DOI: 10.54254/3029-0880/1/2023002

Research analysis of data-driven submarine navigation adaptation zones based on factor analysis

Chen Qiaojun [1,*], Guang Jiahe [2,5], Zhang Lixing [3,6], Zhang Yiyang [4,7], Liu Sitong [5,8],

[1] School of Safety Science and Engineering, Liaoning Technical University, 125000
[2] School of Electronic and Information Engineering, Liaoning Technical University, 125000
[3] School of Software, Liaoning Technical University, 125000
[4] School of Business Administration, Liaoning Technical University, 125000

[5] guangjiahe@qq.com
[6] 1213269775@qq.com
[7] 2215852270@qq.com
[8] 874830061@qq.com

* cqj0925@163.com

Abstract. The study of gravity anomaly adaptation zones holds significant importance in resource exploration, environmental studies, and marine development. This paper initially collected and preprocessed data on gravity anomalies, dividing them into 16 sub-regions. Subsequently, five statistical parameters of gravity field characteristics—standard deviation, roughness, gradient, gravity entropy, and planar correlation—were introduced, and factor analysis was employed to compute scores for each area. Finally, MATLAB software was utilized to construct a three-dimensional distribution map of gravity anomaly values. The results indicate that factor analysis demonstrates promising applications in the research of data-driven submarine navigation adaptation zones.

Keywords: gravity anomalies, factor analysis, MATLAB

1. Introduction:

DOI: 10.54254/3029-0880/1/2023002

The field of Earth sciences has long been dedicated to understanding the Earth's internal structure and crustal dynamic processes. The Earth's gravity field serves as a critical data source for studying subsurface material distribution and crustal deformation. However, the Earth's gravity field is influenced by the heterogeneity of subsurface media, resulting in observed gravity anomalies on the Earth's surface. To accurately interpret and understand these anomalies, researchers need to develop new techniques and methods to determine adaptation zones in a data-driven manner.

Data-driven studies of gravity anomaly [1] adaptation zones can assist geologists in better comprehending subsurface geological structures, such as rock types and underground fault zones. This holds significant importance in resource exploration, such as locating and assessing mineral resources, oil and gas reservoirs. Furthermore, research on gravity anomaly adaptation zones can provide information regarding crustal deformation and precursors to seismic activity. This aids in early warning systems for earthquakes and volcanic activity, thereby reducing disaster risks. Understanding subsurface geological conditions is crucial in urban planning and infrastructure development to ensure the safety and stability of buildings and roads. Research on gravity anomaly adaptation zones enables planners to better understand subsurface conditions, leading to more precise planning and design. Research on gravity anomaly adaptation zones contributes to advancing the forefront of Earth sciences, enhancing our understanding of the Earth's internal structure and crustal evolution. This holds significant importance for theoretical research and academic progress.

In summary, data-driven research on gravity anomaly adaptation zones holds importance in multiple fields, including geology, geophysics, resource exploration, seismology, and environmental science. It provides a robust tool for us to better understand the Earth's internal structure and crustal dynamics, thereby aiding in improving resource management, disaster warning, and infrastructure planning. Hence, this paper conducts an in-depth study of gravity anomaly adaptation zones based on factor analysis.

2. Data Collection and Preprocessing

2.1. Data Collection

This paper initially collected data on gravity anomaly values, and the obtained results are as follows:

Table 1: Data on gravity anomaly values

Longitude	Latitude	Gravity Anomaly Value	Longitude	Latitude	Gravity Anomaly Value
115.0083	11.0068	59.3	116.525	9.8594	19.2
115.025	11.0068	58.1	116.5417	9.8594	15
115.0417	11.0068	52.5	116.5583	9.8594	10.9
115.0583	11.0068	45.5	116.575	9.8594	7.5
115.075	11.0068	38.4	116.5917	9.8594	5.3

116.4417	9.8594	35.5	116.9417	9.0045	-39.1
116.4583	9.8594	34	116.9583	9.0045	-36
116.475	9.8594	30.8	116.975	9.0045	-30.9
116.4917	9.8594	27	116.9917	9.0045	-24.5
116.5083	9.8594	23.2	117.0083	9.0045	-17.6

Note: Data sourced from (http://topex.ucsd.edu/)

2.2. Data Preprocessing

Data cleaning in this study primarily involved the following steps:

STEP 1: Inspection of the original data.

STEP 2: Handling missing values and outliers.

STEP 3: Removal of irrelevant variables.

STEP 4: Division of data into regional segments.

Upon conducting the aforementioned STEPs 1 to 3, it was observed that the provided data did not contain any missing values or outliers. Hence, the data was segmented regionally based on the following ranges: Longitude Range: [115.0083, 117.0083], Latitude Range: [9.0045, 11.0068]. The data was grid-based and had a resolution of [resolution]. Consequently, the data was divided into regional sections, the results of which are presented in Table 2:

Table 2

Zone 1		Zone 2		···	Zone 15		Zone 16	
115.008	11.0068	116.375	10.8922	···	115.475	9.1197	115.542	9.1197
115.025	11.0068	116.392	10.8922	···	115.492	9.1197	115.558	9.1197
115.042	11.0068	116.408	10.8922	···	115.508	9.1197	115.575	9.1197
115.058	11.0068	116.425	10.8922	···	115.525	9.1197	115.592	9.1197
115.075	11.0068	116.442	10.8922	···	116.242	9.2513	115.608	9.1197
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
116.292	10.8922	115.642	10.7613	···	115.458	9.1197	116.942	9.0045
116.308	10.8922	115.658	10.7613	···	115.475	9.1197	116.958	9.0045
116.325	10.8922	115.675	10.7613	···	115.492	9.1197	116.975	9.0045
116.342	10.8922	115.692	10.7613	···	115.508	9.1197	116.992	9.0045
116.358	10.8922	115.708	10.7613	···	115.525	9.1197	117.008	9.0045

3. Statistical Parameters of Gravity Field Characteristics

DOI: 10.54254/3029-0880/1/2023002

Various characteristics of the gravity field describe its changes from different perspectives, serving as essential parameters for selecting adaptation zones. The gravity field data used regional gravity anomaly values, and the following analysis was based on standard grid gravity anomaly values [2].

3.1. Standard Deviation

Standard deviation is a crucial indicator of gravity anomalies and is defined as follows:

$$\delta^2 = Var[f(i,j)] \tag{1}$$

where $f(i,j)$ is the grid coordinate (i,j)

Longitude and latitude standard deviations δ_x reflect the fluctuation degrees of gravity anomalies in the respective directions.

$$\delta_x = \sum_{i=1}^{M} \frac{S_i^2}{M} \tag{2}$$

In the equation, M represents the number of discrete points in the longitude direction for gravity anomaly values.

The latitude standard deviation δ_y reflects the fluctuation dimension of gravity anomalies in the latitude direction. Its definition is as follows:

$$\delta_y{}^2 = \sum_{j=1}^{N} \frac{S_j^2}{N} \tag{3}$$

N denotes the number of discrete points in the latitude direction. While

$$S_i^2 = Var[f(i,:)], S_j^2 = Var[f(:j)] \tag{4}$$

3.2. Roughness

The roughness of gravity anomaly values in both longitude and latitude directions can be defined as

$$\begin{cases} \delta_x{}^2 = \sum_{i=1}^{M} \frac{M_s\left(D_i ff[f(i,:)]\right)}{M} \\ \delta_y{}^2 = \sum_{j=1}^{M} \frac{M_s\left(D_i ff[f(:j)]\right)}{N} \end{cases} \tag{5}$$

The absolute roughness of gravity anomalies can be defined as $\delta^2 = \frac{(\delta_x{}^2 + \delta_y{}^2)}{2}$. Here, M_s means square value function and $Diff$ represents the difference function.

DOI: 10.54254/3029-0880/1/2023002

3.3. Gradient

The gravity anomaly gradient reflects the extent of changes in gravity anomaly values concerning spatial positions. Generally, the larger the gradient of gravity anomaly values, the richer the spatial information regarding gravity characteristics changes. [3]

The gradients in the longitude and latitude directions for gravity anomalies can be defined separately.

$$\begin{cases} G_x^{\,2} = \sum_{i=1}^{M} \dfrac{M_s[\dfrac{Diff(i,:)}{D_x}]}{M} \\ G_y^{\,2} = \sum_{i=1}^{M} \dfrac{M_s[\dfrac{Diff(:,j)}{D_y}]}{M} \end{cases} \tag{6}$$

The gradient of gravity anomaly values can be defined as

$$G^2 = \frac{(G_x^{\,2} + G_y^{\,2})}{2} \tag{7}$$

where M_s represents the mean square value function, $Diff$ represents the difference function, and D_x and D_y denote the longitude and latitude resolutions, respectively.

3.4. Gravity Entropy

Gravity Entropy is a concept commonly used to describe the distribution or clustering degree of geographical spatial data. [4] It is a metric used to measure the uniformity or concentration of data distribution. The definition of gravity entropy is

$$H = -\sum_{i=1}^{M} D_i \lg P_i \tag{8}$$

Here, P_i represents the probability of a specific gravity anomaly value occurrence. And $P_i = \dfrac{n_i}{\sum n_i}$

The longitude and latitude entropy for gravity are defined separately as

DOI: 10.54254/3029-0880/1/2023002

$$\begin{cases} H_x = \sum_{i=1}^{M} \frac{H_i}{M} \\ H_y = \sum_{i=1}^{M} \frac{H_j}{N} \end{cases} \quad (9)$$

where H_i and H_j represent the gravity entropy for each row and column, respectively.

3.5. Plane Correlation

Gravity anomaly plane correlation reflects the correlation of local gravity field features concerning changes in spatial positions of gravity anomaly values. Typically, a larger plane correlation indicates a more pronounced change in gravity field characteristics concerning spatial positions.

The definition of gravity anomaly plane longitude and latitude is as follows:

$$\begin{cases} RD_x^{\,2} = \frac{\sum_{i=1}^{K} R_i^{\,2} / K}{D_x^{\,2}} \\ RD_y^{\,2} = \frac{\sum_{j=1}^{K} R_j^{\,2} / K}{D_y^{\,2}} \end{cases} \quad (10)$$

And the definition of gravity anomaly plane correlation can be expressed as:

$$RP^2 = \frac{\left(RP_x^{\,2} + RP_y^{\,2}\right)}{2} \quad (11)$$

In the above equation:

$$\begin{cases} R_i^{\,2} = Var(f(\lambda,\varphi) - f(\lambda + D_x, \varphi)) \\ R_j^{\,2} = Var(f(\lambda,\varphi) - f(\lambda, D_x + \varphi)) \end{cases}$$

4. Selection of Gravity Adaptation Zones Based on Factors

4.1. Factor Analysis Process

Factor Analysis is a multivariate statistical analysis method used to understand and describe the correlations and structures among variables in observed data. Its main purpose is to reduce the dimensions of multiple observed variables (also known as indicators or items) to identify latent factors or dimensions. These factors are linear combinations of observed variables and can explain the variance-covariance structure of the data. [5] Let the eigenvalue sample data matrix be:

DOI: 10.54254/3029-0880/1/2023002

$$X = \begin{bmatrix} X_{11}, X_{12}, \cdots, X_{1p} \\ X_{21}, X_{22}, \cdots, X_{2p} \\ \vdots \quad \vdots \quad \vdots \quad \vdots \\ X_{n1}, X_{n2}, \cdots, X_{np} \end{bmatrix} \tag{12}$$

The subsequent step is to establish the factor model of the text: $X = \mu + AF + \varepsilon$

Here, $F = (F_1, \cdots F_m)^T$ represents the common factors characterizing all statistical features X of the gravity field, $\varepsilon = (\varepsilon_1, \cdots \varepsilon_p)^T$ represents the features of X, followed by calculating the sample mean, sample covariance matrix, and sample correlation matrix from the sample data matrix X.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_{(i)} = (\bar{X}_1, \bar{X}_2, \cdots \bar{X}_p)' \tag{13}$$

The sample covariance matrix S is defined as:

$$S = \frac{1}{n-1}\sum_{i=1}^{n}(X_{(i)} - \bar{X})(X_{(i)} - \bar{X})' = (S_{ij}) \tag{14}$$

For the sample correlation matrix

$$R = (r_{ij}), \quad r_{ij} = \frac{S_{ij}}{\sqrt{S_{ii}S_{jj}}}(i, j = 1, 2 \cdots p) \tag{15}$$

Compute the eigenvalues and standardized eigenvectors, denoted as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{10} \geq 0$ for eigenvalues and the corresponding unit positive vectors $l_1, l_2 \cdots l_p$.

Then, solve for the factor loading matrix A of the factor model, selecting the desired factors $\frac{\lambda_1 + \lambda_2 + \cdots + \lambda_m}{p} \geq 1$. Let $q_i = \sqrt{\lambda_i} l_i (i = 1, 2 \cdots m)$, ($A = (a_1, \cdots a_m)$) represent the factor loading matrix. Finally, after obtaining the factor loading matrix, derive the determinant of $X = (X_1, X_2 \cdots X_p)^T$.

4.2. Selection of Adaptation Zones Using Factor Analysis

To visualize the features reflected by the aforementioned parameters of the gravity field, MATLAB software is utilized to plot the distribution of gravity anomaly values and the subdivision of sub-regions, as shown in Figure 1.

DOI: 10.54254/3029-0880/1/2023002

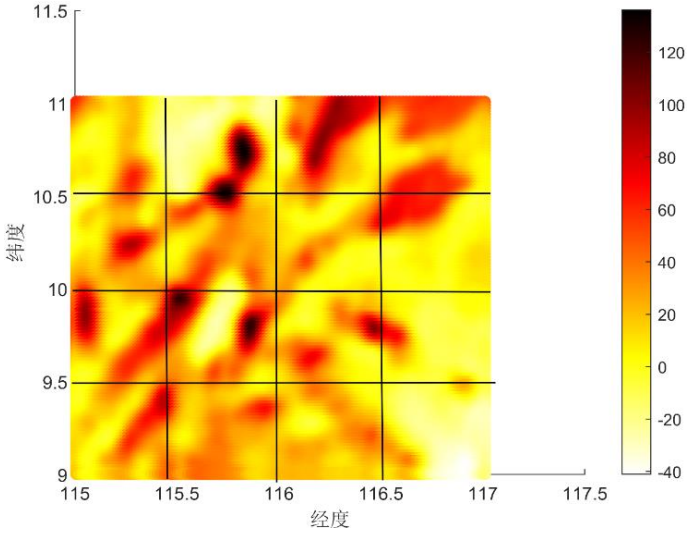

Figure 1. Gravity Anomaly Distribution

Based on the analysis of the above figure, 16 sub-regions were identified. Consequently, adaptive zones were selected based on various characteristics of the gravity regions.

4.3. Result Analysis

Calculation was performed for each of the gravity field characteristic parameters mentioned above. Considering the unit limitations of the characteristics, they were normalized to facilitate subsequent statistical analyses.

Under the aforementioned factor analysis, the results are presented in Table 2:

Table 3: Factor Analysis Results

Component	Extracted Load Sum			Rotated Load Sum		
	Total	Cumulative Contribution Rate	Cumulative Contribution Rate	Total	Cumulative Contribution Rate	Cumulative Contribution Rate
1	11.02	60.20%	60.20%	9.21	48.91%	48.91%
2	4.57	25.11%	85.12%	6.33	34.22%	84.60%
3	1.83	9.20%	91.62%	1.75	10.83%	91.62%

According to the analysis, these three factors contributed 48.91%, 34.22%, and 10.83%, respectively, to all variables. The cumulative contribution rate reached 91.62%, aligning with practical calculation requirements. Calculations were performed for several sub-regions, eventually obtaining factor scores for 16 sub-regions and their total scores. A comparative curve chart was generated as follows:

DOI: 10.54254/3029-0880/1/2023002

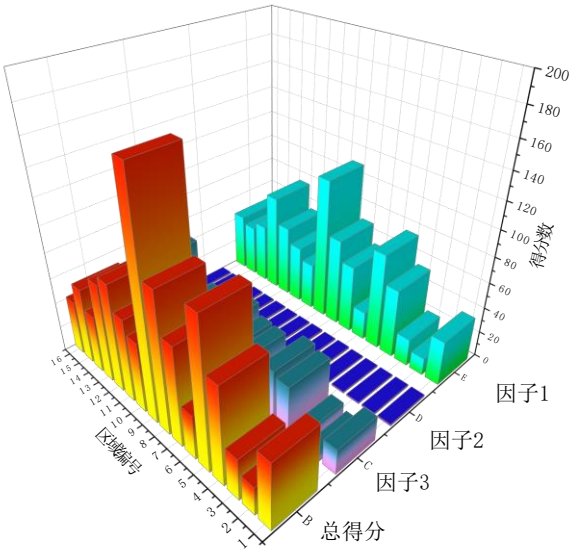

Figure 2. Factor Scores and Total Scores

Higher scores represent higher adaptability in the region, while lower scores indicate lower feasibility for the region as an adaptive zone. Thus, the prioritization of adaptation was determined. Among the 16 sub-regions, sub-regions 5, 6, 7, and 8 exhibited the highest adaptability, sub-regions 1, 4, 12, 13, and 15 showed moderate adaptability, while the remaining sub-regions (2, 3, 6, 10, 11, and 16) demonstrated poor adaptability, resulting in less accurate positioning. MATLAB was employed to create gravity anomaly distribution charts for all sub-regions, as depicted in Figure 3, visualizing the results.

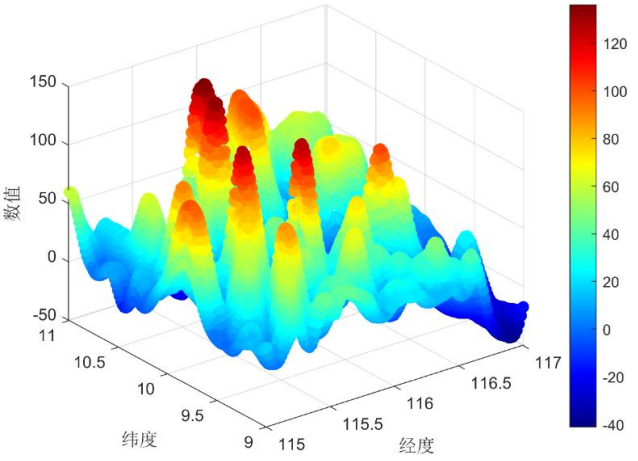

Figure 3 Gravity Anomaly Distribution

5. Conclusion

This paper, concerning the collected data, introduced five statistical parameters for gravity field characteristics. Employing factor analysis facilitated the derivation of factor scores for 16 partitioned sub-regions and vividly presented diverse gravity anomaly distribution charts for different sub-regions. This provides an essential scientific basis for investigating data-driven underwater navigation adaptive zones.

References

[1] Wang H., Wan X., Richard F. A. Prediction of the South China Sea Seabed Topography Based on Convolutional Neural Networks. Journal of Geodesy and Geodynamics, 1-11 [November 5, 2023].

[2] Gong J., Zhang C., Zhou X. A Method for Gravity Adaptive Zone Selection Based on Factor Analysis. Journal of Chinese Inertial Technology, 27(06):732-737 [2019].

[3] Yang F., Shen R., Mei S. Inversion of Emperor Mountain Sea Area Seabed Topography using Joint Gravity Anomaly and Gravity Vertical Gradient Anomaly Data. Acta Oceanologica Sinica, 44(12):126-135 [2022].

[4] Hu P., Liu J., Zhou X. A New Method for Gravity Field Positioning Based on Eigenvalues. Ship Electronic Engineering, 35(10):58-61 [2015].

[5] Fu J., Gui S., Yi Q. Source Tracing and Governance Strategies for Soluble Organic Matter in a River in Nanchang City based on Three-Dimensional Fluorescence and Parallel Factor Analysis. Journal of Environmental Engineering, 1-12 [November 5, 2023].