ORIGINAL PAPER

Enhancing survey field data with artificial intelligence: a real-time kinematic GPS study

Daniel Asenso-Gyambibi[1] · Joseph Agyei Danquah[1] · Edwin Kojo Larbi[1] · Michael Stanley Peprah[2] · Naa Lamkai Quaye-Ballard[1]

Abstract

The significance of adjustment and computation studies has grown in recent years, influencing allied fields like arithmetic and satellite geodesy. This empirical study explores the effectiveness of various soft and traditional regression methods in correcting survey field data. Specifically, it investigates soft computing techniques such as back-propagation artificial neural network (BPANN), radial basis function artificial neural network (RBFANN), generalized regression artificial neural network (GRANN), and traditional regression methods like polynomial regression model (PRM) and least square regression (LSR) techniques. The study aims to fill the knowledge gap regarding soft computing strategies for modifying real-time kinematics (RTK) GPS field data and the ongoing debate between artificial intelligence techniques (ANN) and traditional methods on which technique offers the best results in modifying survey field data. Performance criteria, including horizontal displacement (HE), arithmetic mean error (AME), arithmetic mean square error (AMSE), minimum and maximum error values, and arithmetic standard deviation (ASD), were used to assess each model technique. Statistical analysis revealed that RBFANN, BPANN, and GRANN achieved superior accuracy compared to conventional techniques (PRM and LSR) in adjusting real-time kinematics GPS data. RBFANN outperformed BPANN and GRANN in terms of AME, AMSE, and ASD of their horizontal displacement. These findings suggest that soft computing techniques enhance real-time kinematics GPS field data adjustment, addressing critical issues in accurate positioning, particularly in Ghana. This study contributes to the knowledge base for developing an accurate geodetic datum in Ghana for national and local objectives. This will lay a foundation for the global determination of exact positions in Ghana. RBFANN emerges as a promising option for real-time kinematics GPS field data adjustment in topographic surveys. However, care should be taken to check issues of data overfitting.

Keywords Back-propagation neural network · Generalized regression neural network · Real-time kinematics Global Positioning System (RTK GPS) · Horizontal displacement · Radial basis functions

Introduction

In recent years, the fields of satellite and mathematical geodesy have greatly benefited from advancements in the adjustment and computation of survey field data. These techniques have been instrumental in evaluating error magnitudes and establishing tolerance thresholds (Yakubu et al.

2018). Furthermore, as the collection of survey field data often involves redundant observations, it becomes imperative to modify these data to ensure consistency (Ghilani and Wolf 2006). The pursuit of accurate survey data adjustment remains a focal point for geodesists, geophysicists, surveyors, topographers, and other scholars (Asenso-Gyambibi et al. 2022). GPS observations, as a vital component of surveying, are not excluded from errors. Factors such as satellite orbital errors, signal transmission timing issues due to atmospheric conditions, receiver errors, multipath errors due to reflection from nearby buildings or other surfaces, miscentering errors of the reception antenna over the ground station, and receiver height-measuring errors contribute to inaccuracies (Ghilani and Wolf 2006). These signal errors affect the accuracy of point positions. To mitigate these

✉ Edwin Kojo Larbi
  edlarbi90@gmail.com

1   Geo-Informatics Division, Building and Road Research Institute (CSIR-BRRI), Kumasi, Ghana

2   IHTMOC Consulting Company Ltd, Kumasi, Ghana

errors and enhance point location precision, duplicate observations are made, necessitating the adjustment of field data using both survey methodologies (Ghilani and Wolf 2006).

The field of geodesy demands rigorous adjustment and computation methods. Over the years, various techniques have been employed, including the Kalman filter (KF) (Bashirov et al. 2008; Wang 2014), least squares collocation (LSC) (Ruffhead 2012), total least squares (TLS) (Schaffrin 2006), and kriging methods (Erol and Celik 2005; Kleijnen 2009). The ordinary least square (OLS) method has conventionally served as the go-to technique for adjusting surveying networks (Okwuashi and Eyoh 2012). OLS treats observation equations as stochastic to minimize the square of the sum of residuals, addressing errors solely in the observation matrix. Many geoscientific researchers have successfully applied OLS to tackle various scientific challenges (Annan et al. 2016). To improve OLS efficiency, the total least squares (TLS) method was developed (Annan et al. 2016). However, it is worth noting that the SVD (singular value decomposition) used in TLS does not preserve the structure of the extended data matrix in typical scenarios, potentially affecting its statistical idealness (Golub and Van Loan 1989; Peprah and Mensah 2017). Additionally, least squares collocation (LSC) requires changes in the cross-variance function to maintain its modeling effectiveness (Ophaug and Gerlach 2017). However, these classical methods cannot meet the current precision requirements and require more advanced techniques (Lee et al. 2020).

In parallel to these developments, the field of artificial intelligence (AI) has made significant strides. The ability to simulate human functions, traits, and features has been a hallmark of AI success over other traditional methods (Maidaniuk et al. 2022). Akcin and Celik (2013) conducted a comparative test, pitting local geoid heights derived from trained artificial neural network (ANN) models against heights predicted using Kriging methods in the Bursa Metropolitan Area, Turkey. The results favored the trained ANN models, highlighting their potential in geodesy and survey data adjustment. ANN, particularly back-propagation artificial neural network (BPANN), generalized regression artificial neural network (GRANN), and radial basis functions artificial neural network (RBFANN), has gained traction in various mathematical and satellite geodesy applications (Ziggah et al. 2017). These applications encompass adjusting differential Global Positioning System (DGPS) data, coordinate transformation, modeling DGPS data uncertainties, tidal height predictions, road damage detection after earthquakes, orthometric height predictions in mining, air overpressure predictions, soil nutrient forecasts, rainfall predictions, and more. The study aims to fill the knowledge gap in showcasing the most accurate techniques between artificial intelligence techniques (ANN) and classical methods in adjusting real-time kinematics (RTK) GPS field data.

Despite the advancements, it is evident from the existing literature that ANN approaches have seen limited use in Ghana, and their suitability for modifying RTK GPS data remains underexplored due to the lack of technical know how and poor broad band connectivity. A comprehensive comparison of classical regression techniques and AI model training methods for rectifying RTK GPS data in Ghana is yet to be undertaken. This study presents ANN approach as an effective technique for enhancing RTK GPS survey field data. The study aims to investigate the application of BPANN, GRANN, and RBFANN as alternative technologies to existing classical approaches. It evaluates the effectiveness of these strategies in modifying survey field data, employing various performance criteria indices, including horizontal displacement (HD), arithmetic mean error (AME), arithmetic mean square error (AMSE), minimum and maximum residual values, and arithmetic standard deviation (ASD). These indices are used to assess the strength of each model in adjustment by measuring how the predictions are close to the actual values. Thus, the study offers Ghanaian researchers and practitioners a valuable opportunity to assess the potential of soft computing techniques in addressing some of the country's geodetic challenges and advanced models for survey data adjustment. Beyond its impact on Ghanaian geodesists, this study contributes substantively to the scholarly discourse surrounding the application of soft computing techniques in addressing challenges within the surveying and related fields. By advancing the understanding of these techniques and their efficacy, it paves the way for more precise and efficient surveying practices in a rapidly evolving technological landscape.

Study area

The Greater Kumasi Metropolitan Area (GKMA), which includes the inner Kumasi and other nearby municipalities and districts such as Asokore Mampong, Ejisu, Juaben, Kwabre East, Afigya Kwabre, Atwima Kwanwoma, and Atwima Nwabiagya Districts, is located in the Ashanti Region of Ghana. Its latitudes are 6° 35′N and 6° 40′S and its longitudes are 1° 30′W and 1° 35′E. It has an elevation of between 250 and 350 m above sea level (Ghana Statistical Service 2014). A significant river (Owabi) and streams including Subin, Wiwi, Sisai, Aboabo, and Nsuben cross the undulating ground (Ghana Statistical Service 2014; Caleb et al. 2018). Hence, the intervisibility of various observation points will be affected in geodetic surveys due to the undulating nature of the terrain and the network of streams. With a total population of 3,190,473, it has a land area of 2603 km^2 (Oduro et al. 2014). The area experiences a rapid population growth with its associated rise in built up areas, resulting in increasing multipath errors

over the years. The War Office 1926 ellipsoid serves as the research area's horizontal geodetic datum, and mean sea level (MSL), which approximates the geoid, serves as the study area's vertical datum (Peprah and Kumi 2017). Ghana projected grid, derived from the Transverse Mercator with 1° W Central Meridian and the World Geodetic System 1984 (WGS84) (UTM Zone 30N), is the type of coordinate system utilized in the research region (Yakubu et al. 2018). The city of Metropolis is classified as wet subequatorial. The average maximum temperature is around 30.7 °C, while the average lowest temperature is about 21.5 °C. At sunrise and sunset, the average humidity is about 84.16% and 60%, respectively (Caleb et al. 2018) A clear or an unobstructed view of the sky should be considered for best results in RTK GPS surveys. There is a double maximum rainfall regime in the research area, with roughly 214.3 mm in June and 165.2 mm in September (Ghana Statistical Service 2014). The area experiences a considerable amount of rain which tends to scatter or absorb GPS signals, reducing the positioning accuracy due to bending of signals by water droplets. The Metropolis

is located in the moist semi-deciduous south-East Ecological Zone, which is a transitional forest zone (Ghana Statistical Service 2014). The forest canopy can introduce errors such as receiver and multipath errors in GPS survey; hence, duplicate observations are made, necessitating the quest to find the most accurate technique for adjustment of GPS field data. Middle Precambrian rock predominates in the study area, which also contains two major lithostratigraphic and lithotectonic complexes, namely the Paleoproterozoic supracrustal and intrusive rocks and the Neoproterozoic to early Cambrian lithologically diverse platform sediments. The study area's distinctive geological structure has contributed to the growth of the construction industry in the Metropolis, with few small-scale mining operations and the proliferation of stone quarrying and mining (Osei-Nuamah and Appiah-Adjei 2017). Geodetic data is employed by the construction and mining industries in their activities such as building and road set outs, ore markouts, topo pickups, and demarcation of mine concession. A diagram of the placement of control sites within the research region is shown in Fig. 1.
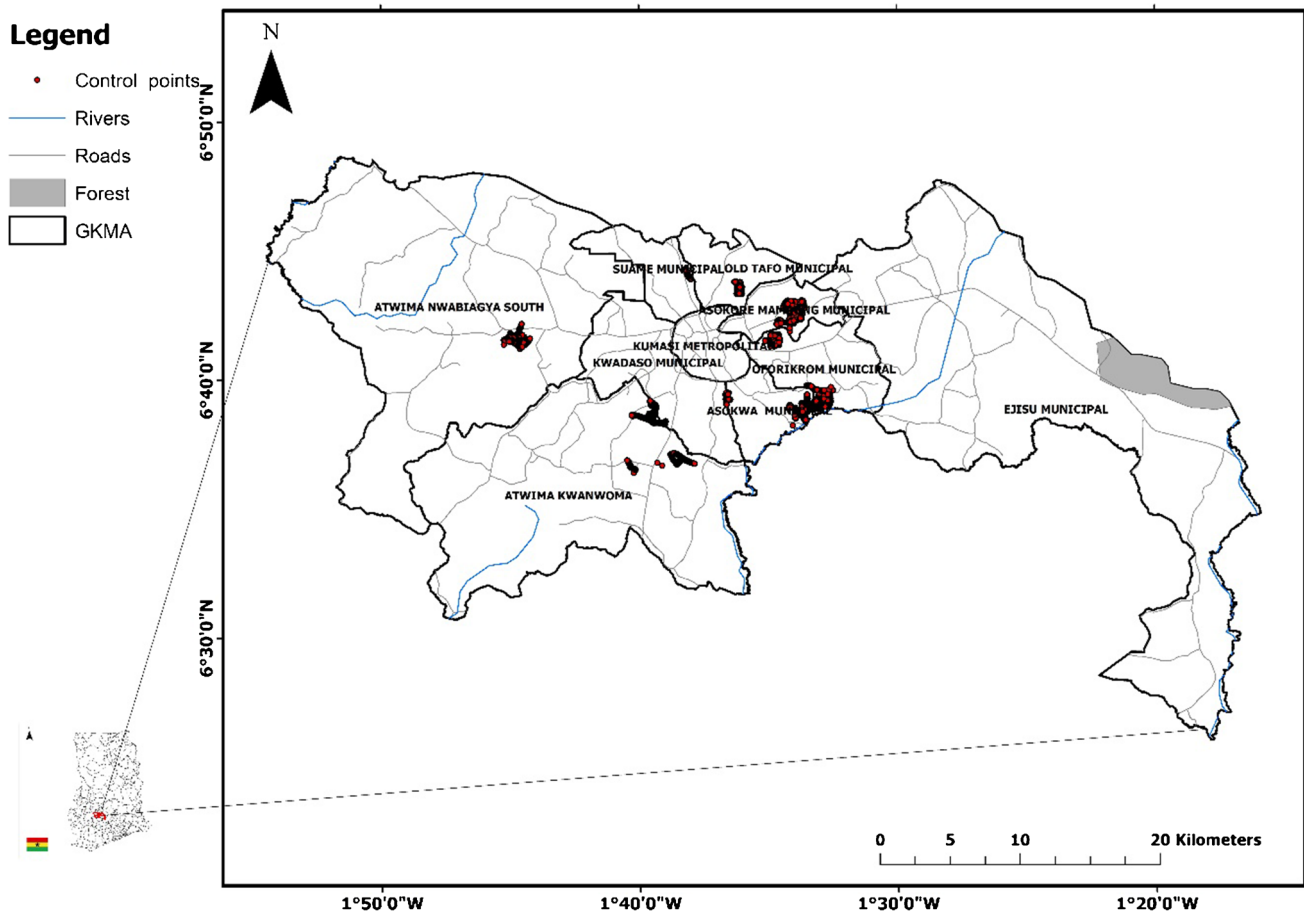

Fig. 1 Map of the study area

Resources and methods used

Resources used

Topographic data from a survey conducted by the authors for the Ghana urban water supply project in the Greater Kumasi Metropolitan Area (GKMA) served as the study's primary source of data. A total of 1000 control points were acquired from September 5, 2022 to October 15, 2022 using real-time kinematics (RTK). GPS equipment make up the sample data. This was to provide enough data to better assess the model performance, identify patterns, trends, and any anomalies. For the chosen controls of the study area, the data consists of three-dimensional coordinates, namely eastings, northings, and ellipsoidal heights designated as (E, N, h), which were recorded using the RTK GPS device. These data forms the actual observation points; hence, the best model for adjustment should be able to predict these values accurately. Table 1 shows sample of the dataset collected from the field.

Methods used

Radial basis function artificial neural network (RBFANN)

An unsupervised learning algorithm built using functional approximation is the RBFANN model. It is made up of an input layer, a hidden layer, and an output layer, each of which serves a different purpose. Sensory elements in the input layer link the network to its surroundings. A non-linear transition from the input space to the hidden space

Table 1   Sample of data used for the study (unit in meters)

ID	Easting	Northing	Elevations (h)
Pt1	658,143.7	741,095.1	311.380
Pt2	658,340.9	741,791.2	308.365
Pt3	658,340.8	741,792.4	308.387
Pt4	658,339.8	741,792.3	308.382
Pt5	658,339.9	741,791.2	308.386
Pt6	658,342.0	741,790.2	308.516
Pt7	658,342.8	741,789.6	308.505
Pt8	658,343.1	741,787.1	308.585
Pt9	658,340.8	741,789.3	308.474
Pt10	658,339.1	741,788.8	308.32
Pt11	658,338.4	741,788.6	308.302
Pt12	658,338.7	741,788.4	307.710
Pt13	658,338.2	741,788.4	308.341
Pt14	658,345.1	741,772.4	309.595
Pt15	658,346.5	741,761.9	310.349
Pt16	658,344.6	741,761.7	309.961
Pt17	658,342.3	741,761.3	309.439

was applied in the second layer by the network's sole hidden layer. The network's response to the activation pattern given to the output layer is provided by the linear output layer. In this study, the input and output variables were the 2D horizontal coordinates denoted as $(N_{i,j}, E_{i,j})$, $(E_{i,j})$, and $(N_{i,j})$ respectively. These variables were chosen in the model formulation because they were directly obtained from the field. The dataset used to create the model was split between training and testing data, with training data making up 70% of the overall dataset. A linear function is utilized in the input neurons because RBFANN is an accurate interpolator (Erdogan 2009), and the connections between the input and hidden layers are not weighted (Kaloop et al. 2017). This study applies the Gaussian function, which provides an optimal approximation and the weighted hidden output layer from Eq. (1) is added to the output neuron.

$$y(x) = \sum_{j=1}^{n} K_j X_j(x) \tag{1}$$

where n is the number of hidden neurons, $x \in R^M$ is the input, K_j is the output layer weights of the radial basis function network, and $\chi_j(x)$ is Gaussian radial basis function given by Eq. (2) (Idri et al. 2010; Srichandan 2012):

$$\chi_j(x) = e\left(\frac{-\parallel x_i - c_j \parallel^2}{\sigma_j^2}\right) \tag{2}$$

where $c_j \in R^M$ and σ are the center and width of jth hidden neurons, respectively, and $\parallel \parallel$ denotes the Euclidean distance.

Back-propagation artificial neural network (BPANN)

BPANN is an efficient and popular multilayer perceptron (MLP) model due to its straightforward implementation (Yilmaz et al. 2017; Ziggah et al. 2016). An input layer with M inputs, a hidden layer with q units, and an output layer with n outputs make up a BPANN. In this investigation, the two-dimensional (2D) horizontal coordinates served as the M inputs $(N_{i,j}, E_{i,j})$, while the q units were achieved by a trial-and-error training by varying the number of hidden neurons, and the n outputs were the estimated (outputs) achieved $(N_{i,j})$ and $(E_{i,j})$ by the BPANN model, respectively. The output of the model (y_i) with a single output neuron is represented by Eq. (3) (Mihalache 2012):

$$y(i) = f\left(\sum_{j=1}^{q} W_j f\left(\sum_{i=1}^{M} w_{j,i} x_i\right)\right) \tag{3}$$

where W_j is the weight between the hidden layer and the output layer, w_{jI} is the weight between the input layer and the hidden layer, and x_i is the input parameter. In this study, the selected input and output variables were normalized into

the interval $[-1, 1]$ using Eq. (4) (Mueller and Hemond 2013) given as

$$Z(i) = \frac{y_{\min} + (y_{\max} - y_{\min}) \times (x_i - x_{\min})}{(x_{\max} - x_{\min})} \quad (4)$$

where $Z(i)$ represents the normalized data, x_i is the measured coordinate values, while x_{min} and x_{max} represent the minimum and maximum values of the measured coordinates with y_{max} and y_{min} values set at 1 and -1, respectively. Based on the lowest arithmetic mean error (AME), arithmetic mean square error ($AMSE$), minimum residual error(r_{min}), maximum residual error(r_{max}), and arithmetic standard deviation (ASD), the best model was found. Their mathematical formulation is described in the section on model performance assessment. In keeping with the approach in the literature, the current investigation used one hidden layer in the BPANN (Hornik et al. 1989). Moreover, the hyperbolic tangent activation function was chosen for the hidden units to incorporate non-linearity into the network, whereas a linear function was used for the output units. Equation (5) gives the definition of the hyperbolic tangent function (Yonaba et al. 2010):

$$Z(x) = \tanh(x) = \frac{2}{1 + e^{-2x}} - 1 \quad (5)$$

where x is the total weighted inputs. The systematic methodology used in this work, which included ANN and traditional regression techniques, is depicted in a flowchart in Fig. 2. The dataset was trained independently using the Levenberg–Marquardt algorithm (trainlm) and the Bayesian regularization training technique (trainbr). Sequential trial-and-error approach was adopted to achieve the optimal results by varying the number of hidden neurons from 1 to 50.

Generalized regression artificial neural network (GRANN)

GRANN consisting of a single-pass neural network based on a general regression theorem GRANN which was first introduced by Specht (1991) is a different kind of radial basis function neural network (RBFANN), which is built on kernel regression network (Hannan et al. 2010) with one pass learning algorithm and highly parallel structure (Dudek 2011). Four layers make up the GRNN: the input layer, the pattern layer (also known as the radial basis layer), the summation layer, and the output layer. The northings and eastings, designated as($N_{i,j}, E_{i,j}$), were the input variables (independent datasets) in this investigation and the output variables (dependent datasets) were the northings and eastings denoted as$\left(N_{i,j}\right)$ and $(E_{i,j})$, respectively. The total number of observational parameters affects how many input units there are in the first layer. A training pattern and its output are delivered to each neuron in the pattern layer, which is coupled to the first layer. The

Fig. 2 Flow chart of methodology
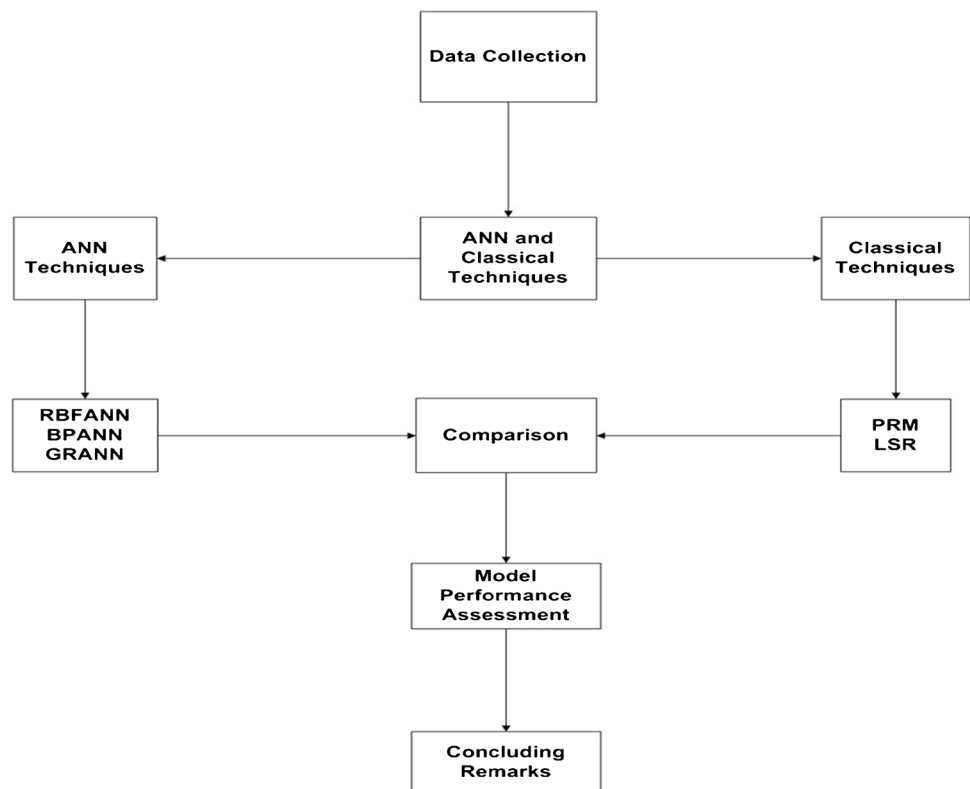
summation layer and the pattern layer are interconnected. The single division unit and summation unit are the two types of summation that make up the summation layer (Hannan et al. 2010). The output datasets are normalized by the summation and output layer together. Radial basis and linear activation functions are utilized in the hidden and output layers of the network during training. Two neurons in the summation layer are coupled to each unit in the pattern layer. The unweighted outputs of pattern neurons are computed by one neuron unit, while the weighted response of the pattern is computed by the other neuron unit. The estimated output variables are produced by dividing the output of each neuron unit by the other in the output layer. Similar with other network architectures, the input layer receives the inputs from the datasets, the pattern layer computes the Euclidean distance, and the summation layer comprises the numerator and denominator parts. The output is estimated using the weighted average of the training dataset outputs in which the weight is computed using the Euclidean distance between the training and test data, respectively. In GRANN model formulation, the spread constant constitutes the adjustable parameter that needs to be varied until the optimal results are achieved. The weighted hidden output layer from Eq. (6) is added to the output neuron in the current study, which applies the Gaussian function (Erdogan 2009) as

$$y(x) = \sum_{j=1}^{n} K_j \chi_j(x) \tag{6}$$

where n is the number of hidden neurons, $x \epsilon R^M$ is the input, K_j are the output layer weights of the radial basis function network, and $\chi_j(x)$ is Gaussian radial basis function given by Eq. (7) as (Idri et al. 2010; Srichandan 2012).

$$\chi_j(x) = e\left(\frac{-\| x_i - c_j \|^2}{\sigma_j^2}\right) \tag{7}$$

where $c_j \epsilon R^M$ and σ are the center and width of jth hidden neurons, respectively, $\|\,\|$ denotes the Euclidean distance.

Least square regression model (LSR)

The least square approach is a statistical method that can identify the line of best fit for a model and looks for the residuals with the lowest sum of squares. Regression analysis and estimation frequently employ this technique (Miller 2006; Peprah and Mensah 2017). Given a set of equations whose least squares solution is indicated by Eq. (8),

$$DZ \approx L \tag{8}$$

where $D \in R^{m \times n}, X \in R^{n \times d}, L \in R^{m \times d}, m \geq n$ (Annan et al. 2016; Schaffrin 2006). D is the design matrix, Z is the matrix of the unknown parameters, and L is the observation matrix.

Equation (9) indicates that the unknown parameters matrix Z can be solved using the ordinary least squares method:

$$Z = \left[D^T D\right]^{-1} \left[D^T L\right] \tag{9}$$

Equation (10) can be used to obtain the relevant error vector:

$$V = DZ - L \tag{10}$$

Kumi-Boateng and Ziggah (2020) improve the estimation performance of classical least squares using artificial neural network techniques in coordinate transformation in Ghana. The findings from their study revealed that the integration of the least square and neural network improved the transformation precision than the independent approach. Moreover, Yakubu et al. (2018) explore the efficiency and performance of ANN to classical least squares in adjusting DGPS survey field data. The results from their investigations showed that both ANN and least squares satisfactory adjusted the data to a good precision and ANN provides an alternative approach in adjusting survey field data.

Polynomial regression model (PRM)

In this study, the northings and eastings were modeled and predicted using a polynomial mathematical model. The independent variables were the horizontal coordinates (N, E). Equation (11), which is abbreviated as, provides the general expression of an m-degree polynomial interpolation (Yilmaz et al. 2017):.

$$Z(N, E) = \sum_{i=0}^{m} \sum_{j=0}^{m-1} a_{ij} N^i E^j \tag{11}$$

where $Z_{(N,E)}$ is the ellipsoidal height information of the point with known horizontal coordinates (N, E) and $a_{i,j}$ is the unknown polynomial coefficients to be estimated ($i, j = 0, \dots, m$). The effectiveness and performance of the simple planar (SP) polynomial model in calculating local heights was suggested by Dawod et al. (2022) and Peprah and Kumi (2017). The general SP polynomial model is denoted by Eq. (12) given as

$$Z_{ij} = a_o + a_1 N + a_2 E \tag{12}$$

where Z_{ij} is the estimated ellipsoidal heights, (N, E) are the horizontal coordinates of the stationary positions, and $a_{i,j}$ are the unknown parameters that can be determined using least square approach. PRM model has proven its superiority in solving majority of problems in geodesy studies comparing with other models. Notable among them are estimating orthometric heights in a mine (Peprah and Kumi 2017) and geoid modeling (Dawod et al. 2022).

Model performance assessment

Statistical error analysis was done to evaluate the accuracy of the models that were being used. Additionally, to produce a more comprehensive model performance analysis consistent with the training and testing results, dimensioned error statistics indicators such as the horizontal displacement (*HE*), arithmetic mean error (*AME*), arithmetic mean square error (*AMSE*), maximum and minimum residuals (*rmin* and *rmax*), and arithmetic standard deviation (*ASD*) were the statistical markers used. Their mathematical expressions are provided by Eqs. (13) through (18), respectively.

$$HE = \sqrt{\left(E_2 - E_1\right)^2 + \left(N_2 - N_1\right)^2} \tag{13}$$

$$AME = \frac{1}{n}\sum_{i=1}^{n}\left(\alpha_i - \beta_i\right) \tag{14}$$

$$AMSE = \frac{1}{n}\sum_{i=1}^{m}\left(\alpha_i - \beta_i\right)^2 \tag{15}$$

$$r_{\max} = \alpha_i - \beta_i \tag{16}$$

$$r_{\min} = \alpha_i - \beta_i \tag{17}$$

$$ASD = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\mu - \overline{\mu}\right)^2} \tag{18}$$

where n is the total number of the observations, α_i and β_i are the measured and adjusted 2D horizontal coordinated from the various techniques, μ denote the residual between the measured and adjusted field data, $\overline{\mu}$ is the mean of the residual, and i is an integer varying from 1 to n.

Results and discussions

Results

In defining the best RBFANN training model, more than one hyperparameter tuning was required in order to achieve the best optimal results. The RBFANN optimal solution leading to a smooth function was reached after sequential training trials with a spread parameter value of 0.8 and a maximum number of hidden neurons set of 20. As shown in Table 2, there is a close to perfect relationship between the observed and predicted output datasets. This is further indicated by the lowest statistical measurement recorded for the training and testing datasets. Moreover,

the error figures recorded in the training and testing results are relatively small, as revealed by the statistical indicators adopted. Hence, values predicted by the RBFANN model are very close to the observed values.

The authors conducted a comprehensive comparison between soft computing techniques (BPANN, RBFANN, and GRNN) and traditional methods (PRM and LSR) using the entire dataset, as summarized in Table 3. The statistical analysis presented in Table 3 demonstrates that the proposed ANN techniques achieved commendable results comparable to classical techniques. This conclusion is supported by the statistical assessments provided. Figures 3, 4, and 5 depict graphical representations of the AME, AMSE, and horizontal displacement model comparisons for all utilized techniques.

Discussions

Developing of artificial neural network models

The Bayesian regularization learning approach was employed to develop a single-layer BPANN model. The hyperbolic tangent activation was used in the hidden layer, while the linear activation functions were used in the output layer since the dataset observed satisfies a regression problem. During the training of the BPANN model using trainbr, both the hidden and output layers featured Tansig and Purelin functions. The ideal model structure, heavily reliant on the number of hidden neurons, was determined through a sequential trial-and-error method based on the lowest statistical indicators of AME, AMSE, rmax, rmin, and ASD. The model was trained with hidden neurons varying in number from 1 to 20 across iterative training sessions, each comprising 5000 epochs, a learning rate of 0.03, a minimum performance gradient of 0.0000001, a target of zero, a maximum number of validation failures of six, and a momentum coefficient of 0.9. The GRANN predictive model consisted of three inputs, a 0.9 width parameter, and two outputs. Validation concluded during neural network training upon reaching the minimum gradient and maximum epoch. The output of both the trained GRANN and RBFANN models was notably influenced by the spread constant's value. A sequential trial-and-error approach was utilized to determine the ideal width parameter values for GRANN and RBFANN in each iterative training session. Gradient descent rule was employed to train the GRANN and RBFANN models.

The ANN models (BPANN, GRANN, and RBFANN) were implemented and programmed using MATLAB (R2018a). After multiple training sessions adjusting the 2D horizontal coordinates (northings and eastings), the BPANN model eventually converged to the optimal model structure, which was denoted as [3 10 2]. Statistical analysis revealed that the RBFANN structure provided the lowest minimum

Table 2 Model results for soft computing techniques (units in meters)

PCI	AME	AMSE	r_{min}	r_{max}	ASD
BPANN (northings)					
Training	0.005327207	0.019865396	−0.035306996	0.012808901	5.89995E−16
Testing	0.00167678	0.000843477	−0.003768905	0.007652645	0.063332897
BPANN (eastings)					
Training	−0.001294345	0.001172731	−0.447174539	0.004457873	9.05791E−17
Testing	0.00043671	5.72147E−05	−0.00164063	0.007652645	0.030032856
GRANN (northings)					
Training	−0.000662951	0.000307653	−0.671967934	0.671926541	5.31205E−16
Testing	0.006130126	0.011273534	−0.428210783	0.57443724	0.117856134
GRANN (eastings)					
Training	0.000617074	0.000266546	−0.614193357	0.625187136	2.2256E−16
Testing	−0.0002536	1.9292E−05	−0.5128985	0.71132381	0.01510547
RBFANN (northings)					
Training	8.04929E−11	4.53537E−18	−1.3039E−08	1.40863E−08	2.60453E−23
Testing	**−1.59877E−10**	7.6682E−18	−6.0536E−09	0	4.17029E−09
RBFANN (eastings)					
Training	−9.48E−12	6.2903E−20	−1.199E−08	1.1059E−08	1.1963E−23
Testing	−1.65698E−10	8.23673E−18	−5.93718E−09	0	2.87477E−09
LSR (northings)					
Training	−27.6033942	533,363.162	−499.517304	313.566744	8.2776E−13
Testing	−288.734	25,010,191	−625.414	170.976	11.48556
LSR (eastings)					
Training	206.0399485	29,716,722.26	−110.823544	648.032361	5.82869E−12
Testing	372.7272	41,677,663	112.2059	602.0103	7.229777
PRM (northings)					
Training	104.1188491	7,588,514.312	−367.8885678	445.3114858	1.92641E−12
Testing	−156.323588	7,331,119.26	−493.666256	302.750807	2.2354E−12
PRM (eastings)					
Training	−90.68795433	5,757,013.542	−407.5571264	351.1524406	6.06949E−12
Testing	75.708192	1,719,519.1	−184.31745	305.430442	5.6542E−13

**Table 3** Statistical analysis of all the models (units in meters)

| PCI	r_{max}	r_{min}	AME	AMSE	ASD
BPANN	0.005732	0.446649	0.00404	0.019055	0.070093
GRANN	0.130091	0.263966	0.006849	0.010969	0.11882
RBFANN	1.79091E−08	9.24311E−09	2.86674E−10	8.75363E−18	5.06513E−09
LSR	149.8338	256.1096	309.7964	2.7242966E7	13.57158
PRM	149.7132	256.2342	309.0598	2.7242966E7	13.57158

Fig. 3 AME model graph of the models
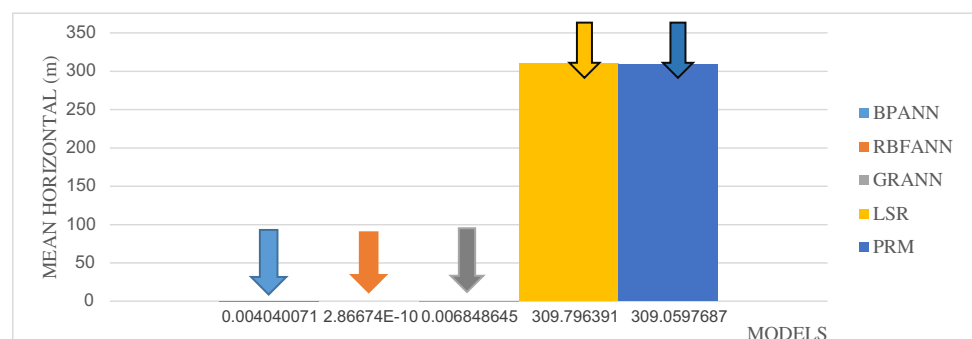
Fig. 4 AMSE model graph of
the models
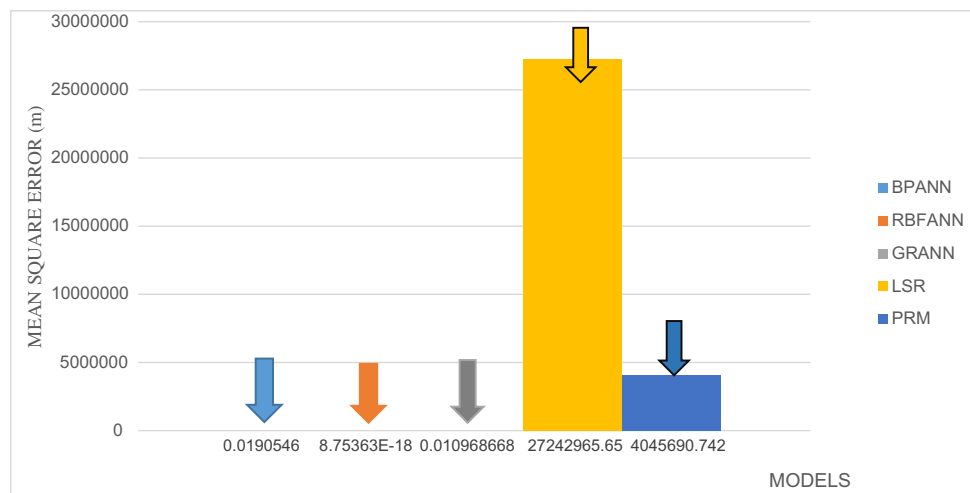
Fig. 5 Horizontal displacement
of models
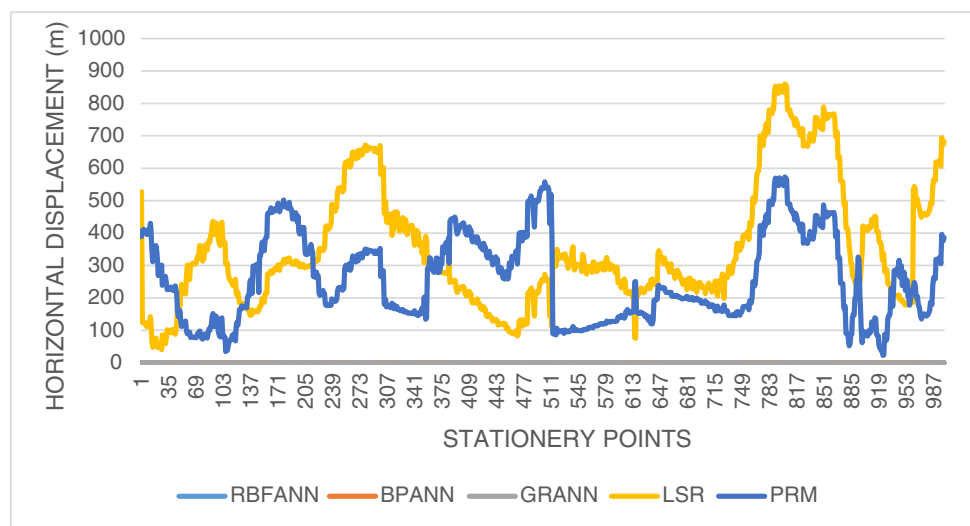
values in terms of AME, AMSE, rminimum, maximum x,
and ASD. The ideal RBFANN prediction model for correct-
ing 2D horizontal data was achieved with a spread param-
eter of 0.1. This optimization was achieved by varying the
spread parameter from 0 to 1 during each training iteration,
resulting in the best model denoted as [3 20 0.1 2], com-
prising a spread constant of 0.1, a maximum of 20 hidden
neurons, 3 input variables (independent variables), and 2
output variables (dependent variables). For modifying the
eastings, the most effective RBFANN model was identified
as [2 20 0.7 1], featuring a hidden layer with a maximum
of 20 hidden neurons, two inputs (independent datasets),
two outputs (dependent datasets), and a width parameter of
0.7. The summarized outcomes of training and testing using
all soft computing methodologies are presented in Table 2.
The statistical findings in Table 2 indicate that soft com-
puting approaches have proven to be effective in modify-
ing survey field data for the study area with significantly

enhanced precision. Encouragingly, the residuals reached
values of $-1.199E-08$ at the minimum and $-5.93718E-09$
at the maximum. Furthermore, the arithmetic mean error
(AME), arithmetic mean square error (AMSE), and arith-
metic standard deviation (ASD) for both training and testing
demonstrated high levels of respectability and encourage-
ment. Consequently, ANN emerged as a highly effective and
practical alternative technique for accurately modifying RTK
GPS data for the research area.

Developing the classical regression models

The Minitab 19 program was used to statistically describe
the data and determine the correlation between the inde-
pendent variables (input datasets) and the dependent dataset
before the PRM model was created (output datasets). Equa-
tion (19) is the ideal PRM equation produced by the Minitab
software for calculating the 2D coordinates ($E(i)$, $N(i)$):

$$E(i) = 308987 + (0.4794N(i)) - 19.26Z(i) \tag{19}$$

$$(i) = 217458 + (0.7838E(i)) + 26.18Z(i) \tag{20}$$

where $E(i)$, $N(i)$ are the dependent datasets, $Z(i)$ is the independent variables (ellipsoidal heights), {308,987, 0.4794, 19.26}, {217,458, 0.7838, 26.18} are the generated unknown parameters by the Minitab software. The software was more direct in its use and suits the author preference as compared to other statistical software. The final estimated $E(i)$ and $N(i)$ values with the given equation and parameters were coded and implemented in MATLAB environment. The LSR model's unknown parameters were estimated using the least squares method, which was developed and put into practice in MATLAB. The statistical performance results of the models are shown in Table 2.

Comparing the predictive performance of the ANN models with the classical models

The ranking effectiveness of the models applied BPANN, RBFANN, GRANN, LSR, and PRM are compared using the statistical indicators of rmin, rmax, AME, AMSE, and ASD. Tables 2 and 3 provide the best option for evaluating the model predicted 2D point data of RTK GPS survey field data. Tables 2 and 3 show that the model with the lowest AMSE recorded statistical indicators of the best absolute fit to the data. It can be revealed in Table 3 that the RBFANN model yielded the best AMSE value after computing the horizontal displacement with a value of 8.75363 E − 18 m, followed by BPANN and GRANN which achieved 0.019055 m and 0.010969 m, respectively. These results can be confirmed in Fig. 4.

The rmin and rmax statistical indicators show the accuracy of the predictive models by revealing the minimum and maximum of their differences between the actual dataset and the predicted values as an error. Tables 2 and 3 indicate performance rmin and rmax results for all the models adopted, respectively. In effect, RBFANN achieved a better rmin and rmax values of 1.79091 E − 08 m and 9.24311 E − 09 m, respectively. In comparison, the BPANN and GRANN models achieved relatively results of 0.005732 m, 0.446649 m, 0.130091 m, and 0.263966 m, respectively. The worst performing testing models in rmin and rmax were the LSR and PRM models recording 149.8338 m, 256.1096 m, 149.7132 m, and 256.2342 m, respectively.

The AMSE is commonly used in model formulation assessment to represent the error loss function and indicates how closely the regression line best fit the set of point. The lower the AMSE values, the better the model and output dataset. Table 3 shows minimal AMSE value of 8.75363 E − 18 m for indicating better predictive model. Conversely,

from Table 3 results, the RBFANN model proved its superiority in enhancing the survey field data over the other models in terms of the AMSE evaluation with LSR and PRM performing poorly in that regard.

The ASD is the degree of deviation of the variables from their true mean. It is observed from Table 3 that the RBFANN recorded 5.06513 E − 09 m which implies that the predicted values of the dataset did not deviate much from their true mean and therefore there is a good relationship between the observed and the predicted values. However, the best predicted model was observed for the RBFANN model as it recorded the best ASD value. Hence, the RBFANN predictions exhibit higher replication on the historical datasets.

A deeper examination of Table 3 reveals that the predictions generated by the proposed ANN models closely matched observed 2D coordinates (eastings and northings), exhibiting superior prediction accuracy over PRM and LSR models, which demonstrated lower accuracy. Specifically, the maximum residual values for conventional techniques and soft computing techniques were 149.7132 m and 0.130091 m, respectively. Results from the indices used for the accuracy assessment attest to the fact that ANN models achieve optimum results when used for survey adjustment in comparison to traditional methods. The ANN models displayed a superior performance due to its ability to learn and model nonlinear relationships.

It is noteworthy that the PRM model displayed an ASD of 13.57158 m and a minimum residual value of 256.2342 m, aligning with findings from prior studies (Chen and Hill 2005; Poku-Gyamfi 2009). This suggests that the PRM model's limitations were associated with increased order and distortions in estimated values arising from the least squares approach, reaffirming the recommendation for maintaining a lower order.

In the context of adjusting observed 2D coordinates, soft computing techniques outperformed traditional methodologies. The comparatively less effective performance of conventional procedures may be attributed to unidentified parameters used in model estimation. The statistical analysis of both computing types consistently demonstrates the superiority of soft computing techniques in determining 2D coordinates for the research region.

Conclusions and recommendations

Enhancing survey field data has been a common practice over the years in geodesy in order to determine the magnitude of errors and the acceptable tolerance level for topographic, astronomic, geodetic, cartographic, navigation, and datum-related studies. This procedure is carried out using the traditional least squares regression

approach. However, existing literature has shown that little or no alternative technique has been tested to serve as a substitute for the traditional least square. The main contribution of this study is to explore the superiority of ANN as a realistic alternative technology for enhancing GPS survey field data. In this study BPANN, RBFANN based on the supervised learning technique as well as GRANN, LSR, and PRM have been presented. The findings from the study reveal that the RBFANN offered satisfactory prediction of the 2D RTK survey data. However, RBFANN compared to BPANN, GRANN, LSR, and PRM showed superior stability and more accurate prediction results. It can therefore be proposed based on the results achieved in this study that the RBFANN should be used instead of BPANN, GRANN, LSR, and PRM within the study area for enhancing survey field data.

On the basis of the results and analysis achieved, it has been revealed that northings and eastings feed into the RBFANN model could successfully produce accurate estimates of the enhanced 2D datasets. Therefore, this study does not have a localized significance but will also open up more scientific discourse into the applications of ANN in enhancing GPS survey field data within the geoscientific community.

In the era of digital revolution marked by innovative discoveries in science and technology, the conventional (traditional) methods for identifying stationary and non-stationary ground truth locations continue to face challenges. In developing countries in the global south such as Ghana, the Global Positioning System (GPS), as an integral component of the Global Navigation Satellite System (GLONASS), plays a pivotal role in applications spanning cadastral surveying, topographical surveys, and advanced engineering projects. However, the assumption that raw RTK GPS field data is inherently exact and accurate, fit for all engineering endeavors, warrants a closer examination.

The critical domains of adjustment and computation in mathematics and satellite geodesy have garnered significant attention over the years as they serve to ascertain the alignment of data with acceptable tolerance levels. While past research has predominantly emphasized traditional methodologies for survey data adjustment, recent investigations have explored the integration of artificial intelligence (AI) to enhance field data modification in RTK GPS data collection.

This study represents an endeavor to bridge these methodologies, employing artificial neural networks (ANN) alongside conventional techniques to adapt field data from an RTK GPS survey using AI. Through rigorous analysis, the study demonstrates that ANN approaches surpass traditional regression procedures in terms of statistical performance. Specifically, the radial basis function artificial neural network (RBFANN) stands out for its superior

accuracy and precision when compared to back-propagation artificial neural network (BPANN) and generalized regression artificial neural network (GRANN), as evident in their respective statistical analyses, and as asserted by Konakoglu and Cakir (2018).

As a result, the application of soft computing methods for RTK GPS data modification in the study area emerges as a more pragmatic approach than reliance on traditional least squares methods. This endeavor contributes significantly to the understanding of Ghanaian geospatial experts, offering valuable insights into the effectiveness of applying soft computing methods to precisely determine stationary points for geodetic purposes.

While this study provides valuable insights, further research in Ghana is warranted to harness various deep learning soft computing methods. These encompass deep learning convolutional neural networks (CNN), least squares support vector machines, extreme learning machine, and various learning algorithms, including Broyden-Fletcher-Goldfarb-Shanno quasi-Newton, resilient back-propagation, Fletcher-Reeves conjugate gradient, scaled conjugate gradient, Polak-Ribiere conjugate gradient, conjugate gradient with Powell/Beale restarts, one-step secant back-propagation, gradient descent, gradient descent with adaptive learning rate, gradient descent with momentum, and gradient descent with momentum and adaptive learning rate. Such investigations will evaluate their performance in addressing more intricate engineering tasks, thus expanding the realm of computational possibilities.

Beyond its impact on Ghanaian geodesists, this study contributes substantively to the scholarly discourse surrounding the application of soft computing techniques in addressing challenges within the surveying and related fields. By advancing the understanding of these techniques and their efficacy, it paves the way for more precise and efficient surveying practices in a rapidly evolving technological landscape.

Data availability The authors confirm that the data supporting the findings of this study are available within the article and its supplementary materials. Raw data that support findings of the study are available from the corresponding author, upon reasonable request.

Declarations

Conflict of interest The authors declare that they have no competing interests.

References

Akcin H, Celik T (2013) Performance of artificial neural networks on kriging method in modeling local geoid. Bol Ciênc Geod 19(1):84–97

Annan RF, Ziggah YY, Ayer J, Odutola CA (2016) Accuracy assessment of heights obtained from total station and level instrument using total least squares and ordinary least squares methods. J Geomat Plan 3(2):87–92

Asenso-Gyambibi D, Lamkai N, Peprah MS, Larbi EK, Asamoah B, Okantey P (2022) Novel ellipsoidal heights predictive models based on artificial intelligence training algorithms and classical regression models techniques : a case study in the Greater Kumasi metropolitan area local geodetic reference. Int J Earth Sci Knowl Appl 3(2021):493–515

Bashirov A, Mazhar Z, Ertürk S (2008) Boundary value problems arising in Kalman filtering. Bound. Value Probl. https://doi.org/10.1155/2008/279410

Caleb M, Atayi J, Kabo-bah A, Švik M, Acheampong D (2018) Assessing the impacts of urbinization on the climate of Kumasi. May 2020. https://doi.org/10.20944/preprints201809.0059.v1

Chen W, Hill C (2005) Evaluation procedure for coordinate transformation. J Surv Eng 131(2):43–49

Dawod GM, Al-krargy EM, Amer HA (2022) Accuracy assessment of horizontal and vertical datum transformations in small-areas. GNSS Surveys in Egypt 8(1):19–28

Dudek G (2011) Generalized regression neural network for forecasting time series with multiple seasonal cycles. Springer-Verlag, Berlin Heidelberg 1:1–8

Erdogan S (2009) A comparison of interpolation methods for producing digital elevation models at the field scale. Earth Surf Process Landforms 34:366–376

Erol B, Celik RN (2005) Modelling local GPS/levelling geoid with the assessment of inverse distance weighting and geostatistical kriging methods. Greece, Athens, pp 1–5

Ghana Statistical Service (2014) Population and housing census estimates. Stats Ghana, Accra Ghana. www.ghanastatisticalservice.com

Ghilani CD, Wolf PR (2006) Adjustment computations: spatial data analysis, 4th edn. Hoboken, NJ: John Wiley & Sons

Golub GH, Van Loan CF (1989) Matrix computations, 2nd edn. Johns Hopkins University Press, Baltimore, MD, p 200

Hannan SA, Manza RR, Ramteke RJ (2010) Generalized regression neural network and radial basis function for heart disease diagnosis. Int J Comput Appl 7(13):7–13

Hornik K, Stinchcombe M, White H (1989) Multilayer feed forward networks are universal approximators. Neural Network 2:359–366

Idri A, Zakrani A, Zahi A (2010) Design of radial basis function neural networks for software effort estimation. Int J Comput Sci 4(7):11–17

Kaloop MR, Rabah M, Hu JW, Zaki A (2017) Using advanced soft computing techniques for regional shoreline geoid model estimation and evaluation. Mar Georesour Geotechnol 1–11. https://doi.org/10.1080/1064119x.2017.1370622

Kleijnen JPC (2009) Kriging metamodeling in simulation: a review. Eur J Oper Res 192(3):707–716. https://doi.org/10.1016/j.ejor.2007.10.013

Konakoglu B, Cakir L (2018) Generalized regression neural network for coordinate transformation. AIST, Montenegro, pp 66–78

Kumi-Boateng B, Ziggah YY (2020) Empirical study on the integration of total least squares and radial basis function neural network for coordinate transformation. GJSTD 7(1):38–57

Lee JM, Min KS, Min WK, Park H (2020) A study on the actively capture of road construction information using spatial analysis. J Korean Soc Cadastre 36(2):149–159

Maidaniuk I, Tsoi T, Hoian I, Doichyk M, Patlaichuk O, Stupak O (2022) The problem of artificial intelligence in contemporary philosophy. Broad Research in Artificial Intelligence and Neuroscience 13(4):436–449. https://doi.org/10.18662/brain/13.4/397

Mihalache RM (2012) Coordinate transformation for integrating map information in the new geocentric European system using Artificial Neural Networks. GeoCAD 1–9

Miller SJ (2006) Methods of least squares, statistics theory. Cornell University, USA 3:1–2

Mueller VA, Hemond FH (2013) Extended artificial neural networks: in-corporation of a priori chemical knowledge enables use of ion selective electrodes for in-situ measurement of ions at environmental relevant levels. Talenta 117:112–118

Oduro C, Kafui O, Peprah C (2014). *Analyzing growth patterns of Greater Kumasi metropolitan area using GIS and multiple regression techniques analyzing growth patterns of Greater Kumasi metropolitan area using GIS and multiple regression techniques*. 7(5). https://doi.org/10.5539/jsd.v7n5p13

Okwuashi O, Eyoh A (2012) Application of total least squares to a linear surveying network. J Sci Arts 4(21):401–404

Ophaug V, Gerlach C (2017) On the equivalence of spherical splines with least-squares collocation and Stokes's formula for regional geoid computation. *Springer*, 1–16. https://doi.org/10.1007/s00190-017-1030-1

Osei-Nuamah I, Appiah-Adjei EK (2017) Hydrogeological evaluation of geological formations In Ashanti Region. Ghana J Sci Technol 37(1):34–50

Peprah MS, Kumi SA (2017) Appraisal of methods for estimating orthometric heights – a case study in a mine. J Geosci Geomat 5(3):96–108

Peprah MS, Mensah IO (2017) Performance evaluation of the ordinary least square (OLS) and total least performance evaluation of the ordinary least square (OLS) and total least square (TLS) in adjusting field data : an empirical study on a DGPS data. S Afr JGeomat 6(1):73–89. https://doi.org/10.4314/sajg.v6i1.5

Poku-Gyamfi Y (2009) Establishment of GPS reference networking Ghana. MPhil Dissertation. Universitat Der Bundeswehr Munchen Werner Heisenberg-Weg, Germany 39(85577):1–218

Ruffhead A (2012) An introduction to least-squares collocation. Surv Rev 29(224):85–94. https://doi.org/10.1179/003962687791512662

Schaffrin B (2006) A note on constrained total least square estimation. Linear Algebra Appl 417:245–258

Specht D (1991) A general regression neural network. IEEE Trans Neural Networks 6(2):568–576

Srichandan S (2012) A new approach of software effort estimation using radial basis function neural networks. ISSN (print) 1(1):2319–2526

Wang J (2014) *Test statistics in Kalman filtering test statistics in Kalman filtering*. June 2008. https://doi.org/10.5081/jgps.7.1.81

Yakubu I, Ziggah YY, Peprah MS (2018) Adjustment of DGPS Data using artificial intelligence and classical least square techniques. J Geomat 12(1):13–20

Yilmaz M, Turgut B, Gullu M, Yilmaz I (2017) Application of artificial of artificial neural networks to height transformation. Tehnicki Vjesnik 24(2):443–448

Yonaba H, Anctil F, Fortin V (2010) Comparing sigmoid transfer functions for neural network multistep ahead stream flow forecasting. J Hydrol Eng 15(4):275–283

Ziggah YY, Tierra A, Youjian H, Konate AA, Hui Z (2016) Performance evaluation of artificial neural networks for planimetric coordinate transformation-a case study, Ghana. Arab J Geosci 9:698–714

Ziggah YY, Youjian H, Laari PB, Hui Z (2017) Novel approach to improve geocentric translation model performance using artificial neural network. Geoid. Bol Ciênc Geod 213–233. https://doi.org/10.1590/S1982-21702017000100014