Data classification on the geoid surface with grey wolf optimization

Ulku Kirici Yildirim, Yasemin Sisman

Department of Geomatic Engineering, Engineering Faculty, Ondokuz Mayıs University (OMU),
55200 Samsun, Türkiye – ulku.kirici@omu.edu.tr, ysisman@omu.edu.tr

Keywords: Data Classification, Outlier Measurement, Optimization, Metaheuristic Optimization, Grey Wolf Optimization

Abstract

Geoid determination is the process of modelling that enables the estimation of the height of a point with a known position. With the advancement of GNSS technologies, geoid determination has become one of the central problems in Geodesy. One of the critical issues in this process is identifying outliers in the data set and classifying data as compatible or outlier. This is because outlier points can significantly deteriorate the accuracy of the model and must be removed from the dataset. Traditionally, the identification of outlier measures in data classification has been carried out through statistical tests within the framework of the Least Squares (LS) method. However, recent developments in artificial intelligence and metaheuristic algorithms have provided powerful alternatives for solving complex optimization problems. One such problem is data classification. Among these algorithms, the Grey Wolf Optimizer (GWO), inspired by the social hierarchy and hunting strategy of grey wolves, has attracted increasing attention. In this study, the applicability of the data classification in GWO algorithm for outlier detection in geoid determination was investigated. The application area is located within the Ondokuz Mayıs University campus in Samsun province, and the dataset consists of 3555 points collected by GNSS/leveling measurements in a relatively flat and structure-free region. A second-degree polynomial surface was fitted to the dataset, and outlier detection was carried out using both LS and GWO methods. The LS method identified 493 outlier points, whereas the GWO classified 472 outlier points. A comparative analysis revealed that 181 points (38.35%) were commonly detected by both approaches, with a higher concentration in the upper and lower parts of the study area. The results demonstrate that GWO is capable of detecting a larger set of anomalous measurements compared to the conventional LS method. This indicates that GWO, by exploring the solution space globally, can capture additional outliers that might remain undetected by traditional statistical approaches. Therefore, the findings suggest that GWO can be considered a robust and complementary tool to classical methods.

1. Introduction

In the applied sciences, the basic principle is to take more measurements than necessary in order to increase the accuracy of the measurements and the results obtained. As data collection methods based on different technologies have increased, so has the amount of data used in geodetic studies. However, this diversity also introduces the possibility of errors in the measurement data (Le Goïc et all.,2012). Especially rough errors can seriously affect the accuracy of data analyses. Therefore, for reliable modelling and engineering applications, it is necessary to effectively remove observations with rough errors from the measurement data. The accuracy of the study highly depends on proper data classification and the effective separation of outlier measurements from the dataset (Yang et all, 2013).

Data classification plays a crucial role in ensuring the reliability of geodetic analyses. In geoid determination and other geospatial applications, the measurement dataset typically includes both consistent observations and potential outliers. If outliers are not correctly identified and removed, they can significantly distort the model and lead to mistake results. Conversely, over-elimination of valid data points may reduce the robustness of the solution. Therefore, developing efficient classification strategies has become a central issue in modern geodesy (Lehmann R, 2013).

With the development of technology, problems in science have started to change and the solutions to these problems have also started to change. In addition, optimization has entered our literature as a problem encountered in many of these developments. Optimization is related to almost every branch of science from business, economics, design and production and is the search for the best solution among different possible solutions to a problem (Gandomi & Roke, 2014). Optimization algorithms are generally separated into heuristic optimization algorithms and mathematical optimization algorithms (Kumar et all, 2017). Heuristic optimization algorithms approach the solution set heuristically and aim to reach the best solution. Nowadays, metaheuristic algorithms have been developed as a result of the effective use of basic heuristic methods in combination. These algorithms are nature-inspired algorithms that are inspired by the behaviour of animals living in herds in nature or plants with different habitats. Nature-inspired algorithms have an advantage over classical methods in that they can be adapted to different problems (Yang, 2011). There are many different metaheuristic algorithms used in the studies. Among the most widely applied are Genetic Algorithms (GA) (Al Qaraghuli, et all, 2022), Particle Swarm Optimization (PSO) (Arora, N., & Banyal, R. K., 2021), Differential Evolution (DE), Ant Colony Optimization (ACO) (Maniezzo, V. et all, 2004), Artificial Bee Colony (ABC) (Karaboga, D., & Basturk, B., 2008), Firefly Algorithm (FA) (Yang, X. S. 2010), and Simulated Annealing (SA) (Nikolaev, A. G., & Jacobson, S. H., 2010). These algorithms are inspired by natural processes such as biological evolution, swarm intelligence, and physical phenomena, and they have been successfully applied to a broad range of optimization problems. In geodetic and geospatial studies, metaheuristic algorithms have been used for tasks such as parameter estimation, geoid modelling, GNSS network adjustment, and outlier detection. Each algorithm has its own strengths and limitations, but they all share the advantage of being able to search complex and multidimensional solution spaces where classical optimization methods may fail.

Among these algorithms, the GWO has recently attracted particular attention. Introduced by Mirjalili et al. (2014), the GWO algorithm mimics the leadership hierarchy and cooperative hunting strategy of grey wolves in nature. Its main advantage over many other metaheuristic methods is its simplicity, as it requires only a few control parameters (Cao M., et all , 2019).

In this study, GWO, a nature-inspired metaheuristic algorithm, is employed to detect and eliminate outlier measurements in geoid determination. The application area is located within the Ondokuz Mayıs University campus in Samsun province, and the dataset consists of 3,555 points collected from a relatively flat and structure-free region to ensure accurate modelling. A second-degree polynomial surface was fitted to the dataset, and both the conventional LS method and the GWO algorithm were applied for data classification. The data were divided into two classes: outlier and compatible. The LS method identified 493 outliers, whereas GWO classified 472 outliers, with 181 points commonly detected by both methods. When the outliers classified by both methods were examined, it was observed that there was a similarity of 38.35% in the number of common points. The spatial distribution of the points classified as outliers by the two methods was analysed, and the results were interpreted accordingly.

2. Geoid Determination and Data Classification

The geoid is a complex surface defined as an equipotential surface of the Earth's gravity field, representing the mean sea level that continues under the continents (Tukka, A. A., et all, 2025). Since the geoid does not exhibit a regular geometric shape due to the irregular distribution of mass within the Earth, its accurate determination remains one of the fundamental challenges in geodesy. Local geoid determination studies focus on estimating the geoid surface in a specific region using various geoid modelling techniques, such as the Polynomial Interpolation Method (Akar, Konakoğlu, & Akar, 2022).

In engineering studies, the relationship between the number of unknowns and the number of measurements is crucial for improving accuracy. If the number of unknown parameters (u) is equal to the number of observations (n), the solution is unique. However, if the number of observations exceeds the number of unknowns (n > u), multiple solutions may exist (Dasgupta & Mishra, 2004). When the number of observations exceeds the number of unknowns in a problem, adjustment calculations are performed to obtain a unique and consistent solution (Montgomery, Peck, & Vining, 2021). Through adjustment computation, the precise values of the observations and unknowns are determined according to mean square errors and objective functions. Measurement errors are classified as rough, systematic, and random, and adjustment calculations are carried out under the assumption that only random measurement errors are present. Random measurement errors are assumed to follow a normal distribution. Rough and systematic errors that are close in magnitude to random errors are referred to as outliers. In cases where measurement errors do not conform to the normal distribution of the data, any observation that deviates from the mean and variance of the measurement set is considered an "outlier." Adjustment is a technique used to estimate the most probable values of unknown parameters when redundant observations are available; in addition, statistical measures such as precision and reliability can be derived as by-products (Ogundare, 2018). Several methods have been developed for adjustment computation, one of the most widely used being the Least Squares (LS) method.

Nowadays, with the developing technology, the amount of data held has increased and it has become important to classify the data obtained. Statistical methods and machine learning are often used in data classification. In this study, the data were categorised as compatible and outlier. Thus, outlier points were identified.

2.1 Conventional data classification method

The LS method explained by Carl Friedrich Gauss in 1795 and Legendre in 1805. This method is used in many different applications (Sisman, 2014). The Least Squares Method is a mathematical approach used to estimate unknown parameters by minimizing the sum of the squares of the residuals the differences between observed and computed values. Unknown parameters calculated with the following equation in this method.

$$\underline{X} = \left(\underline{A}^T \underline{Q}_{\ell\ell}^{-1} \underline{A}\right)^{-1} \underline{A}^T \underline{Q}_{\ell\ell}^{-1} \underline{\ell} \tag{1}$$

Root mean square error (RMSE);

$$m_0 = \pm\sqrt{\frac{\underline{V}^T \underline{PV}}{f}} \quad ; \text{f = n – u} \tag{2}$$

In order to identify outlier measures, the null hypothesis for n number of measures is set as $H_0 : E\{\Delta_i = 0\}$ by considering the presence of at least one outlier measure in the measures. The alternative hypothesis is established with the equation $H_S : E\{\Delta_i \neq 0\}$ if all measures are considered to be compatible (Heinrichs, F., Bastian, P., & Dette, H., 2025).

Then, the test value is calculated.

$$T_i = \frac{|V_i|}{m_{v_i v_i}} \tag{3}$$

This test value is compared with the $q = T_{f,1-\alpha/2}$ table values. If $T_i > T_{f,1-\alpha/2}$, this measurement is accepted as outlier measurements. The measurement has got a biggest value as outlier is removed the measurement group, then this procedure repeated until there are no outlier measurement (Wolf and Ghilani, 1997, Teke and Yalçınkaya, 2005).

3. Grey Wolf Optimization (GWO)

GWO is a population-based algorithm inspired by the social hierarchy and hunting behaviour of grey wolves proposed by Seyedali Mirjalili, Seyed Mohammad Mirjalili and Andrew Lewis in 2014 (Mirjalili & Lewis, 2015). This algorithm is based on the leadership structure of a grey wolf pack and is often used to solve optimization problems. GWO is based on mathematical modelling of the social roles of the wolves in the pack and the effects of these roles on predation. At the heart of the algorithm are three main wolf types that simulate the behaviour of a grey wolf pack: alfa (α), beta (β) ve delta (δ) The remainder are called omega (w).These wolves represent potential solutions in a solution space. The position of each wolf type represents a solution candidate.

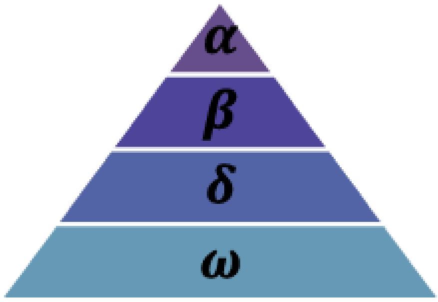

Figure 1. GWO social hierarchy (Mirjalili vd., 2014).

Hunting in the GWO is primarily guided by the alpha, beta, and delta wolves, which represent the best three candidate solutions in the population. These wolves play a leadership role by directing the search process toward promising regions of the solution space. During the optimization, the remaining wolves known as omegas update their positions by encircling and following the alpha, beta, or delta wolves, thereby balancing exploration and exploitation. This hierarchical mechanism not only simulates the natural hunting strategy of grey wolves but also ensures that the algorithm effectively converges toward the global optimum while avoiding premature stagnation in local minima.

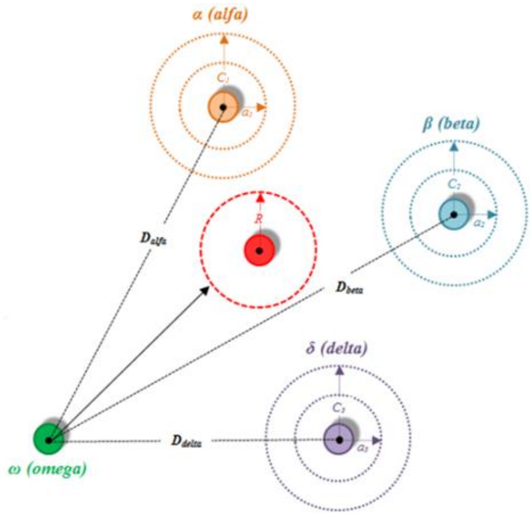

Figure 2. The influence of the leading team during the hunting phase (Karakoyun, 2021)

$$D = \left| C.X_p(t) - X(t) \right| \tag{4}$$

$$X(t+1) = \left| X_p(t) - A.D \right| \tag{5}$$

Where t is the current iteration, A and C are the coefficient vectors, Xp is the location vector of the prey, X is the location of a grey wolf.

$$A = \left| 2a.r_1 - a \right| \tag{6}$$

$$C = \left| 2a.r_2 \right| \tag{7}$$

Grey wolves have the ability to find and surround their prey. Although beta and delta are sometimes involved in this hunt, it is usually guided by alpha.

In GWO, after the location of the prey is determined, the prey is attacked. The attack on the prey takes place after the prey gets tired and stops moving. Considering the modelling process mathematically, the attack process takes place according to the A value. If the value of A is greater than 1, grey wolves move away from the prey and start looking for a more suitable prey. If it is less than 1, grey wolves are forced to attack the prey.

In GWO, fishing continues until either the stopping criterion is met or the specified number of iterations has been reached.

The GWO workflow is as shown in Figure 3.

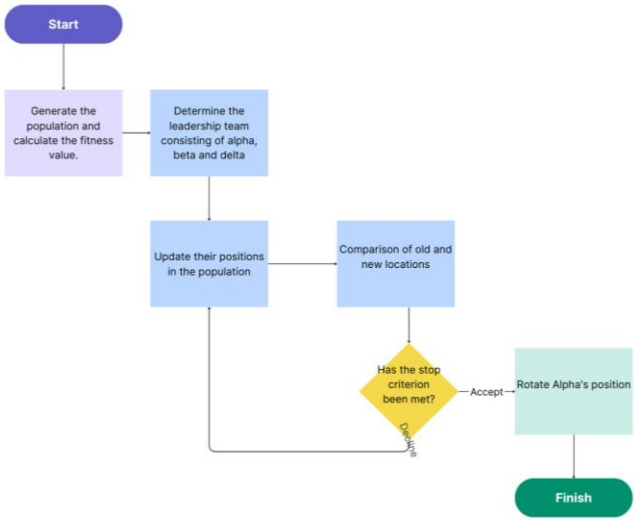

Figure 3. Workflow

4. Case Study

In this study, GWO algorithm is employed as a robust metaheuristic approach for the detection of outlier measurements within geodetic datasets. The selected application area covers the Ondokuz Mayıs University campus, situated in the Samsun province of Türkiye, which provides a suitable test environment due to its diverse topographic and structural features. Ondokuz Mayıs University was established on a very large area. A relatively flat area with no buildings was chosen as the working area (Figure 4). The dataset used in the analysis comprises a total of 3555 points.

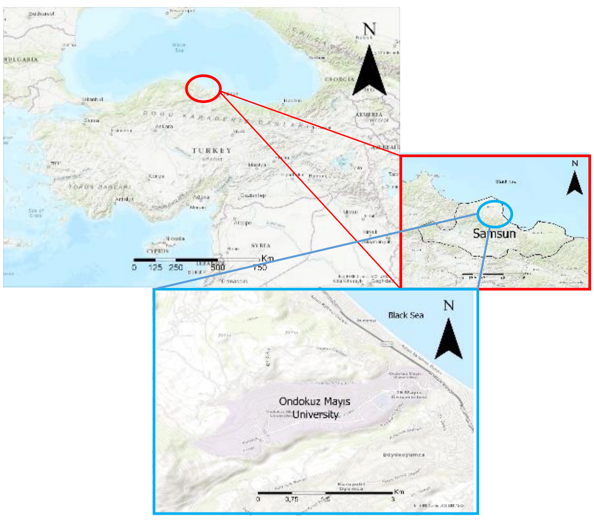

Figure 4. Study area

These 3D point cloud data were obtained through field measurements conducted in the selected area. The study area was carefully selected to improve the accuracy of geoid modelling, prioritizing a relatively flat region without built structures. These 3D point cloud data were obtained through field measurements conducted in the selected area. The distribution of points with known x, y and h values is shown in Figure 5.

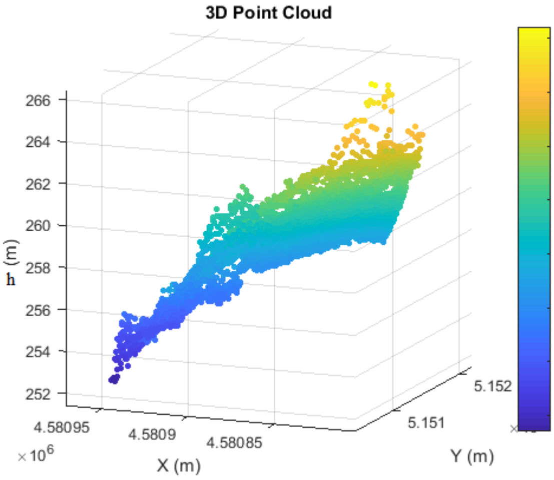

Figure 5. Data set

5. Application

First, a surface model was fitted to the dataset using the conventional least squares approach with a second-degree polynomial, in order to represent the geometric characteristics of the study area. Following the surface fitting process, an outlier detection test was conducted to evaluate the residuals and identify measurements that significantly deviated from the model predictions. In the analysis, the dataset was classified into two categories compatible and outlier measurements according to statistical thresholds determined from the LS residuals. As a result of this classification, a group of 493 points was identified as outliers, indicating potential measurement errors or anomalies in the dataset (Figure 6).

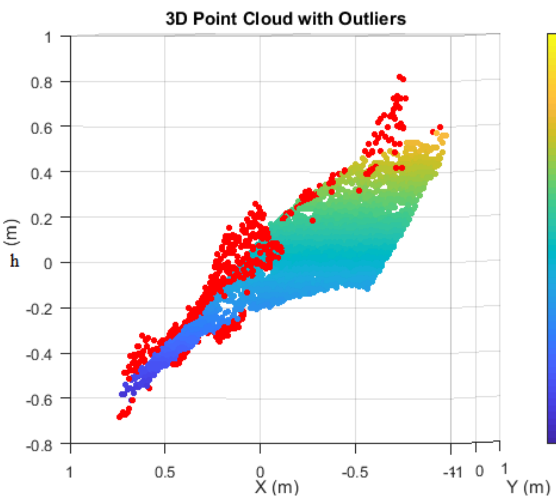

Figure 6. Outlier points of the LS Method

This preliminary step provided a reference baseline for subsequent comparisons with the GWO-based outlier detection method. Subsequently, the GWO algorithm was applied to the data set to identify outliers by classifying the data. The GWO algorithm searches for the optimal solution by minimising the sum of squared residuals, mimicking the hunting behaviour of grey wolves. This process enabled the algorithm identify measurements that deviated significantly from the model surface. As a result, the GWO-based analysis classified 472 points as outliers (Figure 7).

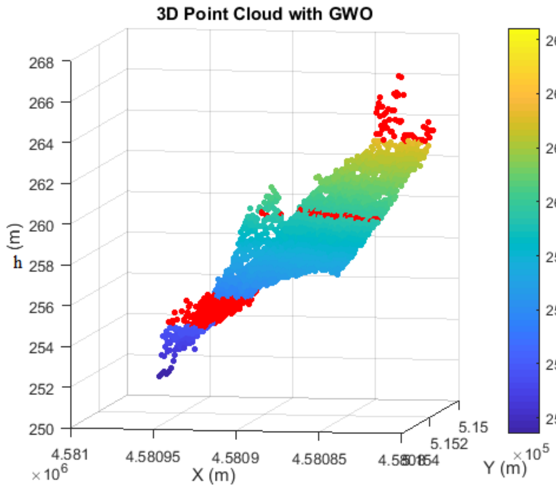

Figure 7. Outlier points of the GWO Method

The figure illustrates the classification of the 3D point cloud obtained from the study area using the GWO algorithm. While the majority of the points exhibit a regular distribution along the surface, the points marked in red were identified as. When the points identified as outliers by the GWO algorithm were examined, it was observed that they did not exhibit a random distribution; rather, they were concentrated in certain areas. An evaluation of the terrain conditions in these regions revealed the presence of slope variations. This suggests that the use of a second-degree polynomial approximation may not be adequate in these areas.
It is observed that these outliers are mostly concentrated in the upper and lower parts of the surface and in grouped cases. This pattern can be attributed to the hunting strategy of the GWO, which allows the algorithm to explore the solution space on a global scale. As a result, it can detect measurement anomalies more effectively, particularly in regions prone to extreme values. The GWO method classifies data by grouping them according to the algorithm's nature, with hunting behaviors serving as the main mechanism guiding this grouping.

6. Conclusion

LS method identified a total of 493 measurements as outliers, whereas the GWO classified 472 measurements as outliers. A comparative analysis revealed that 181 points were commonly detected by both methods. While the LS method relies on statistical thresholds derived from residual analysis, the GWO explores the solution space in a heuristic manner and may therefore capture additional anomalies that classical statistical techniques fail to recognize. The locations of the common points are as shown in Figure 8.

https://doi.org/10.5194/isprs-archives-XLVIII-4-W18-2025-201-2026

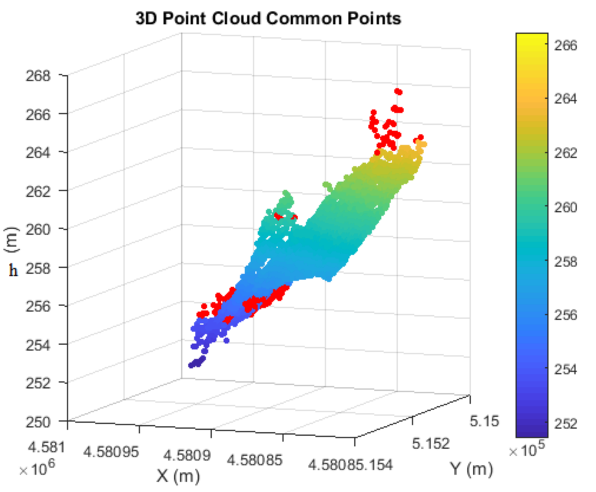

Figure 8. Common points

A comparison of the data classified as outliers by the GWO and LS methods reveals that approximately 38.35% of the points are commonly identified by both approaches. Following the elimination of outliers identified by both methods, the RMSE were calculated in order to assess the accuracy of the remaining dataset. The results indicate that the LS method achieved an RMSE (m_0) of 0.0193 m, while the GWO method yielded a slightly higher RMSE (m_0) value of 0.0427 m. This comparison highlights the difference in the sensitivity of the two approaches, with LS showing lower overall residual errors, whereas GWO provides a broader perspective by capturing clustered anomalies effectively.

When the outliers classified by the GWO method and those identified by the LS method are compared, it is observed that a greater number of common points are detected in the upper and lower parts of the study area. This spatial pattern can be attributed to the working principle of the GWO algorithm, which explores the solution space through heuristic search and therefore tends to identify measurement anomalies that cluster in specific regions.

It has also been observed that the GWO algorithm produces classification results by grouping. This demonstrates that the GWO algorithm can achieve high-accuracy classifications, especially in studies where outlier measurements are concentrated in one section.

7. Discussion

In this study, the performance of the conventional LS method and the GWO algorithm was compared for outlier detection in geodetic measurement data collected within the Ondokuz Mayıs University campus. The LS method identified 493 points as outliers, while the GWO classified 472 points as outliers. A total of 181 points (approximately 38.35%) were commonly detected by both methods, indicating limited compliance in identifying rough measurement errors. By applying the GWO algorithm to the collected measurement data, the aim is to detect and classify measurement errors. During the optimization process, the algorithm searches for the optimal solution space by imitating the hunting strategy of wolves, identifies faulty measurements, and removes them from the dataset.

However, when the spatial distribution of common outliers was examined, a grouped density of points was observed, especially in the upper and lower parts of the study area. This also demonstrates that the GWO algorithm can be effectively used to detect errors that are particularly concentrated in a specific region and that it may yield satisfactory results. This could be attributed to the general algorithm of the GWO method and the movements of wolves in response to prey. On the other hand, the LS method relies on statistical thresholds based on outliers and may be more sensitive to local variations.

While LS provides a deterministic and computationally efficient foundation, GWO offers advanced detection capabilities for complex datasets where anomalies may cluster or exhibit irregular behaviour. The results suggest that integrating metaheuristic algorithms like GWO into geodetic quality control procedures could improve the detection of outlier measurements and ultimately contribute to more accurate geoid modelling and spatial data analysis.

In conclusion, the comparative analysis demonstrates that GWO is a alternative to classical methods for outlier detection in geodetic applications. Its ability to detect additional anomalous points that may be missed by LS highlights the value of incorporating heuristic optimisation techniques into modern geographic data processing workflows.

References

Akar, A., Konakoğlu, B., & Akar, Ö. , 2022 , Prediction of geoid undulations: Random forest versus classic interpolation techniques. *Concurrency and Computation: Practice and Experience*, e7004.

Al Qaraghuli, H., Sheibani, R., & Tabatabaee, H. (2022). Detection of atrial fibrillation using variable length genetic algorithm and convolutional neural network. Concurrency and Computation: Practice and Experience, 34(10), e6789.

Arora, N., & Banyal, R. K. (2021). Workflow scheduling using particle swarm optimization and gray wolf optimization algorithm in cloud computing. Concurrency and Computation: Practice and Experience, 33(16), e6281.

Cao, M., Huang, M., Xu, R., Lü, G., & Chen, M., 2019, A grey wolf optimizer–cellular automata integrated model for urban growth simulation and optimization. Transactions in GIS, 23(4), 672-687.

Dasgupta, M., & Mishra, S. K., 2004, Least absolute deviation estimation of linear econometric models: A literature review. Available at SSRN 552502.

Gandomi, A. H., & Roke, D. A., 2014, Engineering optimization using interior search algorithm. Paper presented at the 2014 IEEE symposium on swarm intelligence.

Heinrichs, F., Bastian, P., & Dette, H., 2025 , Sequential Outlier Detection in Non-Stationary Time Series. arXiv preprint arXiv:2502.18038.

Karaboga, D., & Basturk, B. (2008). On the performance of artificial bee colony (ABC) algorithm. Applied soft computing, 8(1), 687-697.

Karakoyun, M. (2021). Çok amaçlı optimizasyon problemlerinin çözümü için Kurbağa Sıçrama ve Gri Kurt Optimizasyonu algoritmaları tabanlı hibrit bir yöntemin geliştirilmesi. (Doktora). Konya Teknik Üniversitesi,

Kumar, A., Pant, S., & Ram, M., 2017, System reliability optimization using gray wolf optimizer algorithm. Quality and Reliability Engineering International, 33(7), 1327-1335.

Le Goïc, G., Brown, C., Favreliere, H., Samper, S., & Formosa, F. (2012). Outlier filtering: a new method for improving the quality of surface measurements. Measurement Science and Technology, 24(1), 015001.

Lehmann, R., 2013., 3 σ-rule for outlier detection from the viewpoint of geodetic adjustment. Journal of Surveying Engineering, 139(4), 157-165.

Maniezzo, V., Gambardella, L. M., & Luigi, F. d. (2004). Ant colony optimization. In New optimization techniques in engineering (pp. 101-121): Springer.

Mirjalili, S., 2015,. How effective is the Grey Wolf optimizer in training multi-layer perceptrons. Applied Intelligence, 43, 150-161.

Mirjalili, S., Mirjalili, S. M., & Lewis, A., 2014, Grey wolf optimizer. Advances in Engineering Software, 69, 46-61.

Montgomery, D. C., Peck, E. A., & Vining, G. G., 2021. Introduction to linear regression analysis: John Wiley & Sons.

Nikolaev, A. G., & Jacobson, S. H. (2010). Simulated annealing. In Handbook of metaheuristics (pp. 1-39). Springer, Boston, MA.

Ogundare, J. O., 2018, Understanding least squares estimation and geomatics data analysis: John Wiley & Sons.

Sisman, Y., 2014 , Coordinate transformation of cadastral maps using different adjustment methods. Journal of the Chinese Institute of Engineers, 37(7), 869-882.

Teke, K., Yalçinkaya M., 2005, Yerel Jeoid Yüzeyinin Belirlenmesinde Kullanılan Enterpolasyon Yöntemleri , TUJK Düşey Datum Çalıştayı, Karadeniz Teknik Üniveristesi, Trabzon.

Tukka, A. A., Tata, H., & Idowu, O. T. 2025. A Computational Tool for Local Gravimetric Geoid Determination Using Least Squares Collocation. Journal of Spatial Information Sciences, 2(1), 251-274.

Wolf, H., Ghilani, C.D., 1997 , Adjustment Computation: Statistics and Least Squares in Surveying and GIS, John Wiley and Sons Inc., New York, 1997.

Yang, Y., Ni, W., Sun, Q., Wen, H., & Teng, Z. 2013, Improved Cole parameter extraction based on the least absolute deviation method. Physiological measurement, 34(10), 1239.

Yang, X. S., 2011, Metaheuristic optimization. Scholarpedia, 6(8), 11472.

Yang, X.-S., & Slowik, A., 2020, Firefly algorithm. In Swarm intelligence algorithms (pp. 163-174): CRC Press.