches a stable state. For the second technique (RBNN), only Eq. (19) is changed to Eq. (22), and the transfer function used is the Gaussian instead of the hyperbolic tangent sigmoid function. More details about the ANN design procedures can be found in the literature [7].

$$S_k^l = \sum_{i=1}^n \|w_i^l - x_i\| b_0^l \quad (22)$$

In the design stage of the ANN, the number of neurons was determined through the growing method [33]. The avail-

Table 2: Different models of kernel functions.

Kernel Function Type	Formula
Linear	$K(x_i, x_j) = x_i^T x_j$
Polynomial of degree (d)	$K(x_i, x_j) = (t + x_i^T x_j)^d, \quad t \geq 0$
Radial base function (RBF)	$K(x_i, x_j) = e^{-\frac{\ x_i - x_j\ ^2}{2\sigma^2}}$ where σ^2 the variance of Gaussian kernel

able data were divided into three groups: training, validation, and testing. From the previous groups, the number of iterations, hidden layers, and neurons can be determined according to the statistical indicators, as will be shown later in the research methodology section. In this study, both ANN techniques contain an input layer, two hidden layers, and an output layer. The MFFNN has a design of (1-1-1) neurons at an interval distance of 40 km. This design is the optimum one at which four unknowns are estimated. The RBNN has a design of (1-19-1) neurons at an interval distance of 60 km with 40 unknowns. More details about the design steps and mathematical structure of the ANN can be found in the literature [18].

Finally, it is noteworthy that the parametric group can be considered as a kernel technique because it depends primarily on Eq. (2) to represent the spatial correlation relationship. This equation can be considered as a linear kernel equation, because it is a kernel from a kernel [63]. According to Eq. (23), α can be replaced by $\frac{1}{N_r}$ from Eq. (2), and Eq. (23) in this case will be equivalent to Eq. (2).

$$K(x_i, x_j) = \alpha K(x_i^T x_j), \alpha > 0 \quad (23)$$

4 The available data

The study area ranged from latitude 22°N to 32°N, and from longitude 25°E to 37°E. The total available points are 278 GPS/leveling measurements. These points belong to different sources such as the Egyptian Surveying Authority (ESA), and the Survey Research Institute. Spirit leveling was used to obtain the orthometric heights. The observations were performed as closed loops that run between known high-precision benchmarks established by the ESA. The high-precision benchmarks are based on the national vertical datum of Egypt. The origin of the Egyptian vertical datum is based on the mean sea level at the Alexandria tide gauge of 1906. The accuracy of these measurements is unavailable because the ESA did not publish the information. The GPS measurements were performed relative to the ESA national geodetic reference framework of Egypt. The accuracy of the ellipsoidal heights was in the range of 3 to 6 mm. The primary comment related to the spatial distribution of those points is the uneven distribution through the total study area. The points are concentrated primarily at the north coast and Nile valley, where most activities occur, as shown in Fig. 1. More informa-

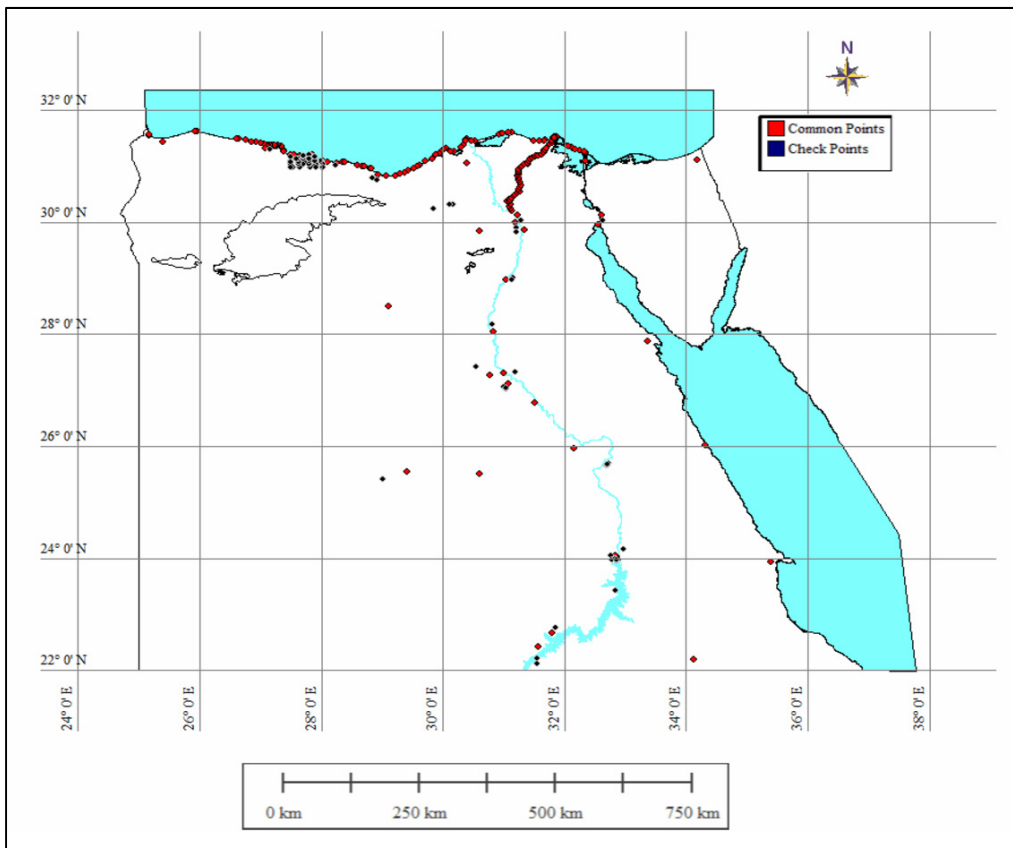


Figure 1: Distribution of common and check points [18].

tion about the available points can be found in the literature [18, 54]. The available raw data will be filtered and divided into two groups: common or (training) points, and the check or (computational) points, as will be shown later in the analysis process.

5 Research methodology

In this section, we will describe the steps that we used to enhance the predictability of LSC. The summarized steps are shown in Fig. 2, and every step is illustrated in the following section.

5.1 Outlier detection

This step contains the following procedures: First, the geoid undulations relative to every point were calculated according to Eq. (24).

$$N^L = h - H \quad (24)$$

where N^L is the measured national geoid undulation, h is the ellipsoidal height, and H represents the orthometric height. The resulted N^L will be compared with the geoid N^G extracted from the global geopotential model EGM2008 at the same location of common points, using the GUT 2.2 software based on the mean-tide condition [23]. The EGM2008 model is chosen as the most appropriate trend surface for Egypt. This choice is based on many studies that assessed different global geoid models with various sources of data. Most studies showed that the EGM2008 geoid model is the most suitable one [4, 5, 13, 18]. The residuals ΔN between the global and national geoid are calculated according to Eq. (25). The average and standard deviation (\bar{x}, σ) are calculated for the residuals' vector ΔN . Subsequently, routine 3σ tests were applied to remove any outliers [28]. Fifteen points were eliminated from the 278 available.

$$\Delta N = N^G - N^L \quad (25)$$

Table 3: Effect of outlier detection process on the available data.

	\bar{x}	σ
Before outlier removal (m)	0.313	0.431
After outlier removal (m)	0.282	0.358
Improvement (%)	9.904	16.937

The remaining 263 points were subdivided into 181 common points and 82 checkpoints. The improvement in ΔN owing to the outlier extraction is depicted in Table 3 [18].

5.2 Calculation of empirical covariance function

Owing to the no gridded data, Eq. (2) will be used instead of Eq. (1) to generate the empirical values through different interval distances r . In this part, the interval distance was chosen as 1, 10, 20, ..., until 1000 km. A representation of the empirical values as an example at different distances r is depicted in (Fig. 3). It is clear from the previous figure that the behavior of the empirical values changed according to the interval distance r . The number of spikes is inversely proportional to the interval distance r . Many details can be missed or cancelled with too long an interval distance, and the distance of zero covariance value is changed relative to the interval distance (r). All these previous characteristics were neither parameterized nor evaluated. No rules are available to guide us in choosing a suitable interval distance that affect the LSC predictability. For the next step, all empirical functions based on various interval distances will be used to generate different covariance functions.

5.3 Generation of ACF and covariance matrix

From the previous step, the covariance functions can be built using the models belonging to the two groups, as mentioned in Section 3. The pre-assumptions used to the build covariance functions are noteworthy. We assumed that the C_0 parameter is separated from the covariance matrix, and it is estimated from the empirical function at a correlation distance of zero. It has a value of 0.0697 m^2 . Moreover, we consider no noise in the observations, as the ESA has not published the records of accuracy. According to the previous reasons, the measurement's autocovariance matrix C_{LL} is equal to the common point signal's autocovariance matrix C_{pp} , as in Eq. (4). The only estimated parameter is the optimum correlation distance (D) [50, 51, 55, 74]. Finally, the VCM associated with every fitting model can be constructed using Eqs. (3) and (4).

5.4 Validation of fitting models

We attempt to draw the whole picture of the various effects resulted from applying different fitting models on the LSC

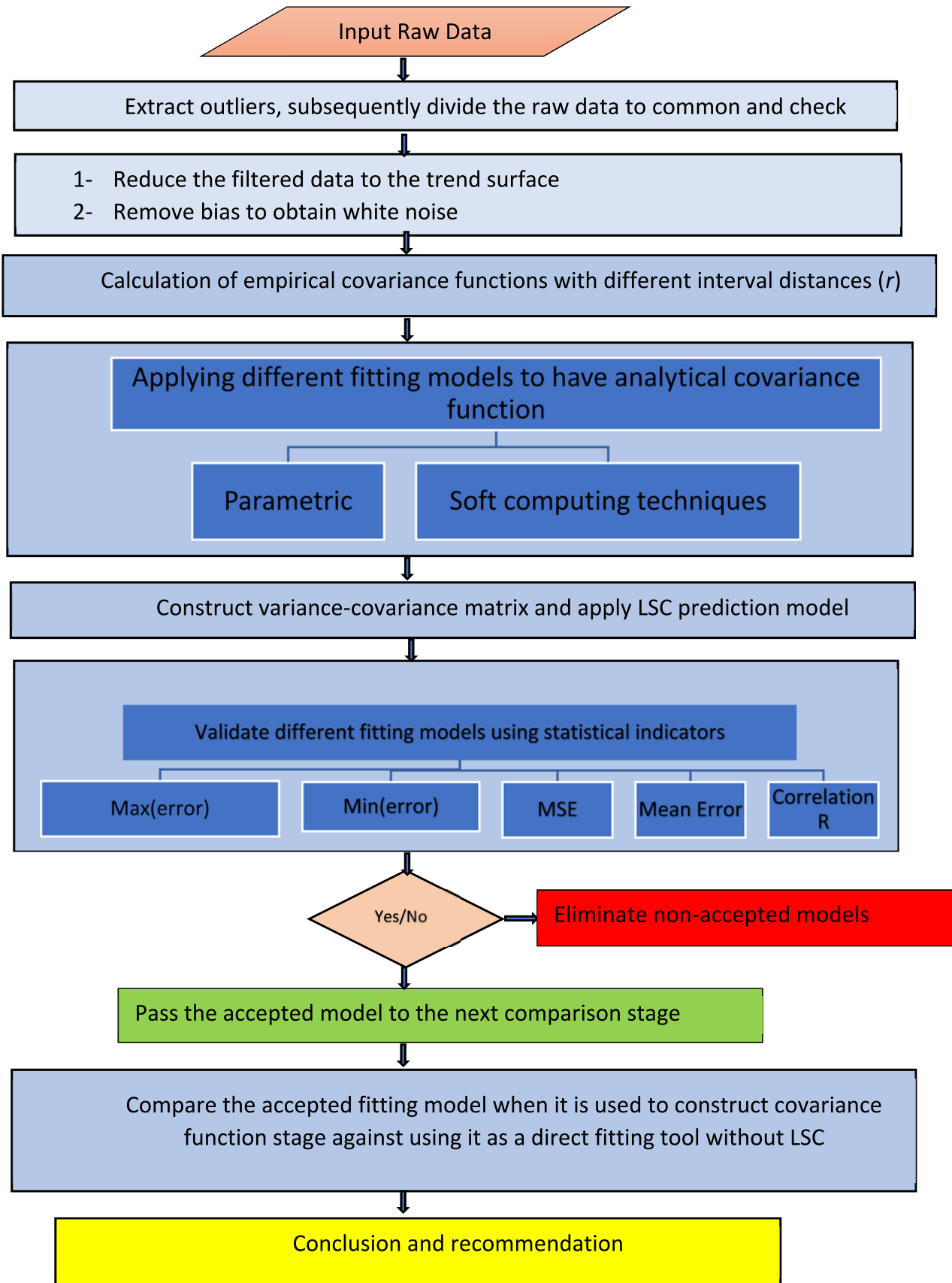


Figure 2: Research methodology flowchart.

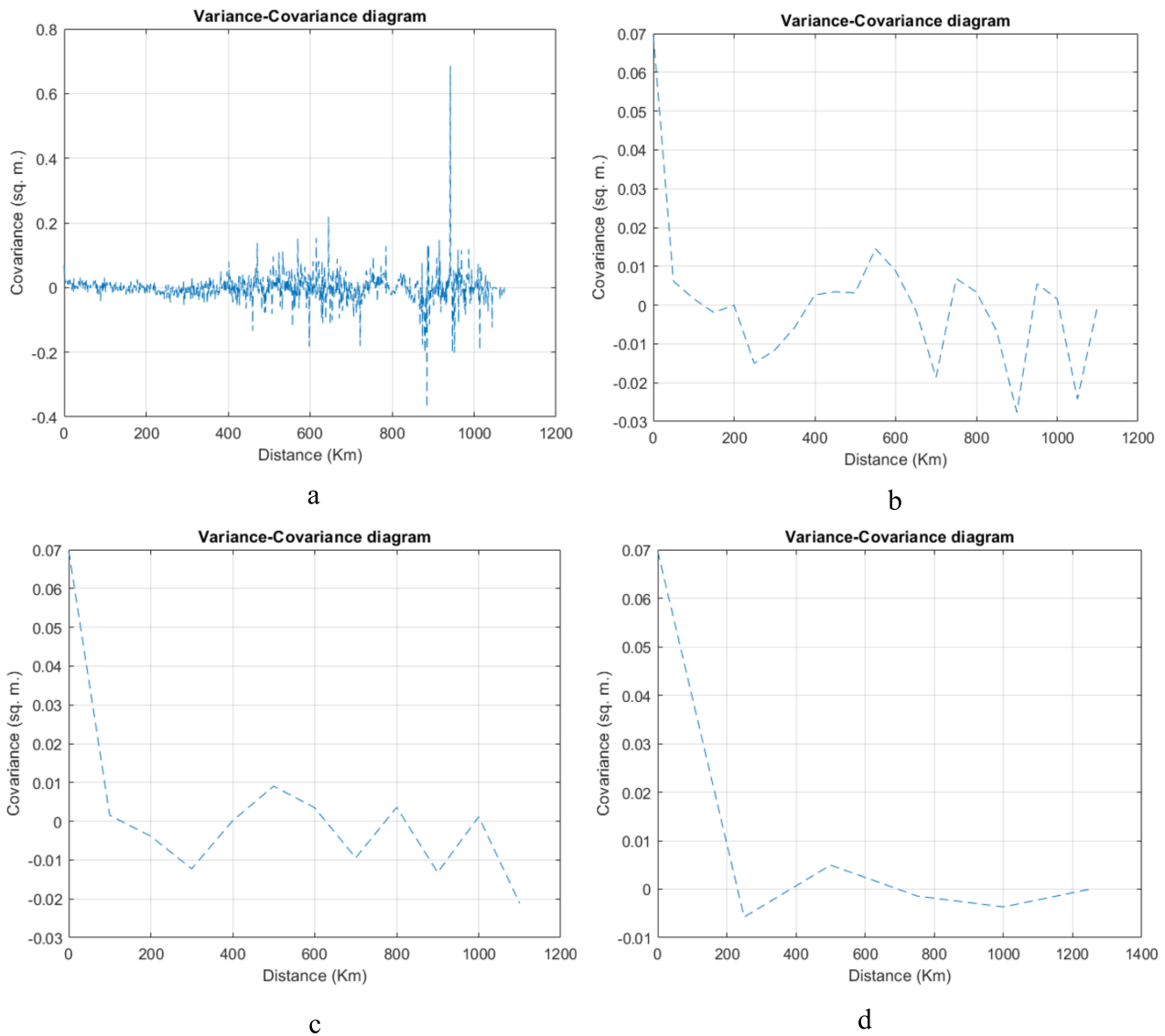


Figure 3: (a, b, c, and d) Empirical values at (1, 50, 100, and 250 km) interval distance r .

predictability. Here, the validation process will depend on a group of statistical indicators (key performance indicators (KPIs)). In this study, six statistical indicators will be adopted. Equation (6) will be used to predict the signals on both the common and computational points. The validation process will operate for the computational points only; in other words, for the external quality check. As depicted from Eq. (6), if the common points P are to be calculated, Eq. (6) will be changed to Eq. (26). Therefore, all fitting covariance functions will have no effect on the internal quality of LSC.

$$P = C_{PP}(C_{PP})^{-1}P \quad (26)$$

The estimated national geoid at the computational points can be calculated from Eq. (27), after restoring the

trend and bias values

$$\hat{N}^L = N^{G*} + s + Q \quad (27)$$

Here, \hat{N}^L is the estimated national geoid, N^{G*} is the global geoid calculated from the trend surface at the computational points, s is the bias value equaling 0.282 m as listed in Table 3, and Q is the estimated noise vector at the computational points from Eq. (6). The residuals error vector for the computational points can be calculated from Eq. (28),

$$e = N^{L*} - \hat{N}^L \quad (28)$$

where N^{L*} represents the measured national geoid vector calculated from Eq. (24) at the computational points. The KPIs adopted for the external quality mathematical formulation are listed in Table 4 [2, 3, 57]. Parameters ($\overline{N^{L*}}$, \hat{N}^L)

are the average values for the measured and estimated national geoid, respectively; $(N_i^{L^*}, \hat{N}_i^{L^*})$ are any elements in the measured and estimated geoid vector, respectively; e_i is any element in the error vector. It is clear that the KPIs represent the accuracy of the model used to build the ACF. A high KPI implies a low error generated by a high accuracy model, whereas a low KPI implies a high error generated by a low accuracy model.

Table 4: Statistical indicators.

Key Performance Indicators (KPI)	Mathematical Expression
Maximum Error	$\max(e)$
Minimum Error	$\min(e)$
Mean Error (\bar{e})	$\sum_i^n e_i / n$
Standard Deviation (σ_e)	$\sqrt{\frac{\sum_i^n (e_i - \bar{e})^2}{n-1}}$
Mean Square Error (MSE)	$e^T e / n$
Correlation (R)	$\frac{\sum_i^n (N_i^{L^*} - \bar{N}^{L^*})(\hat{N}_i^{L^*} - \bar{\hat{N}}^{L^*})}{\sqrt{\sum_i^n (N_i^{L^*} - \bar{N}^{L^*})^2} \times \sqrt{\sum_i^n (\hat{N}_i^{L^*} - \bar{\hat{N}}^{L^*})^2}}$

6 Result analysis

In this section, we present an analysis of the results from applying different mathematical models to build the ACF. Tables 5 and 6 display various values of the interval distance associated with the low and high values of the KPI. We call the interval distance at a high KPI a “good choice,” and at a low KPI a “bad choice.” It is clear that the good and bad choices varied according to the type of mathematical model used to build the ACF, and the choice of the optimum interval distance to yield a high KPI is not governed by any rules. The optimum interval distance can be estimated by testing many values, as we have performed. From Table 5, the worst behavior of the parametric models belongs to Gauss 2 at an interval distance of 70 km; subsequently, the model recovered its quality at the 100-km interval distance, as shown in Table 6. The remaining parametric models display almost no improvement in their quality when the interval distance changed from the bad to good choice; this signifies that the parametric models cannot adopt spatial variability. Meanwhile, soft computing techniques change their behavior based on the variation

Table 5: Bad choice of interval distance associated with low (KPI).

	Model	Interval Distance r (km)	Max. Error (m)	Min. Error (m)	MSE (m)	(\bar{e}) (m)	(σ_e) (m)	(R)
Parametric Models	Cauchy	10	1.408	-2.639	1.567	0.441	1.179	0.933
	Gauss 1	40	0.987	-0.417	0.331	0.484	0.312	0.996
	Gauss 2	70	1.594	0.000	19437.626	18.543	139.031	0.000
	Markov 2D	20	0.987	-0.417	0.331	0.484	0.311	0.996
	Markov 3D	80	0.963	-0.417	0.331	0.491	0.302	0.996
	MFNN	70	0.962	-0.292	0.311	0.482	0.281	0.997
Soft Computing Techniques	RBNN	50	452.196	-430.124	18763.662	7.375	137.624	0.417
	SVM	60	2406.869	-1516.694	213579.516	75.027	458.822	0.058
	LS-SVM	1	34.745	-14.643	24.866	0.734	4.963	0.767

Table 6: Good choice of interval distance associated with high (KPI).

	Model	Interval Distance r (km)	Max. Error (m)	Min. Error (m)	MSE (m)	(\bar{e}) (m)	(σ_e) (m)	(R)
Parametric Models	Cauchy	1	0.960	-0.417	0.324	0.484	0.302	0.996
	Gauss 1	1	0.977	-0.417	0.330	0.486	0.309	0.996
	Gauss 2	100	0.969	-0.417	0.328	0.486	0.305	0.996
	Markov 2D	250	0.968	-0.417	0.328	0.487	0.304	0.996
	Markov 3D	50	0.969	-0.417	0.329	0.487	0.305	0.996
	MFNN	40	0.916	-0.538	0.271	0.419	0.310	0.996
Soft Computing Techniques	RBNN	60	1.057	-0.423	0.291	0.444	0.308	0.996
	SVM	70	0.935	-0.387	0.279	0.434	0.303	0.996
	LS-SVM	70	0.800	-0.706	0.239	0.342	0.301	0.996

in the interval distance to improve their KPIs, as shown in the results of Table 6 compared with Table 5.

The good and bad choices can be close to each other, as in the RBNN and SVM models. Both can differ widely as in the Markov 2D and LS-SVM. From Table 5, it is clear that the incorrect choice of interval distance can deviate the prediction process significantly. From Table 6, it is clear that the parametric models have almost the same value of MSE, ranging from 0.324 to 0.330 m, whereas the soft computing techniques have another category of MSE ranging from 0.239 to 0.291 m. For the mean error (\bar{e}) and its standard deviation (σ_e), it is clear that the LS-SVM has the minimum value of mean error with high accuracy among all models. At a good choice of interval distance, all models have the same correlation factor of 0.996, and this improvement appeared clearly in the soft computing techniques as they can adopt the spatial variability of the data structure. From the absolute value of the maximum and minimum errors, it is clear that the LS-SVM has smaller values. Tables 7 and 8 show the improvement percentages in the MSE and mean error (\bar{e}) values for every soft computing technique against every parametric model; it is clear from these tables that the soft computing techniques supersede the parametric models in enhancing the LSC predictability accuracy at a good choice of interval distance r . The improvement percentage in the MSE value ranges from 10.2 % to 27.7 %, and the mean error (\bar{e}) value ranges from 8.2 to 29.8. In addition, the LS-SVM is the most accurate fitting model among the soft computing techniques, followed by the MFFNN. The average accuracy produced from applying the LS-SVM is improved in terms of the MSE and

mean error (27.2 % and 29.6 % higher than the parametric models, respectively). From the above analysis, it can be concluded that using soft computing techniques to represent the spatial correlation distribution improved the predictability of LSC, and the most appropriate model is the LS-SVM.

According to the previous analysis, the decision is to eliminate the parametric models. The next stage of comparison is designed to determine if it is better to use soft computing techniques as a direct predicting tool independently [49, 57, 64, 75], or to inject them in the inner structure of the LSC technique (as in the present study). This is addressed in the following section.

7 LS-SVM integrated with LSC versus LS-SVM as a direct predicting tool

From the previous stage, the LS-SVM is chosen as the most accurate model to represent the spatial correlation relationship. At this stage, the soft computing technique (LS-SVM) will be used as a direct curve fitting tool. It will be applied to the vector of the geoid residuals ΔN in Eq. (25). Subsequently, the resulted fitting model can be used again to predict the geoid at the computational points, as depicted in Eq. (29).

$$\hat{N}^L = N^{G*} + \Delta N^* \quad (29)$$

Parameters \hat{N}^L, N^{G*} were defined previously in Eq. (27), while ΔN^* represents the predicted vector of geoid residuals at computational points' locations estimated by the LS-SVM model. Again, to assess this step, Eq. (28) will be applied and the quality of the fitting process will be assessed as described previously. Subsequently, the results will be compared to the results obtained from the previous stage. It is noteworthy that all three kernel functions were used with the LS-SVM as a direct fitting tool, and the optimum KPI is achieved with the RBF, similar to as when using the LS-SVM associated with the LSC solution. Table 9 shows the comparison values based on the same external statistical indicators used previously. From these results, it is clear that the integration between the LS-SVM and LSC provides better and more accurate results. The improvement in the accuracy of predictability is up to 20 % and 25 % based on the MSE and mean error (\bar{e}), respectively. In addition, from the maximum and minimum errors values, we observed that the LS-SVM with LSC attempts to adjust its behavior to be centralized within the errors and be in a

Table 7: Improvement in MSE value based on soft computing techniques relative to parametric models.

Model	Cauchy	Gauss 1	Gauss 2	Markov 2D	Markov 3D
MFFNN	16.4	18.0	17.5	17.5	17.8
RBNN	10.2	11.9	11.4	11.4	11.6
SVM	13.8	15.4	14.9	14.9	15.2
LS-SVM	26.2	27.7	27.2	27.2	27.4

Table 8: Improvement in mean error (\bar{e}) value based on soft computing techniques relative to parametric models.

Model	Cauchy	Gauss 1	Gauss 2	Markov 2D	Markov 3D
MFFNN	13.4	13.7	13.8	13.9	13.9
RBNN	8.2	8.6	8.6	8.8	8.8
SVM	10.2	10.6	10.7	10.8	10.8
LS-SVM	29.3	29.6	29.6	29.8	29.7

Table 9: Comparison between LS-SVM with LSC and LS-SVM without LSC.

Model	Max. Error (m)	Min. Error (m)	MSE (m)	(ρ) (m)	(σ_e) (m)	(R)
LS-SVM with LSC	0.850	-0.740	0.239	0.342	0.301	0.996
LS-SVM without LSC	0.952	-0.446	0.296	0.455	0.301	0.996

normal distribution. Meanwhile, the LS-SVM without LSC has a small skew in its mean error value. Both have the same correlation value with the original data. This is due to the small amount of available data.

8 Conclusion

From this study, we conclude that LSC as a predicting tool depended primarily on the correct structure of the covariance function, and could in turn describe the spatial correlation relationship between measurements. Two important factors affected the method to describe this relationship; the interval distance and the mathematical models representing this relationship. A bad or good choice of the interval distance could severely affect the accuracy of LSC predictability. Despite the importance of this factor, the choice process is not governed by any regularization rules. The trial-and-error method is recommended for achieving the desired accuracy. The effect of the interval distance is illustrated in Tables 5 and 6 of this study. The next important factor is the type of mathematical model used to fit the ACF to the empirical data. We noticed that the parametric models could not adopt the spatial variability of the data when the interval distance was changed. Meanwhile, we found that soft computing techniques could modify their behavior to be more accurate than the parametric models. The enhancement in their accuracy ranges from 10.2% to 27.7% and 8.2% to 29.8% in terms of the MSE and mean error, respectively. In addition, we found that the LS-SVM was the most precise soft computing technique to represent the spatial correlation relationship. Finally, it is noteworthy that the results of the integration between the LS-SVM and LSC were more accurate than the results of using the LS-SVM independently as a predicting tool. The improvement in the prediction accuracy with the first solution was better than the second one by 20% and 25% in terms of the MSE and the mean error, respectively. Therefore, using the LS-SVM integrated with LSC is recommended to enhance the accuracy of LSC predictability.

References

- [1] Akyilmaz, O., Özlüdemir, M. T., Ayan, T., and Çelik, R. N., (2009), "Soft computing methods for geoidal height transformation", *Earth Planets and Space*, Vol. 61, No. 7, pp. 825–833.
- [2] Ali, M. H., and Abustan, I., (2014), "A new novel index for evaluating model performance", *Journal of Natural Resources and Development*, Vol. 04, pp. 1–9, DOI: 10.5027/jnrd.v4i0.01.
- [3] Al-Krargy, E. M., Doma, M. I., and Dawod, G. M., (2014), "Towards an Accurate Definition of the Local Geoid Model in Egypt using GPS/Leveling Data: A Case Study at Rosetta Zone", *International Journal of Innovative Science and Modern Engineering (IJISME)* ISSN: 2319-6386, Vol. 2, No. 11.
- [4] Al-Krargy, E., Hosny, M., and Dawod, G., (2015), "Investigating the Precision of Recent Global Geoid Models and Global Digital Elevation Models for Geoid Modelling in Egypt", *Regional Conference on Surveying & Development Sharm El-Sheikh, Egypt*, 3–6 October 2015.
- [5] Amin, M. M., Zaky, K. M., EL Fatairy, S. M., and Habib, M. A., (2013), "Fetching the Most Appropriate Global Geopotential Model for Egypt", *Civil Engineering Research Magazine CERM*, Vol. 35, No. 3, Published by Faculty of Engineering, Al-Azhar University, Cairo, Egypt.
- [6] Arabelos, A., and Tziavos, I. N., (1983), "Determination of Deflection of the vertical using a combination of spherical harmonics and gravimetric data for the area of Greece", *Bull. Géod.* Vol. 57, pp. 240–256.
- [7] Beale, M. H., Hagan, M. T., and Demuth, H. B., (2015). *Neural Network Toolbox User's Guide*, The Math Works, Inc.
- [8] Cakir, L., and Yilmaz, N., (2014), "Polynomial, radial basis functions and multilayer perceptron neural network methods in local geoid determination with GPS/levelling", *J. measurements*, Vol. 57, pp. 48–153.
- [9] Cawley, G. C., and Talbot, N. L. C., (2010), "On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation", *Journal of Machine Learning Research*, Vol. 11, pp. 2079–2107.
- [10] Cross, P. A., (1983), "Advanced least squares applied to position fixing", *North East London Polytechnic Department of Land Surveying*.
- [11] Darbeheshti, N., and Featherstone, W. E., (2009), "Non-stationary covariance function modeling in 2D least-squares collocation", *J. Geod.*, Vol. 83, pp. 495–508, DOI: 10.1007/s00190-008-0267-0.
- [12] Darbeheshti, N., and Featherstone, W. E., (2010), "Tuning a gravimetric quasigeoid to GPS-levelling by non-stationary least-squares collocation", *J. Geod.* Vol. 84, pp. 419–431, DOI: 10.1007/s00190-010-0377-3.
- [13] Dawod, G. M., Mohamed, H. F., and Ismail, S. S., (2010), "Evaluation and Adaptation of the EGM2008 Geopotential Model along the Northern Nile Valley, Egypt: Case Study", *Journal of Surveying Engineering*, Vol. 136, No. 1, DOI: 10.1061/_ASCE_SU.1943-5428.0000002.
- [14] De Brabanter, K., Karsmakers, F., Ojeda, C., Alzate, J., De Brabanter, J., Pelckmans, K., De Moor, Vandewalle, J., and Suykens, J. A. K., (2011), "LS-SVMlab Toolbox User's Guide Version 1.8", *ESAT-SISTA Technical report 10-146*, Katholieke Universiteit Leuven, Belgium, <http://www.esat.kuleuven.be/sista/lssvmlab/>.

- [15] Dermanis, A., (1984), "Kriging and Collocation – A comparison", Manuscript Geodaetica, Vol. 9, pp. 159–167.
- [16] Doganalp S., (2016), "Geoid height computation in strip-area project by using least-squares collocation", Acta Geodyn. Geomater., Vol. 13, No. 2 (182), 167–176, DOI: 10.13168/AGG.2015.0054.
- [17] Egli, R., Geiger, A., Wiget, A., and Kahle, H. G., (2007), "A modified least-squares collocation method for determination of crustal deformation: first results in the Swiss Alps", Geophys. J. Int., Vol. 168, pp. 1–12, DOI: 10.1111/j.1365-246X.2006.03138.x.
- [18] Elshambaky, H. T., (2017), "Application of neural network technique to determine a corrector surface for global geopotential model using GPS/levelling measurements in Egypt", J. Appl. Geodesy, DOI: 10.1515/jag-2017-0017.
- [19] Elshambaky, H. T., Kaloop, M. R., and Hu, J. W., (2018) "A novel three-direction datum transformation of geodetic coordinates for Egypt using artificial neural network approach", Arabian Journal of Geoscience, Vol. 11, p. 110, <https://doi.org/10.1007/s12517-018-3441-6>.
- [20] Emin, A. M., (1993), "Geoid determination in Turkey (TG-91)", Bulletin Géodésique, Vol. 67, pp. 10–22.
- [21] Erol, B., and Erol, S., (2012), "GNSS in practical determination of regional heights", Global navigation satellite systems: Signal, theory and applications, J. Shuanggen, ed., InTech, Croatia, pp. 127–160.
- [22] Espinoza, M., Suykens, J. A. K., and De Moor, B., (2005), "Load forecasting using fixed -size least squares support vector machines", In: Cabestany, J., Prieto, A., Sandoval, F. (eds) "Computational Intelligence and Bioinspired Systems", IWANN 2005, Lecture notes in computer science, Vol. 3512, Springer, Berlin, Hiedelberg, DOI: http://doi.org/10.1007/11494669_125.
- [23] European Space Agency, (2014), "GUT TUTORIAL", Reference: ESA/XGCE-DTEX-EOPS-SW-07- 0001, Version:7.2, Date:19December2014, <https://earth.esa.int/web/guest/software-tools/gut/about-gut/overview>.
- [24] Fan, R. E., Chen, P. H., and Lin, C. J., (2005), "Working set selection using second order information for training support vector machines", Journal of Machine Learning Research, Vol. 6, pp. 1871–1918.
- [25] Fan, R. E., Chen, P. H., and Lin, C. J., (2006), "A study on SMO-type decomposition methods for support vector machines", IEEE Transactions on Neural Networks, Vol. 17, pp. 893–908.
- [26] Featherstone, W. E., (2000), "Refinement of gravimetric geoid using GPS and levelling data", Journal of Surveying Engineering, Vol. 126, No. 2.
- [27] Featherstone, W. E., and Sproule, D. M., (2006), "Fitting AUSGeoid98 to the Australian height datum using GPS-leveling and least squares collocation: Application of cross-validation technique", Survey Review, Vol. 38, No. 301, pp. 573–582.
- [28] Fotopoulos, G., Featherstone, W. E., and Sideris, M. G., (2002), "Fitting a gravimetric geoid model to the Australian height datum via GPS data", IAG Third Meeting of the International Gravity and Geoid Commission, Thessaloniki, Greece, Aug. 26–30, 2002.
- [29] Gullu, M., (2010). "Coordinate Transformation by radial basis function neural network", Scientific Research and Essays, ISSN 1992-2248, Vol. 5, No. 20, pp. 3141–3146, Available online at <http://www.academicjournals.org/SRE>.
- [30] Gullu, M., Yilmaz, M., Yilmaz, I., and Turgut, B., (2011a), "Datum Transformation by Artificial Neural Networks for Geographic Information Systems Applications", International Symposium on Environmental Protection and Planning: Geographic Information Systems (GIS) and Remote Sensing (RS) Applications (ISEPP) 28–29 June 2011, Izmir, Turkey.
- [31] Gullu, M., Yilmaz, M., Yilmaz, I., and Turgut, B., (2011b), "Application of Back Propagation Artificial Neural Network for Modeling Local GPS/Levelling Geoid Undulation: A Comparative Study", Ts007C-Geoid and GNSS Heighting, 5239, FIG Working Week 2011, Bridging the Gap between Cultures, Marrakech, Morocco, 18–22 May 2011.
- [32] Hagan, M. T., H. B. Demuth, and M. H. Beale. (1996). Neural Network Design, PWS Publishing, Boston, MA.
- [33] Haykin, S. (2001). Neural Network: A Comprehensive Foundation, Second Edition. Hamilton, Ontario, Canada.
- [34] Hornik, K. M., Stinchcombe, M., and White, H., (1989), "Multilayer Feedforward Networks are Universal Approximators", Neural Networks, Vol. 2, No. 5, pp. 359–366.
- [35] Hu, W., Sha, Y., and Kuang, S., (2004), "New method for transforming global positioning system height into normal height based on neural network", J. Surv. Eng., Vol. 130, No. 1, pp. 36–39.
- [36] James, G., Witten, D., Hastie, T., and Tibshirani, R., (2013), "An introduction to statistical learning: with application in R", Springer Texts in statistics, Vol. 103, DOI: 10.1007/978-1-4614-7138-7_9.
- [37] Jarmolowski, W., and Bakula, M., (2013), "Two covariance models in Least Squares Collocation (LSC) tested in interpolation of local topography", Contribution to Geophysics and Geodesy, Vol. 43, No. 1, pp. 1–19, DOI: 10.2478/congeo-2013-0001.
- [38] Jarmolowski, W., and Bakula, M., (2014), "Precise estimation of covariance parameters in least -squares collocation by restricted maximum likelihood", Stud. Geophys. Geod., Vol. 58, 1, DOI: 10.1007/s11200-013-1213-z.
- [39] Jordan, S. K., (1972), "Self-consistent statistical models for the gravity anomaly, vertical deflection, and undulation of the geoid", J. Geophys. Res., Vol. 77, No. 20, pp. 3660–3670.
- [40] Kararup, T., (1969), "A contribution of the mathematical foundation of physical geodesy", The Danish Geodetic Institute, Meddelelse, No. 44.
- [41] Kavazoglu, T., and Saka, M. H., (2005), "Modeling local GPS/Leveling geoid undulations using artificial neural networks", J. Geodesy, Berlin, Vol. 78, No. 9, pp. 520–527.
- [42] Kotsakis, C., (2007), "Least -squares collocation with covariance-matching constraints", J. Geodesy, DOI: 10.1007/s00190-007-0133-5.
- [43] Kuroishi, Y., Ando, H., and Fukuda, Y., (2002), "A new hybrid geoid model for Japan, GSIGEO2000", J. Geod., Vol. 76, pp. 428–436, DOI: 10.1007/s00190-002-0266-5.
- [44] Ligas, M., and Banasik, P., (2014), "Least squares collocation alternative to HELMERT's transformation with HAUSBRANDT's post-transformation correction", Reports on Geodesy and Geoinformatics, Vol. 97, pp. 23–34, DOI: 10.2478/rgg-2014-0009.
- [45] Lin, L. S., (2007), "Application of a back-propagation artificial neural network to regional grid-based geoid model generation

- using GPS and leveling data", *J. Surv. Eng.*, 133, pp. 81–89.
- [46] Mikhail, E. M., and Ackermann, F., (1976), "Observations and Least Squares", Dun Donnelly, New York.
- [47] Moritz, H., (1978), "Least – Squares Collocation", *Review of Geophysics and Space Physics*, Vol. 16, pp. 421–430.
- [48] Moritz, H., (1980), "Advanced physical geodesy", Abacus, Tunbridge Wells Kent.
- [49] Mosbeh, R. Kaloop, Rabah, Mostafa, Hu, Jong Wan, and Zaki, Ahmed, (2017), "Using advanced soft computing techniques for regional shoreline geoid model estimation and evaluation", *Marine Georesources & Geotechnology*, DOI: 10.1080/1064119X.2017.1370622.
- [50] Pardo-Igúzquiza, E., (1998), "Maximum likelihood estimation of spatial covariance parameters", *Math. Geol.*, Vol. 30, No. 1, pp. 95–108.
- [51] Pardo-Igúzquiza, E., Mardia, K. V., and Chica-Olomo, M., (2009), "A computer program for maximum likelihood inference with spatial Matérn covariance model", *Comput. Geosci.*, Vol. 35, pp. 1139–1150.
- [52] Pikaridas, C., and Fotiou, A., (2011), "Estimation and evaluation of GPS geoid heights using artificial neural network model", *Appl. Geomat.*, Vol. 3, pp. 183–187, DOI: 10.1007/s12518-011-0052-2.
- [53] Platt, J., (1998), "Sequential minimal optimization: A fast algorithm for training support vector machines", Technical report, MSR-TR-98-14.
- [54] Powell, S. M., (1997), "Results of the Final Adjustment of the New National Geodetic Network", Technical report, Egyptian Surveying Authority, Egypt.
- [55] Ruffhead, A., (1987), "An introduction to least squares collocation", *Survey Review*, Vol. 29, No. 224, pp. 85–94, DOI: 10.1179/sre.1987.29.224.85.
- [56] Sampson, P. D., and Guttorp, P., (1992), "Nonparametric estimation of nonstationary spatial covariance structure", *J. of the American Statistical Association*, Vol. 87, No. 417, pp. 108–119, DOI: 10.2307/2290458.
- [57] Samui, P., Kim, D., and Aiyer, B. G., (2015), "Pullout capacity of small ground anchor: a least square support vector machine approach", *Journal of Zhejiang University – Science A (Applied Physics & Engineering)*, ISSN 1673-565X(print); ISSN 1862-1775(online), www.zju.edu.cn/jzus; www.springerlink.com.
- [58] Schwarz, K. P., and Lachapelle, G., (1980), "Local characteristics of the gravity anomaly covariance function", *Bull. Géod.*, Vol. 54, pp. 21–36.
- [59] Sorkhabi, O. M., (2015), "Geoid determination based on Log Sigmoid Function of artificial neural network: (A case study: Iran)", *J. of Artificial Intelligence in Electrical Engineering*, Vol. 3, No. 12.
- [60] Stopar, B., Ambrozic, T., Kuhar, M., and Turk, G., (2006), "GPS-derived geoid using artificial neural network and least squares collocation", *Survey Review*, Vol. 38, p. 300.
- [61] Suykens J. A. K., and Vandewalle J., (1999), "Least squares support vector machine classifiers", *Neural Processing Letters*, Vol. 9, No. 3, pp. 293–300.
- [62] Suykens J. A. K., (2001), "Support vector machines: A nonlinear modelling and control perspective", *European Journal of Control*, Vol. 7, pp. 311–327.
- [63] Suykens J. A. K., Van Gestel T., De Brabanter J., De Moor, B., and Vandewalle, J., (2002), "Least Squares Support Vector Machines", World Scientific, Singapore.
- [64] Szu-Pyng, K., Chao-Nan, C., Hui-Chi, H., and Yu-Ting, S., (2014), "Using a least squares support vector machine to estimate a local geoid model", *Bol. Ciênc. Geod.*, Vol. 20, No. 2, DOI: 10.1590/S1982-21702014000200025.
- [65] Tierra, A., Dalazoana, R., and De Freitas, S., (2008), "Using Artificial Neural Network to Improve the Transformation of Coordinates Between Classical Geodetic Reference Frames", *Computers & Geosciences*, pp. 181–189, DOI:10.1016, Netherlands.
- [66] Tscheerning, C. C., (2010), "The use of Least-Squares Collocation for the processing of GOCE data", *Vermessung & Geoinformation*, Vol. 1, pp. 21–26.
- [67] Tziavos, I. N., (1987), "Determination of geoidal heights and deflections of the vertical for the Hellenic area using heterogeneous data", *Bull. Géod.*, Vol. 61, pp. 177–197.
- [68] Vapnik, N. V., (1998), "Statistical Learning Theory", John Wiley & Sons, New York.
- [69] Veronez, M. R., Thum, A. B., and De Souza, G. C., (2006), "A new method for obtaining geoidal undulation through artificial neural networks", 7th International Symposium on Spatial Accuracy Assessment in Neural Resources and Environmental Sciences, pp. 306–316.
- [70] Vestøl, O., (2006), "Determination of postglacial land uplift in Fennoscandia from leveling, tide-gauges and continuous GPS stations using least squares collocation", *J. Geod.*, Vol. 80, pp. 248–258, DOI: 10.1007/s00190-006-0063-7.
- [71] Wang, J., Y. Hu, and J. Zhou, 2009, Combining model for regional GPS height conversion based on least squares support vector machines, *Proceedings – 2009 International Conference on Environmental Science and Information Application Technology*, ESIAT 2009, Vol. 2(2), pp. 639–641. DOI: 10.1109/ESIAT.2009.182.
- [72] Yang, Y., Zeng, A., and Zahang, J., (2009), "Adaptive collocation with application in height system transformation", *J. Geod.*, Vol. 83, pp. 403–410, DOI: 10.1007/S00190-008-0226-9.
- [73] Ye, J., and Xiong, T., (2007), "SVM versus least squares SVM", *Proc. 7th artificial intelligence and statistics*, 21–24 March 2007, San Juan, Puerto Rico, Vol. 2, pp. 644–651.
- [74] You, R. J., and Hawang, H. W., (2006), "Coordinate transformation between two geodetic datums of Taiwan by least-squares collocation", *Journal of Surveying Engineering*, Vol. 132, No. 2, pp. 64–70.
- [75] Zaletnyik, P., Völguesi, L., and Paláncz, B., (2008), "Modelling local GPS/levelling geoid undulations using Support Vector Machine", *Periodica Polytechnica, Civil Engineering*, Vol. 52, No. 1, pp. 39–43, web: <http://www.pp.bme.hu/ci>.
- [76] Ziggah, Y. Y., Youjian, H., Tierra, A., and Konaté, A. A., (2016a), "Performance evaluation of artificial neural networks for planimetric coordinate transformation – a case study, Ghana", *Arab. Geosci.*, Vol. 9, p. 698, DOI: 10.1007/s12517-016-2729-7.
- [77] Ziggah, Y. Y., Youjian, H., Yu, X., and Basommi, L. P., (2016b), "Capability of artificial neural network for forward conversion of geodetic coordinates (ϕ , λ , h) to Cartesian coordinates (x , y , z)", *Math. Geosci.*, DOI: 10.1007/s11004-016-9638-x.