



## ORIGINAL PAPER

GEOID UNDULATION PREDICTION USING GAUSSIAN PROCESSES REGRESSION:  
A CASE STUDY IN A LOCAL REGION IN TURKEYBerkant KONAKOGLU <sup>1)</sup> \* and Alper AKAR <sup>2)</sup><sup>1)</sup>Amasya University, Technical Sciences Vocational School, Department of Architecture and Urban Planning, Amasya, Turkey<sup>2)</sup>Erzincan Binali Yıldırım University, Vocational School, Department of Architecture and Urban Planning, Erzincan, Turkey\*Corresponding author's e-mail: [berkantkonakoglu@amasya.edu.tr](mailto:berkantkonakoglu@amasya.edu.tr)

## ARTICLE INFO

## Article history:

Received 21 October 2020

Accepted 28 December 2020

Available online 14 January 2021

## Keywords:

Gaussian Process Regression

Generalized regression neural network

Geoid

GNSS/leveling

Radial basis function neural network

Inverse distance to a power

## ABSTRACT

In order to convert ellipsoidal heights obtained by the Global Navigation Satellite System (GNSS) to orthometric heights, it is necessary to know the distance between the ellipsoidal and geoid surface, called the geoid undulation. The geoid undulation can be predicted using emerging mathematics tools and algorithms. The objective of this study was to develop a model for predicting the geoid undulation using Gaussian Process Regression (GPR), one of the soft machine learning algorithms having different covariance functions. This method was then compared with the radial basis function neural network (RBFNN), generalized regression neural network (GRNN), and the interpolation method of inverse distance to a power (IDP) with the power of 1, 2, 3, 4, and 5. First, 70 % of GNSS/leveling data (422 points) were used in the training phase. The remaining 185 points were used as testing data to check the effectiveness of the constructed model. In the GPR modeling, ten covariance functions (Materniso d= 1, 3, 5; Maternard d= 1, 3, 5; SEiso; SEard; RQiso; and RQard) were tested for prediction on this dataset. The GPR based on the Materniso (d = 1) covariance function model was introduced as an effective method for predicting geoid undulation and provided the best results (RMSE = 8.32 cm, MAE = 5.51 cm, R<sup>2</sup> = 0.98968) when compared with the other developed GPR models. In addition, the statistical findings showed that the accuracy of all the GPR models was also better in predicting geoid undulation than the RBFNN, GRNN, and IDP with the power of 1, 2, 3, 4, and 5.

## 1. INTRODUCTION

The use of the Global Navigation Satellite System (GNSS) can provide geocentric coordinates ( $X, Y, Z$ ) or geodetic coordinates ( $\varphi, \lambda, h$ ) for any point on earth, based on the WGS84 ellipsoid. In geodetic applications, elevation is measured with reference to the surface of a geoid as orthometric height ( $H$ ). Hence, GNSS-derived ellipsoidal heights must be transformed into orthometric heights. The mathematical relation between these heights is as follows (Heiskanen and Moritz, 1967):

$$H = h - N \quad (1)$$

where  $h$  is the ellipsoidal height,  $H$  is the orthometric height, and  $N$  is the geoid undulation or geoid height. The geoid undulation determined by the GNSS/leveling technique can also be used as data in the determination of the geoid via various estimation techniques such as the interpolation, least-squares collocation (LSC) and geophysical gravity data inversion (GGI) methods (Kotsakis and Sideris, 1999; Tscherning et al., 2001; Erol et al., 2008; Trojanowicz, 2012, 2015; Doganalp and Selvi, 2015; Doganalp, 2016; Karaaslan et al., 2016; Tusat and Mikailsoy, 2018; Trojanowicz et al., 2020). Several soft-

computing methods have been applied as alternative techniques to produce accurate results in the prediction of geoid undulation, including the artificial neural network (ANN), least square support vector machines (LS-SVM), and adaptive neuro-fuzzy inference systems (ANFIS). Among these methods, the ANN has been adopted by the geodetic community. An important characteristic of ANNs is their ability to learn from a dataset (training data). When the network is trained with a sufficient number of training data, predictions can be made based on previous learning. A few studies have used the ANN with a different network topology in the modeling of a local GNSS/leveling geoid surface (Lin, 2007; Akcin and Celik, 2013; Erol and Erol, 2013; Albayrak et al., 2020). For example, Kavzoglu and Saka (2003) predicted geoid undulation for Istanbul, Turkey using a feed-forward ANN learning algorithm for the application. The derived results were compared with polynomial surface fitting and least-squares collocation (LSC). The prediction accuracy was evaluated using the root mean square error (RMSE). The results stated that the ANN was capable of yielding results that were as accurate as the polynomial surface fitting and LSC methods. Stopar et

al. (2006) proposed a method of geoid undulation approximation that coupled the LSC and the ANN (ANN-LSC). It was concluded from this research that the ANN-LSC method could be considered as an alternative to the existing geoid or quasi-geoid models, especially for areas where correct models are not available. Veronez et al. (2011) developed an ANN model (trained using the Levenberg-Marquardt training algorithm) to estimate the geoid of the state of São Paulo, Brazil. The results of the model indicated that the ANN had improved the geoid undulation prediction accuracy by 40 % and that it could be used in predicting geoid undulation. Cakir and Yilmaz (2014) compared multi-layer perceptron neural network (MLPNN) estimates with polynomials and radial basis functions (RBFs) including multiquadric (MQ) and thin plate spline (TPS). Based on the results of this research, the MLPNN model was found to be more accurate than the other methods tested. Another soft-computing method is represented by the least square support vector machine (LS-SVM), which is the least squares formulation of the support vector machine (SVM) and has been applied to detect a geoid model (Zaletnyik et al., 2008; Kao et al., 2014; Kaloop et al., 2018). In addition to ANN and LS-SVM methods, the ANFIS approach has also been used by geodesy researchers in geoid modeling (Akyilmaz et al., 2008; Yilmaz, 2010; Yilmaz and Arslan, 2013).

Gaussian Process Regression (GPR) is a non-parametric probabilistic kernel-based approach for regression problems. Handling modeling problems with noisy data is the main advantage of GPR. The GPR has been utilized in predicting total electron content (TEC) values (Ackermann et al., 2011; Albert et al., 2020; Inyurt et al., 2020). However, few studies have been carried out on the use of GPR in estimating geoid undulation. Kaloop et al. (2019) used GPR for the first time in modeling geoid undulation. The authors utilized the GPR model for the prediction of geoid undulation and compared this model with the least square support vector regression (LSSVR), kernel ridge regression (KRR), and multivariate adaptive regression splines (MARS) methods. Subsequently, the accuracy of these models was compared with a local gravity geoid model and global geopotential models (GGMs) that had been applied in Kuwait. The results revealed that the KRR model was the best for estimating the geoid undulation of Kuwait. The GPR method also gave acceptable results, whereas the MARS yielded the worst results. To date, no study has been found in the literature comparing the GPR technique with any of the ANN or classical interpolation methods in order to verify its reliability in predicting geoid undulation. Hence, in order to fill this gap, this study aimed to compare the GPR model with the radial basis function neural network (RBFNN), generalized regression neural network (GRNN), and one interpolation method including inverse distance to a power (IDP) with the power of 1, 2, 3, 4, and 5.

The structure of this paper is as follows. In Section 2 a brief description of the study area and the data used in this study are given and the basic theoretical aspects of GPR, RBFNN, GRNN and IDP are provided. Section 3 presents a description of the statistical indicators used in this study to evaluate the prediction performance of the models. Section 4 provides the results and discussion. Finally, the conclusions of this study are presented in Section 5.

## 2. MATERIAL AND METHODS

### 2.1. DESCRIPTION OF DATA

This study was carried out in the province of Ardahan, which is located in northeastern Turkey at the latitude of 41°6'46.62"N and longitude of 42°42'8.21"E using GNSS/leveling data. The topography is irregular with orthometric height changes between 1306.49 m and 2959.81 m and geoid undulations between 23.46 m and 26.54 m. The orthometric heights of the points were determined using the geometric levelling method in the datum of Turkey National Vertical Control Network 1999 (TNVCN99). The geographic coordinates including ellipsoidal heights ( $h$ ) were determined in static positioning mode and referred to the Turkish National Fundamental GPS Network (TNFGN). The GNSS coordinates of the points were derived in the ITRF96 datum. The accuracies of the orthometric and ellipsoidal heights were  $\pm 2$  cm and  $\pm 3.4$  cm, respectively. The 607 C3 (third order densification) points covering the entire area of Ardahan measured by the Erzincan XXIV Regional Directorate of the Turkish Land Registry and Cadastre were used for the determination of the local geoid surface in Ardahan, as presented in Figure 1. The distribution of the points, with a density of 1 point per 13 km<sup>2</sup>, demonstrates a good characterization of the topography.

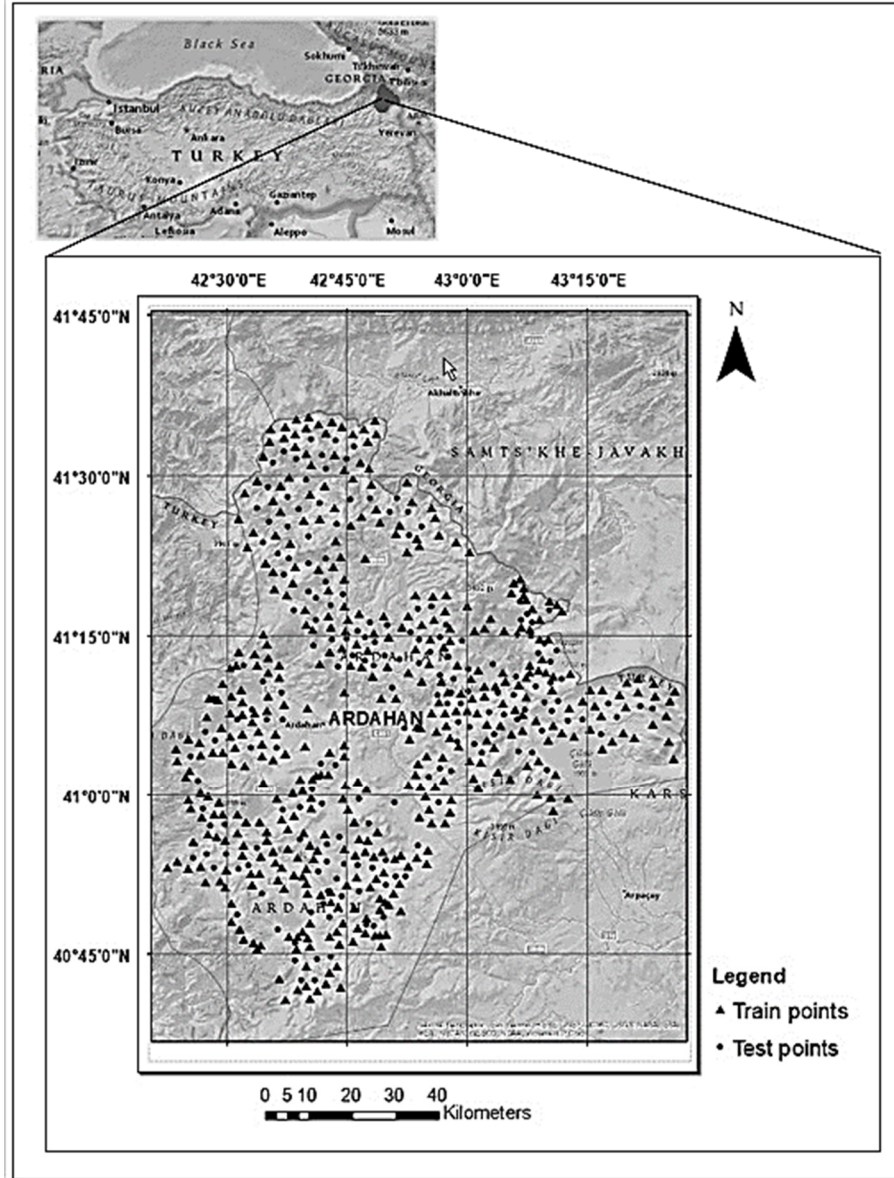
For training, 422 GNSS/leveling points (i.e., train points) were used, remaining 185 GNSS/leveling points ( $\sim 30$  % of the data) were used for testing the performance.

### 2.2. GAUSSIAN PROCESS REGRESSION (GPR)

Gaussian Process Regression (GPR) is a flexible non-parametric method that uses the prior of a Gaussian process to perform regression of data (Yuan et al., 2008). The kernel function based on Bayes linear regression is adopted instead of a linear kernel function. This model can be used to solve complex, nonlinear regression problems and classification problems (Lin et al., 2019). The statistical characteristics of the Gaussian process is characterized by its mean and covariance function. Here, we define the mean function  $m(x)$  and covariance function  $k(x, x')$  of a real process  $f(x)$  as follows:

$$m(x) = E[f(x)] \quad (2)$$

$$k(x, x') = E[(f(x) - m(x))(f(x') - m(x'))] \quad (3)$$



**Fig. 1** The distributions of the train and test points in the study area (map produced using ArcGIS).

where  $x$  and  $x' \in R^K$  are independent variables, and  $E[\cdot]$  denotes the expectation. The Gaussian process can be written as below

$$y = f(x) \sim GP(m(x), k(x, x')) \quad (4)$$

The general regression model with noise can be expressed as follow:

$$y_i = f(x_i) + \varepsilon, \varepsilon \sim N(0, \sigma_n^2) \quad (5)$$

where  $x_i$  is the input vector,  $f$  is the function value, and  $y_i$  represent the noisy observations. It is assumed that the additive white noise  $\varepsilon$  is independent and subjected to a Gaussian distribution with a mean of zero and a variance of  $\sigma_n^2$  (Lin et al., 2019).

Given a training set  $D = (X, y)$ , the target is to predict the function values  $f_*$  given inputs  $X_*$ . Being a linear combination of Gaussian variables,

$y$  is also Gaussian, with distribution  $y \sim N(m(X), K(X, X) + \sigma^2 I)$ , where  $I$  is the identity matrix. The data-set is generally preprocessed with a mean of zero ( $m(X) = 0$ ) (Rasmussen and Nickisch, 2010; Chen et. al., 2013). The joint Gaussian prior distribution of the training data and the predicted output is given using Equation (6) (Williams and Rasmussen, 2006).

$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim N \left( 0, \begin{bmatrix} m(X) \\ m(x_*) \end{bmatrix}, \begin{bmatrix} K(X, X) + \sigma^2 I & K(X, x_*) \\ K(x_*, X) & K(x_*, x_*) \end{bmatrix} \right) \quad (6)$$

where  $K(X, X)$  is the covariance matrix of  $N \times N$ ,  $K(x_*, X)$  is the covariance matrix of  $1 \times N$ , and  $[K(x_*, X)]_{1j} = k(x_*, x_j)$ ;  $K(X, x_*)$  is the covariance matrix of  $N \times 1$  and  $[K(X, x_*)]_{j1} = k(x_j, x_*)$  is the covariance of the test input vector  $x_*$ .

The principle of joint Gaussian distributions allows the prediction results for the target to be inferred from the mean function  $\bar{y}_*$  and the covariance function  $cov(y_*)$  given by Equations (7) and (8), respectively.

$$\bar{f}_* = m(x_*) + K(x_*, X)[K(X, X) + \sigma^2 I]^{-1}(y - m(X)) \quad (7)$$

$$cov(f_*) = K(x_*, x_*) - K(x_*, X)[K(X, X) + \sigma^2 I]^{-1}K(X, x_*) \quad (8)$$

where  $\bar{f}_*$  is the mean value of  $f_*$ , and  $cov(f_*)$  is the covariance of  $f_*$ . If the noise  $\varepsilon$  is considered, the posterior distribution of  $y_*$  is as follows:

$$y_*|X, y, x_* \sim N(\bar{f}_*, cov(f_*) + \sigma^2) \quad (9)$$

The appropriate covariance function  $k(x, x')$  plays an important role in the modeling process and should be chosen carefully because the covariance function affects the smoothness, length, scale, amplitude, etc. Under the Gaussian process view, the covariance function is used to measure nearness or similarity among the input observations. It is a fundamental similarity assumption that input points are likely to have similar target values, and thus, the training points that are close to a test point should provide informative about the prediction. There are several common covariance functions available in the literature; however, the squared exponential (SE) is the one most frequently utilized. This covariance function is basically of two types: the squared exponential covariance function with isotropic distance measure (covSEiso) and the squared exponential covariance function with automatic relevance detection (covSEard), respectively, as expressed in Equations (10) and (11) (Williams and Rasmussen, 2006):

$$k_{SEiso}(x, x') = \sigma_f^2 \exp\left[-\frac{(x-x')^T(x-x')}{2l^2}\right] \quad (10)$$

and

$$k_{SEard}(x, x') = \sigma_f^2 \exp\left[-\frac{(x-x')^T \Lambda^{-2}(x-x')}{2l^2}\right] \quad (11)$$

where  $\sigma_f^2$ ,  $\alpha$ ,  $\Lambda$ , and  $l$  are the signal variance and the shape parameter for the rational quadratic covariance and  $\Lambda$  is  $l$  times the unit matrix and the characteristic length scale, respectively.

Similarly, for the covariance functions, in addition to the basic two above, four different forms were fitted: the Matérn covariance function with automatic relevance detection (covMaternard), the Matérn covariance function with isotropic distance measure (covMaterniso), the rational quadratic covariance function with automatic relevance detection (covRQard), and the rational quadratic covariance function with isotropic distance measure (covRQiso). These additional four covariance functions were used in this study as follows (Rasmussen and Nickisch, 2010):

- Matérn covariance function with automatic relevance detection (covMaternard)

$$k_{Maternard}(x, x') = \sigma_f^2 f_d(r_d) \exp(-r_d), \quad r_d = \sqrt{d(x-x')^T \Lambda^{-2}(x-x')} \quad (12)$$

where  $f$  is a function of  $r_d$ ,  $d = 1, 2, 3$ .

- Matérn covariance function with isotropic distance measure (covMaterniso)

$$k_{Materniso}(x, x') = \sigma_f^2 f_d(r_d) \exp(-r_d), \quad r_d = \sqrt{\frac{d}{l^2}(x-x')^T(x-x')} \quad (13)$$

- Rational quadratic covariance function with automatic relevance detection (covRQard)

$$k_{RQard}(x, x') = \sigma_f^2 \left(1 + \frac{1}{2\alpha}(x-x')^T \Lambda^{-2}(x-x')\right) \quad (14)$$

where  $\alpha$  is the shape parameter for the rational quadratic covariance.

- Rational quadratic covariance function with isotropic distance measure (covRQiso)

$$k_{RQiso}(x, x') = \sigma_f^2 \left(1 + \frac{1}{2\alpha l^2}(x-x')^T(x-x')\right) \quad (15)$$

The hyper parameters in the covariance function can be compressed into a matrix denoted by  $\phi$ . These parameters are all unknown and can be obtained from the training data by inference. In order to get the appropriate hyper parameters, optimization with the maximization of the log-likelihood function can be utilized (Rasmussen and Nickisch, 2010).

$$\log p(y|X, \theta) = -\frac{1}{2} y^T K(X, X)^{-1} y - \frac{1}{2} \log |K(X, X)| - \frac{n}{2} \log(2\pi) \quad (16)$$

### 2.3. RADIAL BASIS FUNCTION NEURAL NETWORK

The radial basis function neural network (RBFNN) was proposed by D. S. Broomhead and D. Lowe in the late 1980s. It has three layers consisting of an input layer, one hidden layer, and an output layer. A non-linear RBF activation function is used in the hidden layer. As with the multi-layer perceptron neural network (MLPNN), the RBFNN has the advantage of not suffering from local minima (Haykin, 1999). In addition, the RBFNN is good at modeling nonlinear data and can be trained in one stage rather than using an iterative process as with MLPNN. It also learns the given application quickly (Venkatesan and Anitha, 2006). The Gaussian function used in this study is one of the most popular activation functions in the literature (Bishop, 1995). The mathematical formulation of this function can be found in Kisi (2009). For more details about the RBFNN, see Kisi et al. (2015).



#### 2.4. GENERALIZED REGRESSION NEURAL NETWORK

The generalized regression neural network (GRNN), as a modified form of the RBFNN, was proposed by Specht (1991). The GRNN approximates any arbitrary function between the input and output variables, drawing the function estimate directly from the training data (Zounemat-Kermani, 2014). Like the RBFNN, this network does not require an iterative procedure. It has four layers including the input layer, pattern layer, summation layer, and output layer (Specht, 1993). The GRNN has the advantages of fast training, simple structure, and global convergence (Yuan et al., 2019). The main parameter in the GRNN training process is the spread parameter ( $\sigma$ ), which directly affects the prediction accuracy. This parameter is often chosen by trial-and-error experiments (Haidar et al., 2011). Further details regarding the GRNN can be found in Kisi (2008) and Ladlani et al. (2012).

#### 2.5. INVERSE DISTANCE TO A POWER

The inverse distance to a power (IDP) method predicts the value for any test point by interpolating the measured values in the neighboring reference location (Franke and Nielson, 1980). The weight of a reference point is a function of the horizontal distance from the test point (Yanalak, 2004). The weighting function is assigned to the data using the weighting power ( $k$ ). This value is chosen by the user (i.e.,  $k = 1, 2, 3, 4$ , or  $5$ ) (Karaaslan et al., 2016). More descriptive information can be found in the article by Yanalak and Baykal (2003).

#### 3. PERFORMANCE EVALUATION OF MODELS

The performance of the GPR, RBFNN, GRNN, and IDP models was assessed based on three statistical indicators: the root means square error (RMSE), mean absolute error (MAE), and the coefficient of determination ( $R^2$ ). The equations for the statistical indicators are given as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (O_i - P_i)^2}{n}} \quad (17)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |O_i - P_i| \quad (18)$$

$$R^2 = \left( \frac{\sum_{i=1}^n (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2 \sum_{i=1}^n (P_i - \bar{P})^2}} \right)^2 \quad (19)$$

where  $n$  is the number of samples,  $O_i$  are the observed geoid undulation values,  $P_i$  are the predicted geoid undulation values,  $\bar{O}$  is the mean of the observed geoid undulation values, and  $\bar{P}$  is the mean of the predicted geoid undulation values. Low RMSE and MAE values indicate good prediction, whereas the coefficient of determination ( $R^2$ ) value closest to one (1) indicates the best prediction.

#### 4. RESULTS AND DISCUSSION

In this study, Gaussian Process Regression (GPR) was used to predict the geoid undulation. The optimum GPR model was then compared to those of the RBFNN, GRNN, and IDP and the best model was determined. The geodetic latitudes ( $\varphi$ ) and longitudes ( $\lambda$ ) of the points were chosen as inputs, while the geoid undulation ( $N$ ) was selected as the output of the models. Gaussian Processes for Machine Learning (GPML) program that enables users to construct a GPR model. In addition, a computer code for developing the RBFNN and GRNN models was written in MATLAB software by the authors. Surfer software was utilized for the interpolation-based computation (IDP). Before training the models, it was necessary to normalize data to avoid reduction in network speed and accuracy. In this present work, the GNSS/leveling data were normalized between 0 and 1.

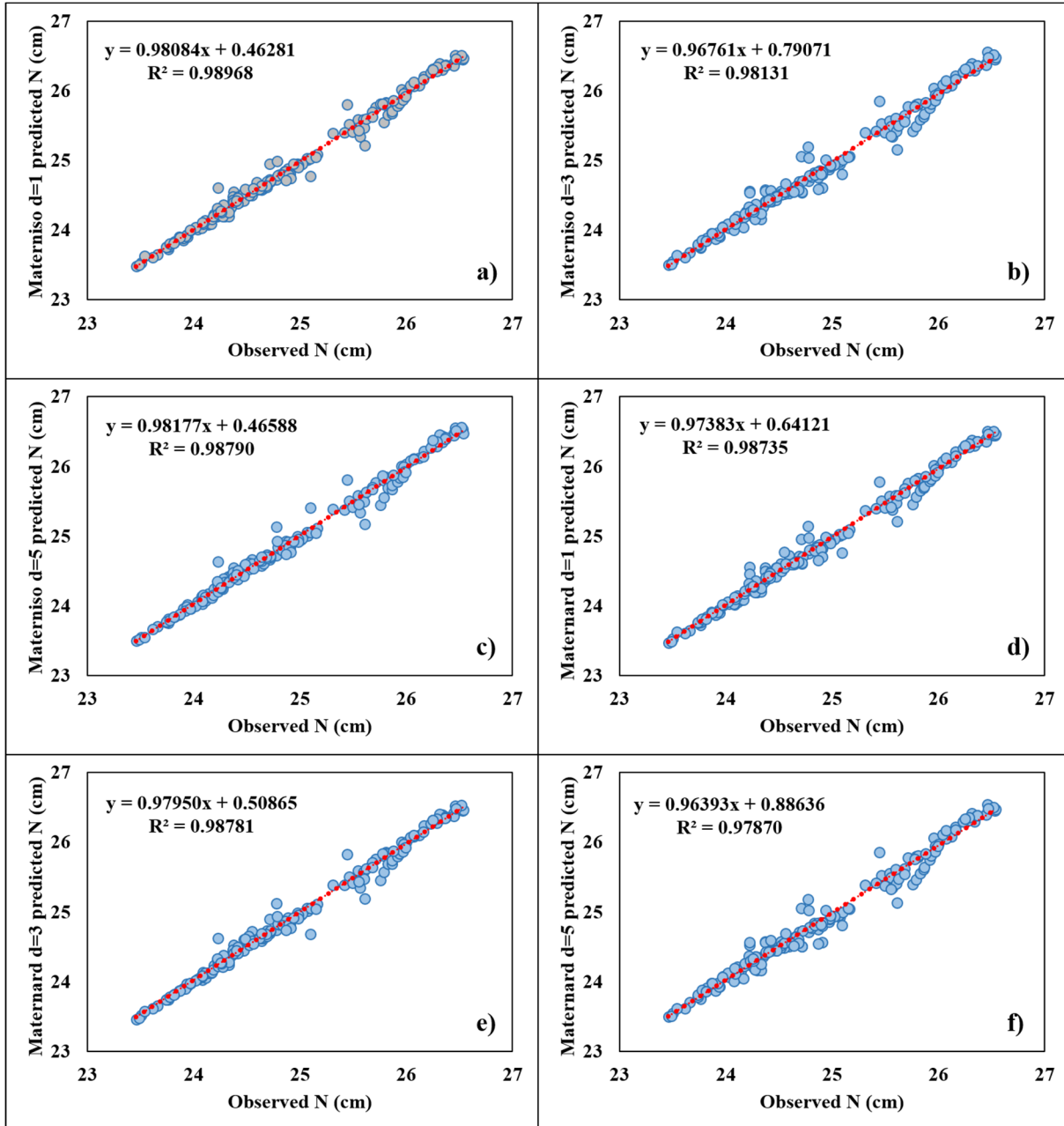
##### 4.1. GPR MODEL RESULTS BASED ON DIFFERENT COVARIANCE FUNCTIONS

During the GPR model development stage, it was necessary to select a covariance function. Determining the optimum covariance function is vital in producing reliable predictions close to actual true values. Thus, in this study, when developing the GPR models to predict the geoid undulation based on the GNSS/leveling data, different covariance functions were considered, including the Matérn covariance function with automatic relevance detection (covMaternard) ( $d = 1, 3, 5$ ), the Matérn covariance function with isotropic distance measure (covMaterniso) ( $d = 1, 3, 5$ ), the rational quadratic covariance function with automatic relevance detection (covRQard), the rational quadratic covariance function with isotropic distance measure (covRQiso), the squared exponential covariance function with automatic relevance detection (covSEard), and the squared exponential covariance function with isotropic distance measure (covSEiso). The hyper parameters were obtained by maximizing the value of the log-likelihood function. The prediction results of the developed GPR models using different covariance functions are given in Table 1. Table 1 shows that the covariance functions utilized gave similar predictive results. However, the developed GPR model based on the Materniso ( $d = 1$ ) covariance function performed the best (RMSE = 8.32 cm, MAE = 5.51 cm,  $R^2 = 0.98968$ ) compared to the other covariance functions. This can also be confirmed by Figures 2 and 3, which show the correlation between the measured geoid undulation and the prediction from the GPR models. In contrast, the same figures show that the GPR based on the SEard covariance function model was found to be the worst (RMSE = 14.68 cm, MAE = 10.74 cm,  $R^2 = 0.97004$ ) compared to the other covariance function models. From a comprehensive perspective, the Matérn-based GPR models performed slightly better in their RMSE, MAE, and  $R^2$  values than the SE-based and RQ-based GPR models.

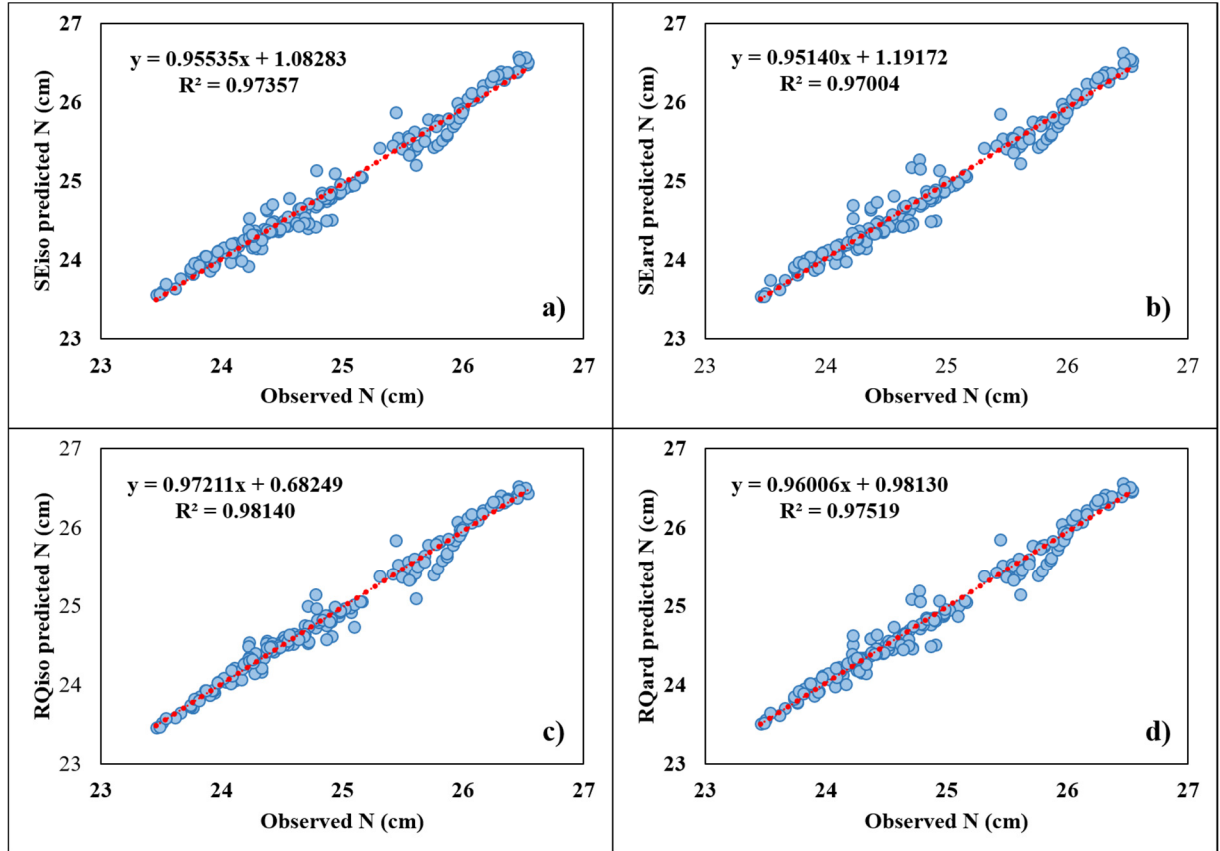
**Table 1** Performance of GPR models based on different covariance functions in predicting geoid undulation.

Covariance functions	RMSE (cm)	MAE (cm)	R <sup>2</sup>
<b>Materniso (d=1)</b>	<b>8.32</b>	<b>5.51</b>	<b>0.98968</b>
Materniso (d=3)	11.60	7.85	0.98131
Materniso (d=5)	9.28	6.08	0.98790
Maternard (d=1)	9.53	6.26	0.98735
Maternard (d=3)	9.25	5.61	0.98781
Maternard (d=5)	12.32	8.58	0.97870
SEiso	13.99	10.68	0.97357
SEard	14.68	10.74	0.97004
RQiso	11.49	7.75	0.98140
RQard	13.30	9.42	0.97519

RMSE: root mean square error; MAE: mean absolute error; R<sup>2</sup>: coefficient of determination; Boldface: the best results



**Fig. 2** Scatter plots of the observed geoid undulation values versus the predicted values of geoid undulation: (a) Materniso d=1, (b) Materniso d=3, (c) Materniso d=5, (d) Maternard d=1, (e) Maternard d=3, (f) Maternard d=5.



**Fig. 3** Scatter plots of the observed geoid undulation values versus the predicted values of geoid undulation: (a) SEiso, (b) SEard, (c) RQiso, (d) RQard.

**Table 2** Comparison of the best GPR model with several other models.

Types of models	RMSE (cm)	MAE (cm)	R <sup>2</sup>
<b>GPR (Materniso (d=1))</b>	<b>8.32</b>	<b>5.51</b>	<b>0.98968</b>
RBFNN	28.02	19.18	0.89095
GRNN	18.30	7.42	0.95244
IDP (k=1)	15.18	10.65	0.97033
IDP (k=2)	13.12	7.66	0.97605
IDP (k=3)	14.48	7.19	0.96996
IDP (k=4)	17.24	7.58	0.95762
IDP (k=5)	20.33	8.27	0.94188

RMSE: root mean square error; MAE: mean absolute error; R<sup>2</sup>: coefficient of determination; Boldface: the best results

Figure 2 illustrates the scatter plots of the observed geoid undulation versus the predicted outputs of the GPR based on the Matérn covariance function model with automatic relevance detection (covMaternard) ( $d = 1, 3, 5$ ) and the GPR based on the Matérn covariance function model with isotropic distance measure (covMaterniso) ( $d = 1, 3, 5$ ); Figure 3 depicts the scatter plots of the measured geoid undulation versus the predicted outputs of the GPR based on the rational quadratic covariance function model with automatic relevance detection (covRQard), the rational quadratic covariance function model with isotropic distance measure (covRQiso), the squared exponential covariance function model with automatic relevance detection (covSEard), and the squared exponential covariance

function model with isotropic distance measure (covSEiso).

The relation of the observed and predicted geoid undulation values can be seen in the points scattered around the dotted red line. For each scatterplot, the slope of the line reaches closer to the value 1 and the intercept moves closer to the value 0, indicating that the predicted values follow their corresponding observed ones very closely. It is clear from Figure 2 that the predictions provided by the GPR based on the Materniso ( $d = 1$ ) covariance function model were highly correlated and less scattered than the other GPR predictions based on the Matérn covariance function models, whereas the predictions produced by the GPR based on the Maternard ( $d = 5$ ) covariance function model were less correlated and more highly scattered

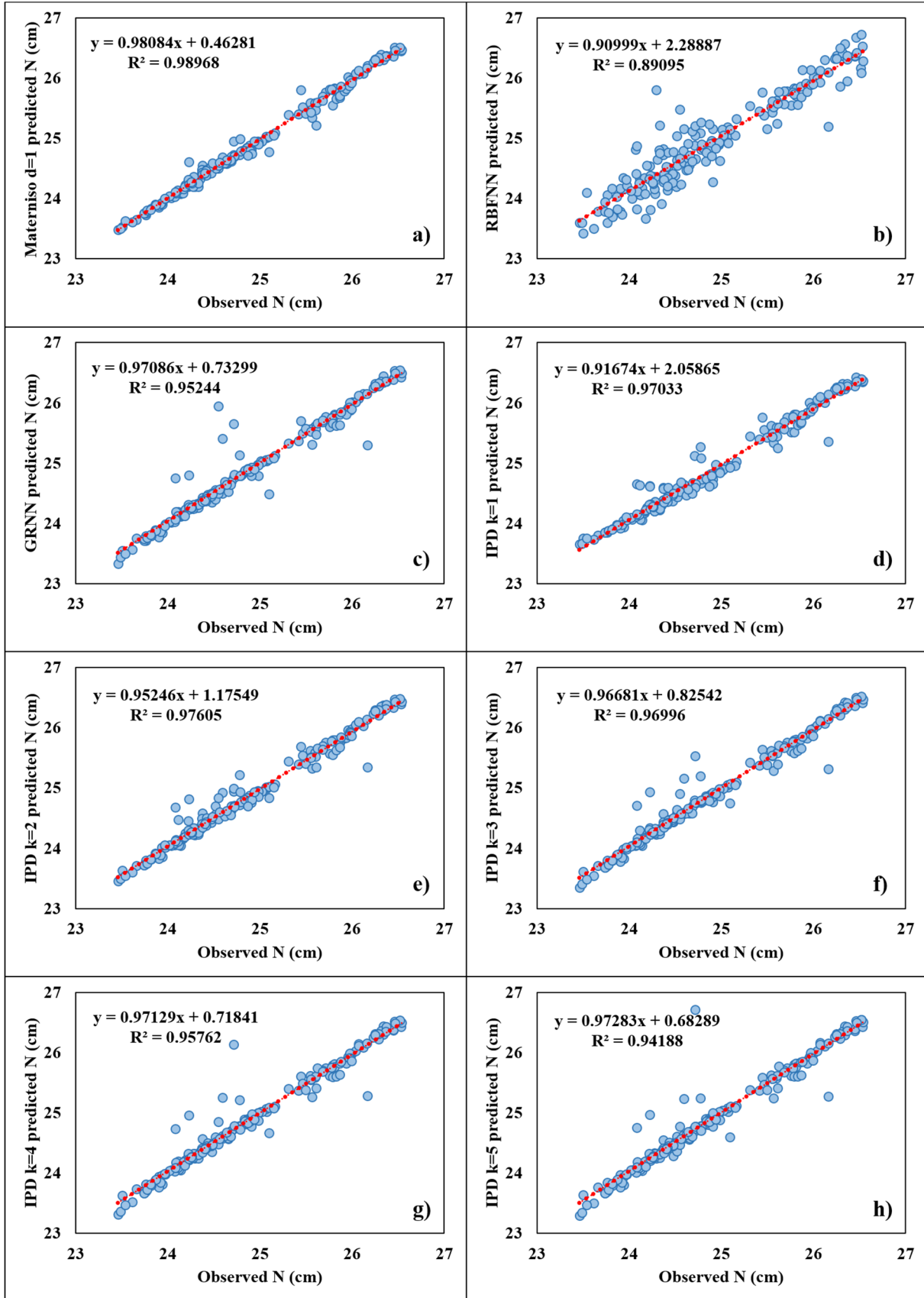


Fig. 4 Scatterplots of the different models in predicting geoid undulation: (a) Materniso d=1, (b) RBFNN, (c) GRNN, (d) IDP k=1, (e) IDP k=2, (f) IDP k=3, (g) IDP k=4, (h) IDP k=5.



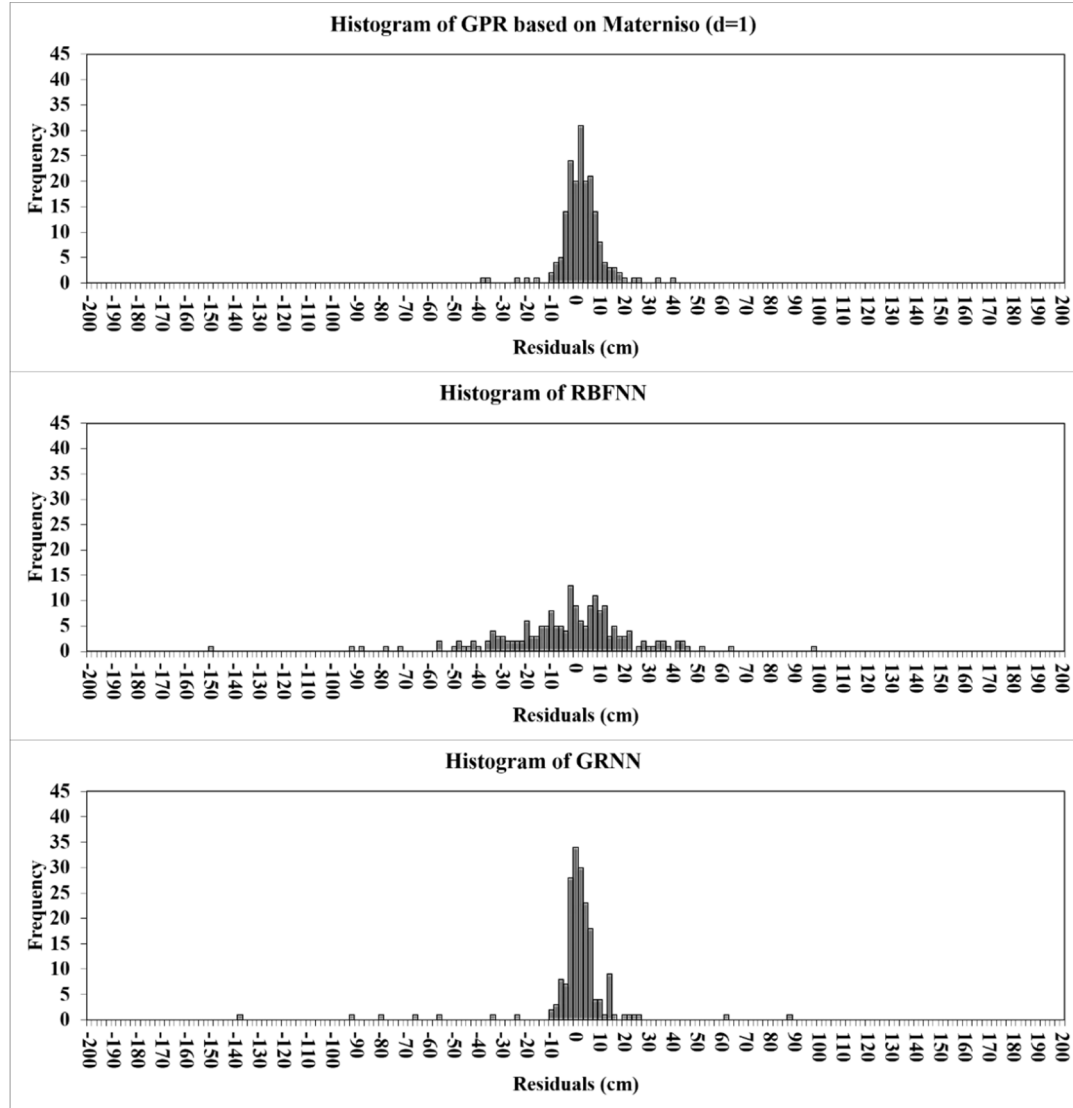


Fig. 5 Residuals versus number of samples in GPR based on the Materniso ( $d=1$ ), RBFNN and GRNN models.

than the other GPR predictions based on the Matérn covariance function models. In addition, from Figure 3, it can be seen that the predictions provided by the GPR based on the RQiso covariance function model were highly correlated and less scattered than the other GPR predictions based on the RQ and SE covariance function models, whereas the predictions produced by the GPR based on the SEard covariance function model were less correlated and more highly scattered compared to the other RQ and SE covariance function models. In general, the slope and intercept ratios of the Matérn-based GPR models were closer to 1 and 0, respectively, compared to the RQ- and SE- based GPR models.

#### 4.2. COMPARISON OF BEST GPR MODEL AND OTHER INVESTIGATED MODELS

For prediction, the GNSS/leveling dataset was used in the GPR as well as in the RBFNN, GRNN, and IDP models. The RBFNN has three layers: input, hidden, and output. In the RBFNN, the spread parameter ( $\sigma$ ) and the number of neurons in the hidden

layer are the parameters that must be determined during the training phase. The spread parameter and the number of neurons were selected using the results giving the lowest RMSE and MAE, and the highest  $R^2$  values. As a result of the experiments, the optimum RBFNN model led to a structure with 2 inputs, 10 hidden neurons and 1 output, i.e., [2-10-1]. Unlike the structure of the RBFNN, a basic GRNN has four layers: input, pattern, summation, and output. Before training the GRNN model, we needed to determine the optimal spread parameter ( $\sigma$ ), which affects the fitness in the GRNN model. The optimal spread parameter ( $\sigma$ ) was found to be 0.032, with the lowest RMSE and the highest  $R^2$ , as in the RBFNN. The error statistics for the predicted geoid undulation were calculated using different weighting power ( $k$ ) parameter choices (i.e., 1, 2, 3, 4, and 5) in order to identify the optimal weighting power parameter. A summary of the results of the best GPR model developed compared with those of the other investigated models based on the RMSE, MAE, and  $R^2$  is presented in Table 2.

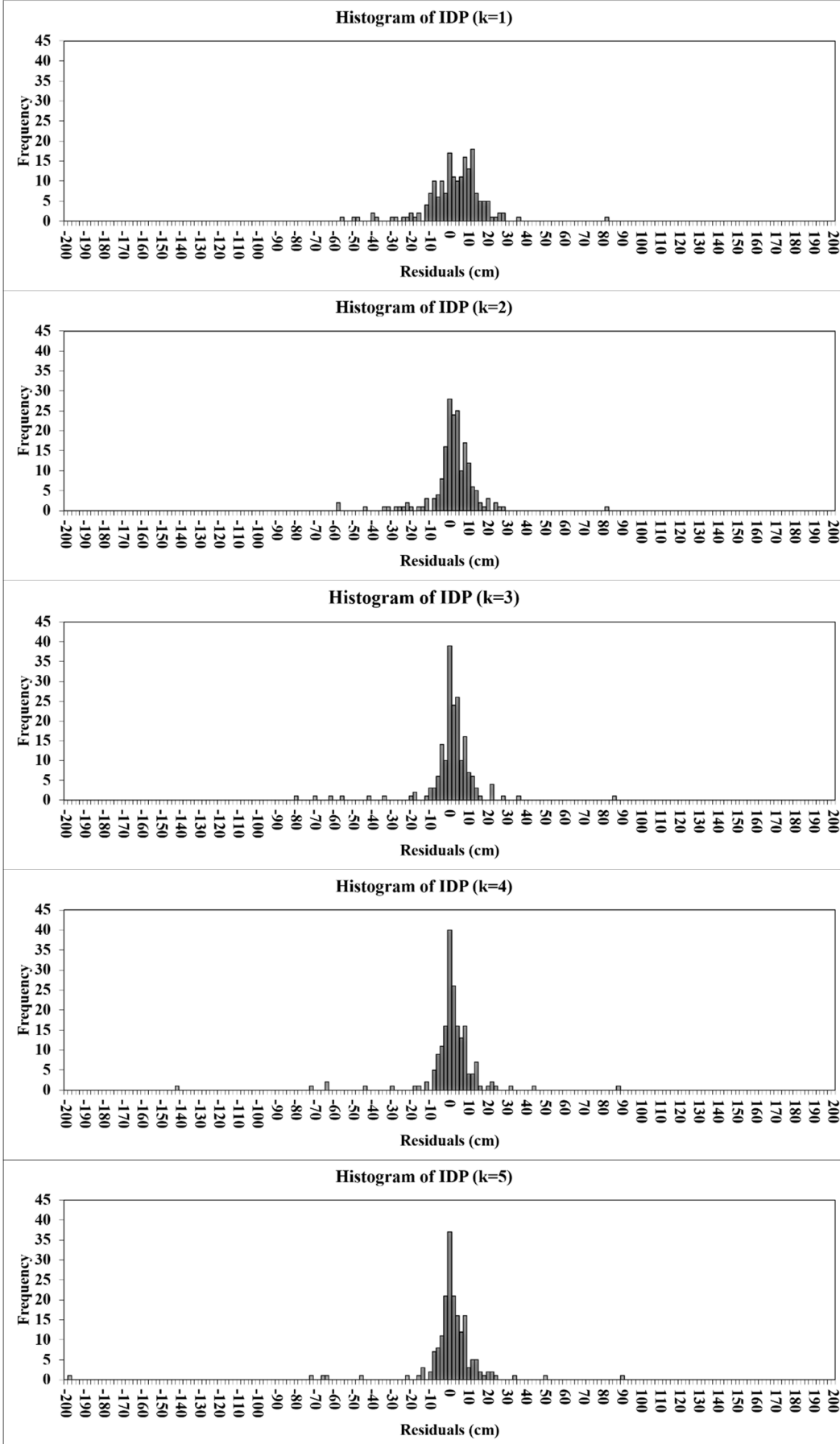
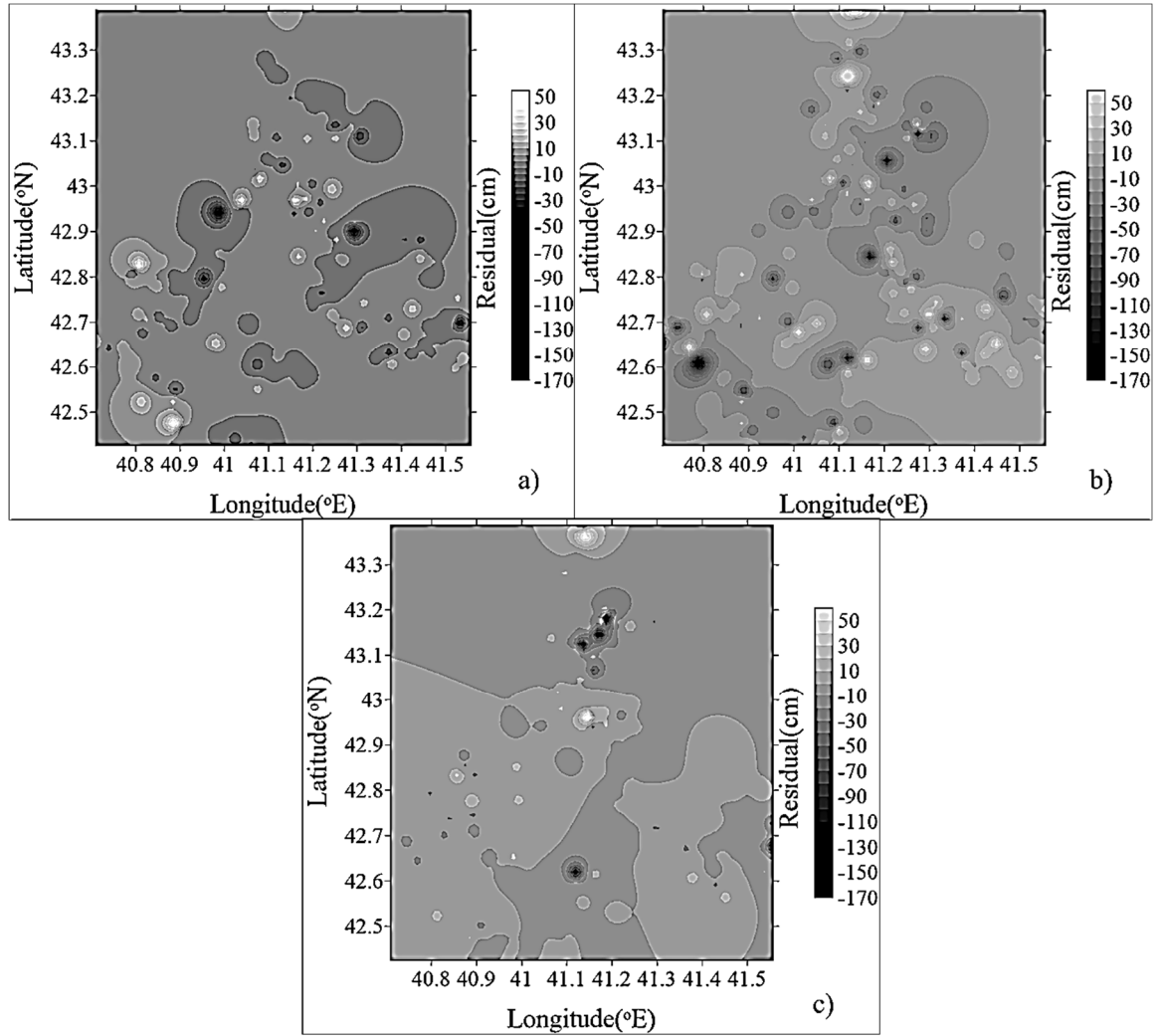


Fig. 6 Residuals versus number of samples in IDP models ( $k=1, 2, 3, 4$ , and  $5$ ).



**Fig. 7** Residuals between the observed and the predicted values a) GPR based on Materniso ( $d=1$ ), b) RBFNN, c) GRNN.

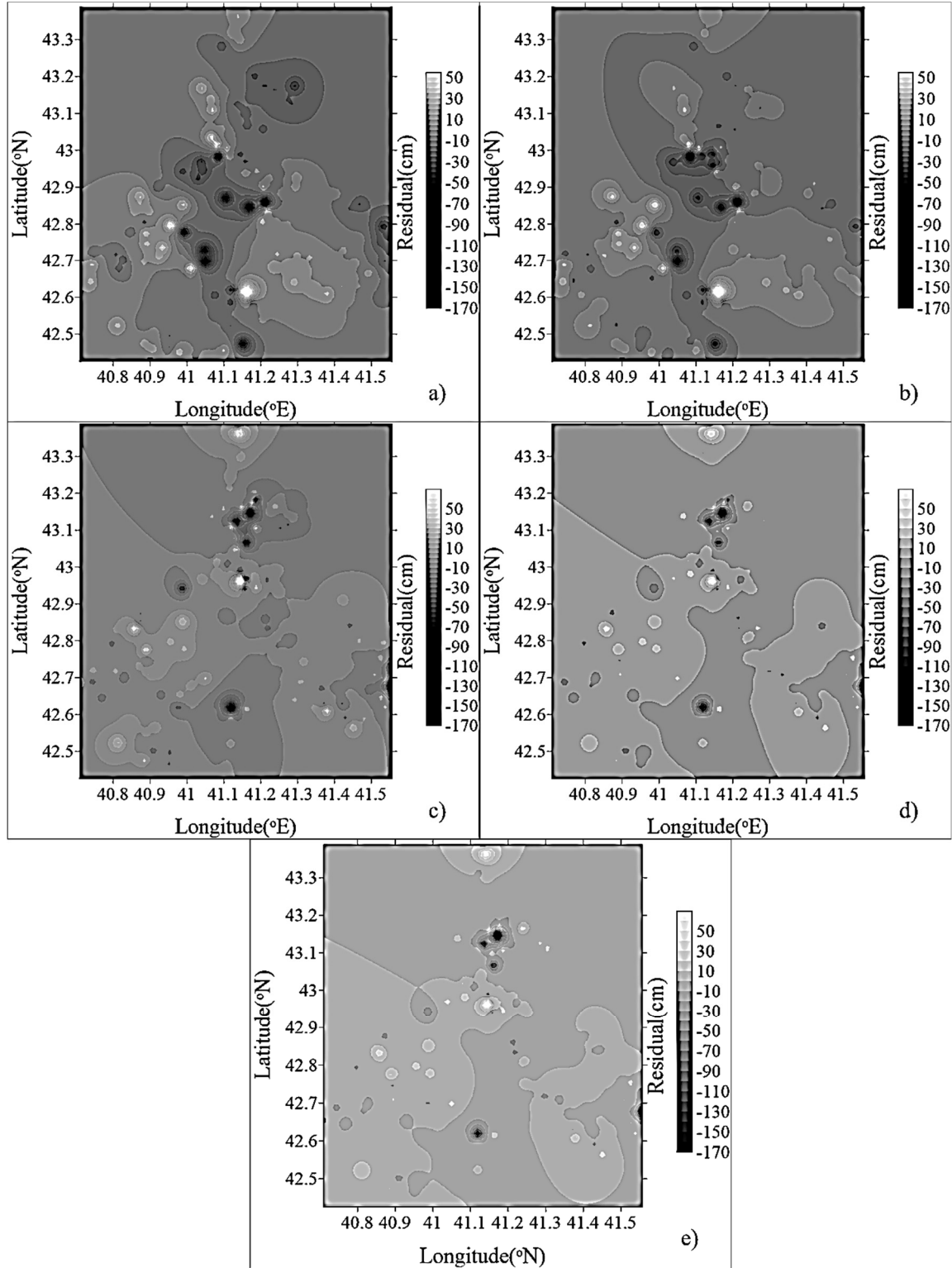
Table 2 shows that the GPR based on the Materniso ( $d = 1$ ) model had the best results (i.e.,  $RMSE = 8.32$  cm,  $MAE = 5.51$  cm,  $R^2 = 0.98968$ ) and thus was better than the other models, with the IDP ( $k = 2$ ) ranking second-best. On the other hand, the RBFNN model yielded the worst results (i.e.,  $RMSE = 28.02$  cm,  $MAE = 19.18$  cm, and,  $R^2 = 0.89095$ ). When the ANN models were compared with each other, the GRNN predicted better than the RBFNN. A comparison of the IDP models with each other demonstrated that the IDP ( $k = 2$ ) model achieved the highest  $RMSE$  (13.12 cm) and  $R^2$  (0.97605) values, whereas the IDP ( $k = 2$ ), IDP ( $k = 3$ ) and IDP ( $k = 4$ ) models gave very similar results in terms of MAE.

Figure 4 illustrates the comparisons of the predicted results of all the different models with the actual values of geoid undulation. It can be seen immediately from this figure that the IDP produced better results compared to the RBFNN and GRNN. The selection of the weighting power ( $k$ ) parameter significantly influenced the results of the IDP. The IDP ( $k = 2$ ) model produced results that were closer to the true geoid undulation values than the other IDP

models. In general, the best straight line in the scatter plots showed that the slope and intercept coefficients for the GPR based on the Materniso ( $d = 1$ ) model were, respectively, closer to 1 and 0 than those of the other models.

The residual values specifying the prediction limitations of the developed models were calculated by subtracting the predicted geoid undulation values from their corresponding observed geoid undulation values. Histogram graphics were drawn of the GPR based on the Materniso ( $d = 1$ ) covariance function, the RBFNN, GRNN, and IDP ( $k = 1, 2, 3, 4$ , and 5). Figure 5 shows the histogram of the frequency against the residual values for GPR based on the Materniso ( $d = 1$ ) covariance function, the RBFNN, and GRNN; Figure 6 indicates the distribution of geoid undulations for the IDP models ( $k=1, 2, 3, 4, 5$ ).

Looking at the histograms, these figures confirm that the GPR based on the Materniso ( $d = 1$ ) covariance function model had a higher predictive capacity than the other models. The RBFNN model performed the worst since the residual values were much greater than in the other models.



**Fig. 8** Residuals between the observed and the predicted values a) IDP ( $k=1$ ), b) IDP ( $k=2$ ), c) IDP ( $k=3$ ), d) IDP ( $k=4$ ), e) IDP ( $k=5$ ).

Figures 7 and 8 also show maps of residuals between the observed and the predicted values by all models.

Compared the maps of residuals given in Figures 7 and 8, it is clearly seen that, for GPR based on the Materniso ( $d = 1$ ) covariance function model, the residuals are much smaller over the whole study area (Fig. 7a). This conclusion confirms once again

the superior performance of the GPR based on the Materniso ( $d = 1$ ) covariance function model in predicting geoid undulation among the all tested models.

## 5. CONCLUSION

In this study, the GPR was successfully applied to predict geoid undulation. The zero mean function

was simultaneously modeled with ten different covariance functions: the Matérn covariance function with automatic relevance detection (Maternard  $d = 1, 3, 5$ ), the Matérn covariance function with isotropic distance measure (Materniso  $d = 1, 3, 5$ ), the rational quadratic (RQ) covariance function with automatic relevance detection, the RQ function with isotropic distance measure, the squared exponential (SE) covariance function with automatic relevance detection, and the SE covariance function with isotropic distance measure. In this way, ten GPR models were developed to predict geoid undulation. Among these models, the GPR based on the Materniso ( $d = 1$ ) covariance function model gave the best results, based on the lowest RMSE and MAE and the highest  $R^2$  values. The prediction results of the GPR models constructed with the Matérn covariance functions were better than the results obtained with GPR models using other covariance functions (i.e., RQ and SE). The performance of the best GPR model was then compared with the ANN models (RBFNN and GRNN), and the interpolation method of inverse distance to a power (IDP). By assessing the results of the models, it was concluded that the GPR based on the Materniso ( $d = 1$ ) model was capable of providing more accurate geoid undulation predictions in comparison to the RBFNN, GRNN, and IDP. The RBFNN model provided the worst geoid undulation estimates. The GPR model may be tested using different data. In the present work, the zero mean function was used and therefore, this had no effect on the prediction output. Thus, the prediction depended solely on the covariance function. If other means functions (constant mean and linear mean) had been used, the GPR models might have produced even better results. This may be the subject of another study. Improper division of data may cause predicted results to deviate from observed values. This leads to reduced accuracy of results for the predictive performance of models. To overcome this problem, the cross-validation technique for evaluating predictive models (e.g., k-fold) can be used. This issue will be addressed in a future work.

## ACKNOWLEDGMENTS

We are grateful to the Erzincan XXIV Regional Directorate of Land Registry and Cadastre Directorate in Turkey for providing the data-sets for our study. The GPML (Gaussian Processes for Machine Learning) open-source toolbox was employed to perform the analyses of the GPR. We would also like to thank the anonymous reviewers for their constructive comments and suggestions that helped to improve the paper.

## REFERENCES

Ackermann, E.R., De Villiers, J.P. and Cilliers, P.J.: 2011, Nonlinear dynamic systems modeling using Gaussian processes: Predicting ionospheric total electron content over South Africa. *J. Geophys. Res., Space Phys.*, 116, No. A10. DOI: 10.1029/2010JA016375

Akcin, H. and Celik, C.T.: 2013, Performance of artificial neural networks on Kriging method in modeling local geoid. *Bol. Cienc. Geod.*, 19, No. 1, 84–97. DOI: 10.1590/S1982-21702013000100006

Akyilmaz, O., Özlüdemir, M.T., Ayan, T. and Çelik, R.N.: 2009, Soft computing methods for geoidal height transformation. *Earth Planets Space*, 61, No. 7, 825–833. DOI: 10.1186/BF03353193

Albayrak, M., Özlüdemir, M.T., Aref, M.M. and Halicioğlu, K.: 2020, Determination of Istanbul geoid using GNSS/levelling and valley cross levelling data. *Geod. Geodyn.*, 11, No. 3, 163–173. DOI: 10.1016/j.geog.2020.01.003

Albert, J.G., Oei, M.S.S.L., van Weeren, R.J., Intema, H.T. and Röttgering, H.J.A.: 2020, A probabilistic approach to direction-dependent ionospheric calibration. *Astron. Astrophys.*, 633. DOI: 10.1051/0004-6361/201935668

Bishop, C.: 1995, *Neural networks for pattern recognition*, Oxford: Clarendon Press.

Cakir, L. and Yilmaz, N.: 2014, Polynomials, radial basis functions and multilayer perceptron neural network methods in local geoid determination with GPS/levelling. *Measurement*, 57, 148–153. DOI: 10.1016/j.measurement.2014.08.003

Chen, N., Qian, Z., Nabney, I.T. and Meng, X.: 2013, Wind power forecasts using Gaussian processes and numerical weather prediction. *IEEE Trans. Power Syst.*, 29, No. 2, 656–665. DOI: 10.1109/TPWRS.2013.2282366

Doganalp, S. and Selvi, H.Z.: 2015, Local geoid determination in strip area projects by using polynomials, least-squares collocation and radial basis functions. *Measurement*, 73, 429–438. DOI: 10.1016/j.measurement.2015.05.030

Doganalp, S.: 2016, Geoid height computation in strip-area project by using least-squares collocation. *Acta Geodyn. Geomater.*, 13, No. 2(182), 167–176. DOI: 10.13168/AGG.2015.0054

Erol, B. and Erol, S.: 2013, Learning-based computing techniques in geoid modeling for precise height transformation. *Comput. Geosci.*, 52, 95–107. DOI: 10.1016/j.cageo.2012.09.010

Erol, B., Erol, S. and Çelik, R.N.: 2008, Height transformation using regional geoids and GPS/levelling in Turkey. *Surv. Rev.*, 40, No. 307, 2–18. DOI: 10.1179/003962608X253394

Franke, R. and Nielson, G.: 1980, Smooth interpolation of large sets of scattered data. *Int. J. Numer. Meth. Eng.*, 15, No. 11, 1691–1704. DOI: 10.1002/nme.1620151110

Haidar, A.M., Mustafa, M.W., Ibrahim, F.A. and Ahmed, I.A.: 2011, Transient stability evaluation of electrical power system using generalized regression neural networks. *Appl. Soft Comput.*, 11, No. 4, 3558–3570. DOI: 10.1016/j.asoc.2011.01.028

Haykin, S.: 1999, *Neural networks: A comprehensive foundation*. Prentice-Hall, Upper Saddle River, New Jersey.

Heiskanen W.A. and Moritz H.: 1967, *Physical geodesy*. W.H. Freeman and Co., San Fransisco.

Inyurt, S., Kashani, M.H. and Sekertekin, A.: 2020, Ionospheric TEC forecasting using Gaussian process regression (GPR) and multiple linear regression (MLR) in Turkey. *Astrophys. Space Sci.*, 365, No. 6, 1–17. DOI: 10.1007/s10509-020-03817-2

Kaloop, M.R., Rabah, M., Hu, J.W. and Zaki, A.: 2018, Using advanced soft computing techniques for



- regional shoreline geoid model estimation and evaluation. *Mar. Georesour. Geotech.*, 36, No. 6, 688–697. DOI: 10.1080/1064119X.2017.1370622
- Kao, S.P., Chen, C.N., Huang, H.C. and Shen, Y.T.: 2014, Using a least squares support vector machine to estimate a local geometric geoid model. *Bol. Cienc. Geod.*, 20, No. 2, 427–443. DOI: 10.1590/S1982-21702014000200025
- Karaaslan, Ö., Kayıkçı, E.T. and Aşık, Y.: 2016, Comparison of local geoid height surfaces, in the province of Trabzon. *Arab. J. Geosci.*, 9, No. 6, 431. DOI: 10.1007/s12517-016-2470-2
- Kavzoglu, T. and Saka, M.H.: 2005, Modelling local GPS/levelling geoid undulations using artificial neural networks. *J. Geod.*, 78, No. 9, 520–527. DOI: 10.1007/s00190-004-0420-3
- Kisi, O., Tombul, M. and Kermani, M.Z.: 2015, Modeling soil temperatures at different depths by using three different neural computing techniques. *Theor. Appl. Climatol.*, 121, No. 1-2, 377–387. DOI: 10.1007/s00704-014-1232-x
- Kisi, O.: 2008, The potential of different ANN techniques in evapotranspiration modelling. *Hydrol. Process.*, 22, No. 14, 2449–2460. DOI: 10.1002/hyp.6837
- Kiş, Ö.: 2009, Daily pan evaporation modelling using multi-layer perceptrons and radial basis neural networks. *Hydrol. Process.*, 23, No. 2, 213–223. DOI: 10.1002/hyp.7126
- Kotsakis, C. and Sideris, M.G.: 1999, On the adjustment of combined GPS/levelling/geoid networks. *J. Geod.*, 73, No. 8, 412–421. DOI: 10.1007/s001900050261
- Ladlani, L., Houichi, L., Djemili, L., Heddam, S. and Belouz, K.: 2012, Modeling daily reference evapotranspiration (ET<sub>0</sub>) in the north of Algeria using generalized regression neural networks (GRNN) and radial basis function neural networks (RBFNN): a comparative study. *Meteorol. Atmos. Phys.*, 118, No. 3–4, 163–178. DOI: 10.1007/s00703-012-0205-9
- Lin, C., Li, T., Chen, S., Liu, X., Lin, C. and Liang, S.: 2019, Gaussian process regression-based forecasting model of dam deformation. *Neural. Comput. Appl.*, 31, No. 12, 8503–8518. DOI: 10.1007/s00521-019-04375-7
- 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, No. 2, 81–89. DOI: 10.1061/(ASCE)0733-9453(2007)133:2(81)
- Rasmussen, C.E., and Nickisch, H.: 2010, Gaussian processes for machine learning (gpml) toolbox. *J. Mach. Learn. Res.*, 11, 3011–3015.
- Specht, D.F.: 1991, A general regression neural network. *IEEE Trans. Neural Netw.*, 2, No. 6, 568–576. DOI: 10.1109/72.97934
- Specht, D.F.: 1993, The general regression neural network-rediscovered. *Neural Netw.*, 6, No. 7, 1033–1034. DOI: 10.1016/S0893-6080(09)80013-0
- Stopar, B., Ambrožič, T., Kuhar, M. and Turk, G.: 2006, GPS-derived geoid using artificial neural network and least squares collocation. *Surv. Rev.*, 38, No. 300, 513–524. DOI: 10.1179/sre.2006.38.300.513
- Trojanowicz, M.: 2012, Local modelling of quasigeoid heights with the use of the gravity inverse method—case study for the area of Poland. *Acta Geodyn. Geomater.*, 9, No. 1(165), 5–18.
- Trojanowicz, M.: 2015, Assessment of the accuracy of local quasigeoid modelling using the GGI method: case study for the area of Poland. *Stud. Geophys. Geod.*, 59, No. 4, 505–523. DOI: 10.1007/s11200-014-0527-9
- Trojanowicz, M., Osada, E. and Karsznia, K.: 2020, Precise local quasigeoid modelling using GNSS/levelling height anomalies and gravity data. *Surv. Rev.*, 52, No. 370, 76–83. DOI: 10.1080/00396265.2018.1525981
- Tscherning, C.C., Radwan, A., Tealeb, A.A., Mahmoud, S.M., Abd El-Monum, M., Hassan, R., El-Syaed, I. and Saker, K.: 2001, Local geoid determination combining gravity disturbances and GPS/levelling: a case study in the Lake Nasser area, Aswan, Egypt. *J. Geod.*, 75, No. 7–8, 343–348. DOI: 10.1007/s001900100185
- Tusat, E. and Mikailsoy, F.: 2018, An investigation of the criteria used to select the polynomial models employed in local GNSS/leveling geoid determination studies. *Arab. J. Geosci.*, 11, No. 24, 801. DOI: 10.1007/s12517-018-4176-0
- Venkatesan, P. and Anitha, S.: 2006, Application of a radial basis function neural network for diagnosis of diabetes mellitus. *Curr. Sci.*, 91, No. 9, 1195–1199.
- Veronez, M.R., Florêncio de Souza, S., Matsuoka, M.T., Reinhardt, A. and Macedônio da Silva, R.: 2011, Regional mapping of the geoid using GNSS (GPS) measurements and an artificial neural network. *Remote Sens.*, 3, No. 4, 668–683. DOI: 10.3390/rs3040668
- Williams, C.K. and Rasmussen, C.E.: 2006, *Gaussian processes for machine learning*, MA: MIT Press, Cambridge.
- Yanalak, M. and Baykal, O.: 2001, Transformation of ellipsoid heights to local leveling heights. *J. Surv. Eng.*, 127, No. 3, 90–103. DOI: 10.1061/(ASCE)0733-9453(2001)127:3(90)
- Yanalak, M. and Baykal, O.: 2003, Digital elevation model based volume calculations using topographical data. *J. Surv. Eng.*, 129, No. 2, 56–64. DOI: 10.1061/(ASCE)0733-9453(2003)129:2(56)
- Yanalak, M.: 2004, Sibson (natural neighbour) and non-Sibsonian interpolation for digital elevation model (DEM). *Surv. Rev.*, 37, No. 291, 360–376. DOI: 10.1179/sre.2004.37.291.360
- Yilmaz, M.: 2010, Adaptive network based on fuzzy inference system estimates of geoid heights interpolation. *Sci. Res. Essays*, 5, No. 16, 2148–2154. DOI: 10.5897/SRE.9000130
- Yuan, J., Wang, K., Yu, T. and Fang, M.: 2008, Reliable multi-objective optimization of high-speed WEDM process based on Gaussian process regression. *Int. J. Mach. Tools Manuf.*, 48, No. 1, 47–60. DOI: 10.1016/j.ijmachtools.2007.07.011
- Yuan, Q., Li, S., Yue, L., Li, T., Shen, H. and Zhang, L.: 2019, Monitoring the variation of vegetation water content with machine learning methods: Point–surface fusion of MODIS products and GNSS-IR observations. *Remote Sens.*, 11, No. 12, 1440. DOI: 10.3390/rs11121440
- Yılmaz, M. and Arslan, E.: 2008, Effect of the type of membership function on geoid height modelling with fuzzy logic. *Surv. Rev.*, 40, No. 310, 379–391. DOI: 10.1179/003962608X325439
- Zaletnyik, P., Völgyesi, L. and Paláncz, B.: 2008, Modelling local GPS/levelling geoid undulations using support vector machines. *Period. Polytech. Civ. Eng.*, 52, No. 1, 39–43. DOI: 10.3311/pp.ci.2008-1.06
- Zounemat-Kermani, M.: 2014, Principal component analysis (PCA) for estimating chlorophyll concentration using forward and generalized regression neural networks. *Appl. Artif. Intell.*, 28, No. 1, 16–29. DOI: 10.1080/08839514.2014.862771