Advances in Engineering Research : Possibilities and Challenges
ISSN Print: 3079-5192
ISSN Online: 3079-5206
https://doi.org/10.63313/AERpc.9046

Evaluation and Suggestions on Logistics Level of Yangtze River Delta Urban Agglomeration Based on Factor Analysis and Cluster Analysis

Liu Yang*. Zhongying He
School of Economics and Management, Southwest Petroleum University, Chengdu, China
Email: 19182532570@163.com

How to cite this paper: Liu, Y., & He, Z. Y. (2025). Evaluation and Suggestions on Logistics Level of Yangtze River Delta Urban Agglomeration Based on Factor Analysis and Cluster Analysis. Advances in Engineering Research :Possibilities and Challenges, 2(2), 26-36. ISSN Print: 3079-5192; ISSN Online: 3079-5206.
https://doi.org/10.63313/AERpc.9046
Published: 2025-08-18

Abstract

As the "neural network" of economic integration, the development level of the logistics system determines the efficiency of cross-regional flow and allocation of factor resources. Currently, although the economic aggregate of the Yangtze River Delta region accounts for nearly one-fourth of the national total, the imbalance in logistics development has become a key bottleneck restricting in-depth regional collaboration. To comprehensively evaluate the logistics development level of the Yangtze River Delta urban agglomeration, this paper constructs an evaluation system for regional logistics level of the Yangtze River Delta urban agglomeration by selecting 9 indicators from 4 dimensions. The comprehensive scores of each city are calculated through factor analysis. The results show that Shanghai, Wuxi, and Nanjing have the highest logistics development level, followed by Jiaxing, Ningbo, and Nantong. Based on the scores of each common factor and the comprehensive scores obtained from the above factor analysis, a cluster analysis can be conducted on the regional logistics level of the cities in the Yangtze River Delta. According to logistics scale, openness, logistics network density, and industrial contribution, the 14 cities are divided into three categories. By analyzing and identifying the advantages and disadvantages of urban logistics development, revealing the spatial differentiation law of logistics development in the Yangtze River Delta, and proposing differentiated countermeasures, this paper puts forward policy suggestions for promoting the integration of the Yangtze River Delta region.

Keywords

Yangtze River Delta Urban Agglomeration; Logistics Level Evaluation; Factor Analysis; Cluster Analysis; Regional Collaboration

1. Introduction

The Yangtze River Delta Urban Agglomeration refers to a cluster of closely connected cities with Shanghai as the core, distributed within the administrative regions of Shanghai, Jiangsu Province, Zhejiang Province, and Anhui Province. These cities are mainly located in the optimized development and key development zones under the national "two horizontal and three vertical" urbanization pattern. In 2020, the Central Committee of the Communist Party of China and the State Council issued the Outline of the Development Plan for the Integration of the Yangtze River Delta Region, which clearly proposed strengthening the construction of a collaborative industrial system in the Yangtze River Delta, promoting the formation of a new pattern of regional coordinated development, and improving the modern logistics infrastructure network [1]. In recent years, the country has attached great importance to the high-quality development of the logistics industry. As a crucial engine for China's economic development and a key hub for opening up to the outside world, the efficient operation of the logistics system in the Yangtze River Delta Urban Agglomeration is vital to the regional integration strategy.

Regarding the logistics evaluation of the Yangtze River Delta Urban Agglomeration, domestic research results are mainly reflected in the interpretation of policies and the selection of methods for measuring logistics levels. On the one hand, for logistics research under the background of Yangtze River Delta integration, Ding Yijing (2025) pointed out that the logistics industry in the Yangtze River Delta region is a key factor driving China's modernization process, and its development plays a significant role in multiple fields [2]. On the other hand, in terms of the construction and measurement of logistics development systems, Shen Yufang (2011) and others used statistical analysis and GIS-based methods such as the location Gini coefficient and location entropy to analyze the evolution process of regional logistics in the Yangtze River Delta. They found that the current spatial organization of regional logistics in the Yangtze River Delta presents a central hierarchical diffusion pattern, and the trend of regional logistics spatial agglomeration is becoming stronger [3]. Based on the entropy weight-fuzzy matter-element model, Zhang Jinyi (2024) evaluated and analyzed the green logistics development level of each province and city in the Yangtze River Delta region from both temporal and spatial dimensions, and used the obstacle degree model to identify the main obstacle factors affecting the regional green logistics development level [4]. Yang Chongqin et al. (2024) evaluated the internal system of the logistics industry in the Yangtze River Delta based on the entropy weight-coupling coordination degree model, and concluded that the comprehensive index and coordination degree of the logistics industry in the Yangtze River Delta have increased year by year, but the coupling coordination degree is not high [5]. Building on the above research results, this study improves the method system for urban agglomeration logistics evaluation and verifies the applicability of factor analysis-cluster analysis in regional logistics research.

2. Construction of Logistics Level Evaluation Index System for the Yangtze River Delta Urban Agglomeration

2.1. Evaluation Dimension Division and Indicator Selection

To objectively measure the comprehensive logistics level of cities in the Yangtze River Delta, it is necessary to consider both the construction of regional logistics infrastructure and logistics operation capabilities, and also not ignore the economic achievements of logistics development. Based on an extensive review of relevant literature, combined with the economic conditions and social logistics development level of the Yangtze River Delta region, and with reference to the research ideas of outstanding scholars and existing statistical data, this study selects 14 cities as statistical samples, including Nanjing, Changzhou, Nantong, Wuxi, Ningbo, Hangzhou, Shaoxing, Jiaxing, Huzhou, Wuhu, Ma'anshan, Tongling, Anqing, and Shanghai. A first-level index system is constructed from four dimensions: infrastructure, operation capability, economic contribution, and information openness. Following the principles of data availability and operability, 9 factors are selected as secondary indicators, as shown in Table 1. These indicators include the number of civil vehicles owned, freight volume, freight turnover, port cargo throughput, postal and telecommunications business revenue, the proportion of the logistics industry in GDP, total import and export volume, and the number of mobile phone users. The data are sourced from public data of official platforms such as the 2024 National Statistical Yearbook, 2024 Statistical Yearbooks of Various Cities, and statistical bulletins on national economic and social development.

Table 1. Logistics Performance Evaluation Indicator System

Dimension	Evaluation Indicators		
	Specific Indicators	Unit	Variable Symbol
Infrastructure	Highway Density	km/km^2	X_1
	Number of Civil Vehicles Owned	10,000 tons	X_2
Operational Capacity	Freight Volume	10,000 tons/km	X_3
	Freight Turnover Volume	10,000 tons	X_4
	Port Cargo Throughput	10,000 tons	X_5
Economic Contribution	Postal and Telecommunications Business Revenue	100 million yuan	X_6
	Proportion of Logistics Industry in GDP	—	X_7
	Total Import and Export Volume	100 million yuan	X_8
Information Openness	Number of Mobile Phone Users	10,000 users	X_9

2.2. Main

Due to the differences in dimensions among logistics evaluation indicators, direct calculation would lead to indicators with larger dimensions dominating the analysis results. Therefore, data standardization is required to eliminate the impact of dimensional differences. This study adopts the Z-score method to standardize the original data, with the formula as follows:

$$\mathbf{Z_{ij}} = \frac{\mathbf{x_{ij} - \mu_i}}{\mathbf{\sigma_j}} \tag{1}$$

Among them, Xij represents the original value of the i-th city on the j-th indicator, $\bar{X}$j is the mean value of the j-th indicator, Sj is the standard deviation of the j-th indicator, and Zij denotes the standardized data.

For missing values, multiple imputation (for missing proportions > 5%) and mean imputation within the same region (for missing proportions ≤ 5%) are used for filling.

3. Comprehensive Evaluation of Logistics Level Based on Factor Analysis

As a complex system with multiple dimensions and indicators, the logistics system of the Yangtze River Delta urban agglomeration requires solving the problems of high correlation and information redundancy among indicators in its evaluation. Therefore, the application of factor analysis to process data has a solid scientific basis and reliability.3.1. Abbreviations and Acronyms

3.1. Suitability Test

Based on the 9 indicators identified in this paper, the KMO test and Bartlett's test of sphericity were conducted on the standardized data using SPSS 27.0 software, and the results are shown in Table 2. A larger KMO value indicates a stronger correlation between variables; meanwhile, factor analysis is only suitable when the KMO value is greater than 0.6. When the significance level of Bartlett's test of sphericity is less than 0.05, it indicates that there is a correlation between variables. According to the test results, the KMO test statistic is 0.627, which is greater than 0.6, so factor analysis is suitable based on the KMO measurement standard. In Bartlett's test, the significance value is 0.000, which is less than 0.05, indicating that there is a certain correlation among various factors, and thus the factor analysis is effective and reasonable.

Table 2. KMO and Bartlett's Test

KMO Sampling Adequacy Measure		0.627
Bartlett's Test of Sphericity	Approx. Chi - Square	203.668
	Degrees of Freedom	36
	Significance	0.000

3.2. Extraction and Interpretation of Common Factors

After systematic verification using factor analysis methods, it was found that the original 9 subdivided indicators do not require further refinement, which indicates that these indicators already have a clear hierarchical structure. Furthermore, the analysis was conducted based on the four main dimensions identified in this study. As shown in Table 3, the factor analysis scale successfully extracted 2 main components, with variance contribution rates of 68.244% and 17.541% respectively. The cumulative variance contribution rate reached 85.785%, exceeding the 60% benchmark. This result indicates that the effect of principal component extraction is relatively ideal, with less loss of data information, and can effectively explain most of the information in the data.

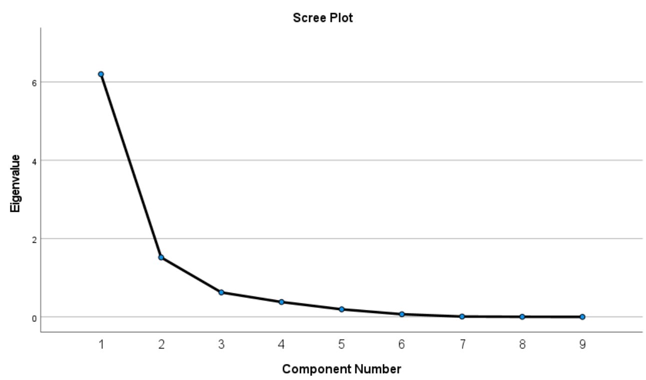
Table 3. Table of Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.201	68.902	68.902	6.201	68.902	68.902	6.142	68.244	68.244
2	1.519	16.883	85.785	1.519	16.883	85.785	1.579	17.541	85.785
3	0.626	6.951	92.736						
4	0.380	4.225	96.961						
5	0.193	2.149	99.110						
6	0.068	0.756	99.865						
7	0.009	0.101	99.966						
8	0.003	0.031	99.997						
9	0.000	0.003	100.000						

a. Extraction Method: Principal Component Analysis.

In addition, according to the scree plot in Figure 1, the line becomes flat starting from factor 3, while the lines for factors 1 and 2 are relatively steep. This indicates that the first two factors can cover most of the information among all indicators and can comprehensively reflect the logistics level of cities. Therefore, it is appropriate to extract two factors.

Figure 1. Scree Plot.



3.3. Factor Loadings and Naming Interpretation

To better explain the extracted common factors, the varimax method was used for factor rotation, and the rotated components are shown in Table 4.

Table 4. Rotated Component Matrix.

	Component 1	Component 2
Mobile Phone Users	0.997	-0.009
Freight Turnover Volume	0.972	0.111
Total Import and Export Volume	0.963	0.116
Postal and Telecommunications Business Revenue	0.957	0.075
Freight Volume	0.930	-0.058
Port Cargo Throughput	0.874	0.265
Number of Civil Vehicles Owned	0.853	-0.025
Highway Density	0.040	0.895
Proportion of Logistics Industry in GDP	0.059	0.820

Extraction Method: Principal Component Analysis. Rotation Method: Kaiser Normalization with Varimax Rotation.

a.Rotation converged after 3 iterations.

As can be seen from Table 4, Factor F1 has high loadings on indicators such as mobile phone users, freight turnover volume, total import and export volume, postal and telecommunications business revenue, freight volume, port cargo throughput, and number of civil vehicles owned. This factor comprehensively reflects the overall scale of regional logistics as well as logistics-related infrastructure and service capabilities, and can be named the Logistics Scale and Openness Factor.

Factor F2 has high loadings on highway density and the proportion of the logistics industry in GDP. Highway density directly reflects the network coverage of logistics infrastructure, while the proportion of the logistics industry in GDP characterizes the contribution of the logistics industry to the regional economy. These two indicators together reflect the penetration and economic value of the logistics system, and can be named the Logistics Network Density and Industrial Contribution Factor.

3.4. Factor Score Calculation and Ranking

The factor component score matrix obtained using SPSS 27.0 is shown in Table 5.

Table 5. Component Score Coefficient Matrix

	Component	Component
	1	2
Highway Density	0.997	-0.009
Number of Civil Vehicles Owned	0.972	0.111
Freight Volume	0.963	0.116
Freight Turnover Volume	0.957	0.075
Port Cargo Throughput	0.930	-0.058
Postal and Telecommunications Business Revenue	0.874	0.265
Proportion of Logistics Industry in GDP	0.853	-0.025
Total Import and Export Volume	0.040	0.895
Mobile Phone Users	0.059	0.820
Extraction Method: Principal Component Analysis. Rotation Method: Kaiser Normalization with Varimax Rotation. Component Scores.		

Based on Table 5, the standardized original data are substituted into the expression of common factors to calculate F1 and F2. Thus, a factor score model for evaluating the logistics capability of urban agglomerations in the Yangtze River Delta can be constructed.

$$F1=-0.043X_1+0.144X_2+0.159X_3+0.157X_4+0.132X_5+0.156X_6-0.036X_7+0.155X_8+0.167X_9$$

$$F2=0.581X_1-0.063X_2-0.089X_3+0.019X_4+0.124X_5-0.004X_6+0.531X_7+0.022X_8-0.061X_9$$

To accurately analyze the logistics regional level of the 11 cities, a weighted calculation is required to obtain a comprehensive score model, which is as follows:

$$F=\frac{0.68244}{0.85785}F_1+\frac{0.17541}{0.85785}F_2=0.796F_1+0.204F_2$$

On this basis, the individual factor score results are substituted to calculate the comprehensive score of logistics level for each sample city. Finally, based on the respective scores, the sample cities are ranked in terms of two aspects—logistics scale and openness, as well as the penetration and economic value of the logistics system—and their overall logistics level. The results are shown in Table 6.

Table 6. Ranking of Logistics Levels of 14 Cities in the Yangtze River Delta

City	F	Ranking	F1	F2
Nanjing	0.280	3	0.223	0.500
Changzhou	0.200	7	0.039	0.828
Nantong	0.207	6	0.129	0.511
Wuxi	0.289	2	0.184	0.701
Ningbo	0.237	5	0.225	0.284
Hangzhou	0.143	11	0.142	0.148
Shaoxing	0.103	13	0.079	0.199
Jiaxing	0.264	4	0.066	1.039
Huzhou	0.111	12	0.051	0.344
Wuhu	0.184	8	0.023	0.814
Ma'anshan	0.159	10	-0.026	0.883
Tongling	0.161	9	-0.052	0.991
Anqing	0.085	14	-0.005	0.436
Shanghai	0.876	1	0.900	0.780

3.5. Result Analysis

From the perspective of the comprehensive scores of each city, the logistics level of the Yangtze River Delta urban agglomeration presents an obvious "core-periphery" gradient distribution. As the core city of the Yangtze River Delta, Shanghai leads by an absolute advantage. Its international ports, high-end industrial chains and free trade zone policies have jointly promoted its leading position in logistics scale and openness, which confirms the positive correlation between economic strength and logistics level. Sub-central cities such as Nanjing, Wuxi and Ningbo form the second echelon relying on their industrial foundations and transportation hub status, but their F2 scores indicate that there is still room for improvement in their logistics network density and industrial contribution. Although cities in Anhui have relatively high F2 scores, their F1 scores are generally low, reflecting that the logistics scale is limited by the total economic volume.

Analyzing from the F1 dimension of logistics scale and service capacity, the gradient differences among cities are more significant. Shanghai takes the lead with a high score of 0.9, followed by regional central cities such as Nanjing and Ningbo. In contrast, cities like Tongling and Ma'anshan are obviously insufficient in terms of logistics scale. To a large extent, this difference reflects the gap in regional economic development level and degree of opening up to the outside world. Coastal port cities score higher by virtue of their natural location advantages, while provincial capitals occupy advantageous positions relying on their comprehensive service capabilities. However, some economically developed industrial cities such as Changzhou and Wuxi still have room for improvement in logistics specialization.

In addition, regarding logistics network density and industrial contribution, the distribution pattern among cities shows different characteristics. Cities like Jiaxing and Tongling perform prominently, while economically strong cities such as Hangzhou and Shaoxing are relatively backward. This unexpected contrast indicates that infrastructure investment is not completely positively correlated with economic strength. Further analysis reveals that cities with a high degree of logistics specialization often perform excellently in industrial contribution, while comprehensive cities have relatively balanced logistics industry contributions due to the diversification of economic structures. Some cities such as Ma'anshan show obvious characteristics of industrial dependence on logistics.

4. Research on Urban Agglomeration Classification Based on K-means Clustering

4.1. Clustering Method

As a classic partitioning clustering method, the core principle of the K-means algorithm is to divide n observation samples into k clusters through iterative calculation, such that each sample belongs to the cluster corresponding to the nearest mean (i.e., cluster center). Ultimately, it achieves the goal of maximizing intra-cluster similarity and inter-cluster dissimilarity.

To better evaluate the regional logistics level of the Yangtze River Delta urban agglomeration and propose targeted suggestions, on the basis of the aforementioned factor analysis method, SPSS 27 software was used for K-means clustering analysis. Here, principal component F1, principal component F2, and comprehensive score F were adopted as variables for analysis. Cities were classified into three levels: high, medium, and low, with the number of clusters set to 3. The logistics levels of the 14 cities in the Yangtze River Delta were thus divided into 3 categories with significant differences. The resulting clustering chart and clustering analysis results are shown in Figure 2 and Table 7, respectively.

Figure 2. Clustering Chart

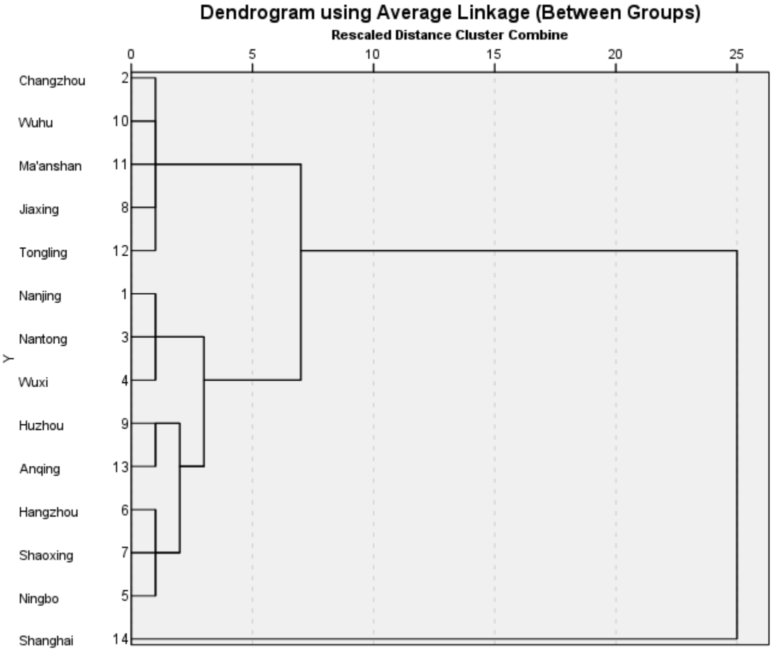


Table 7. Cluster Analysis Results

Category	Cities
Category 1	Shanghai
Category 2	Changzhou, Wuxi, Jiaxing, Wuhu, Ma'anshan, Tongling
Category 3	Nanjing, Nantong, Ningbo, Hangzhou, Shaoxing, Huzhou, Anqing

4.2. Analysis of Clustering Results

From the perspective of cluster center characteristics, the three categories of the Yangtze River Delta urban agglomeration present a stepped distribution pattern.

The first category includes only Shanghai. Its comprehensive score and scores of the two common factors are significantly higher than those of other cities, showing the "benchmarking" feature of all-round leadership. Its leading logistics development level stems from the superimposed effect of multiple advantages. Geographically, Shanghai is located at the estuary of the Yangtze River, with two international hubs —Yangshan Deep-Water Port and Pudong International Airport—forming a three-dimensional "sea-land-air" transportation network.

The second category includes 6 cities: Changzhou, Wuxi, Jiaxing, Wuhu, Ma'anshan, and Tongling. These cities perform prominently in F2 scores, showing a "facility-driven" development model. Geographically, they are all located along the Yangtze River or on the periphery of the Shanghai-Nanjing-Hangzhou core area. Wuhu and Ma'anshan has the golden waterway of the Yangtze River, while Wuxi and Changzhou lie on the Shanghai-Nanjing logistics corridor.

The third category includes 7 cities: Nanjing, Nantong, Ningbo, Hangzhou, Shaoxing, Huzhou, and Anqing. These cities belong to the "balanced development type," but their overall development level needs to be improved. The logistics development of the 7 third-category cities shows contradictory characteristics. In terms of economic aggregate, the GDP of both Nanjing and Hangzhou exceeds 1.5 trillion yuan, but their logistics industry faces the problem of being "large but not strong."

It can be seen that the logistics development of the Yangtze River Delta urban agglomeration has the problem of unbalanced regional development, mainly reflected in the following aspects: First, Shanghai stands out as a single pole, with a large gap from the last-ranked cities, showing a "core-periphery" differentiation, and cities in Anhui are generally lagging behind. Second, there is an imbalance in factor allocation: the second-category cities focus on infrastructure, while the third-category cities have a mismatch between provincial capital economy and logistics capacity. Third, there is a lack of collaborative mechanisms: administrative barriers have hindered the river-sea intermodal transportation model, Hangzhou's digital economy has not effectively radiated to the logistics field, and cross-provincial industrial linkage is insufficient.

5. Countermeasures and Suggestions for Improving Logistics Levels in the Yangtze River Delta Urban Agglomeration

In response to the current imbalance in logistics development across the Yangtze River Delta urban agglomeration, differentiated development strategies and systematic collaborative plans need to be adopted. For cities with high logistics levels such as Shanghai and Ningbo, the focus should be on strengthening their international hub functions. Relying on Yangshan Deep-Water Port and Pudong International Airport, they should develop high-end logistics services, build global supply chain management centers, and simultaneously promote policy innovations in free trade zones to enhance the facilitation of cross-border logistics.

For medium-level cities like Wuxi and Changzhou, the key is to optimize the multimodal transportation network, improve connecting facilities for road, railway, water, and air transportation, focus on resolving navigation bottlenecks such as the Nanjing Yangtze River Bridge, build a number of regional logistics hubs, strengthen the application of smart logistics technologies, and upgrade the level of logistics services.

Cities with lower logistics levels, such as those in northern Anhui, are in urgent need of addressing infrastructure shortcomings. They should accelerate the construction of expressway networks and dedicated railway lines, improve urban-rural distribution networks, actively undertake the spillover of logistics functions from core cities, and develop characteristic industrial logistics.

In terms of regional collaboration, a shared logistics information platform for the Yangtze River Delta should be co-established to promote data interconnection among customs, transportation, commerce, and other departments, and to build a unified logistics big data center. At the same time, a green distribution system should be constructed, promoting new energy logistics vehicles, piloting the application of hydrogen fuel cell heavy-duty trucks in port collection and distribution, deploying intelligent shared warehousing nodes in major urban agglomerations, and developing co-distribution models. It is suggested to establish an inter-provincial logistics collaborative development fund to provide financial support for key corridor construction and informatization projects, improve the regional interest compensation mechanism, and promote the optimal allocation of logistics resources.

6. Research Conclusions

This study systematically evaluated the logistics development level of the Yangtze River Delta urban agglomeration through factor analysis and cluster analysis. Factor analysis extracted two core dimensions: the Logistics Scale and Openness Factor (F1), which reflects cities' comprehensive strength in terms of freight volume, port throughput, and other indicators; and the Logistics Network Density and Industrial Contribution Factor (F2), which embodies the improvement of infrastructure and the level of industrial collaboration.

The clustering results show an obvious spatial differentiation pattern: the phenomenon of Shanghai standing out as a single core is prominent; cities along the Yangtze River are relatively leading in infrastructure construction; peripheral cities are generally lagging in development; and the mismatch between the logistics capacity and economic status of provincial capitals deserves attention. This differentiation pattern not only reflects differences in resource endowments and development stages among cities but also exposes the deep-seated contradiction of insufficient regional collaboration.

References

[1] Editorial Department. The Outline of the Yangtze River Delta Regional Integration Development Plan Released [J]. Paper and Paper Making, 2020, 39(01): 18.

[2] Ding Yijing. Research on the Impact of Logistics Industry Development in the Yangtze River Delta on the Process of Chinese-Style Modernization [N]. Economic Herald, 2025-05-19(007). DOI: 10.28420/n.cnki.njjdb.2025.000488.

[3] Shen Yufang, Wang Nengzhou, Ma Renfeng, et al. Research on Spatial Layout and Evolution Characteristics of Regional Logistics in the Yangtze River Delta [J]. Economic Geography, 2011, 31(04): 618-623. DOI: 10.15957/j.cnki.jjdl.2011.04.015.

[4] Zhang Jinyi. Research on Evaluation of Green Logistics Development Level and Influencing Factors in the Yangtze River Delta [D]. Anhui University of Science and Technology, 2024. DOI: 10.26918/d.cnki.ghngc.2024.001123.

[5] Yang Chongqin, Zhang Guangsheng. Research on Logistics Industry Development Evaluation Based on Entropy-Weighted Coupling Coordination Degree Model: A Case Study of the Yangtze River Delta [J]. Economic & Trade Update, 2024, 21(02): 37-40. DOI: 10.19463/j.cnki.sdjm.2024.02.029.

[6] Dai Ningfan, Jiang Hua. Research on Evaluation of Regional Logistics Level in Hebei Province Based on Factor Analysis and Cluster Analysis [J]. Logistics Technology, 2025, 48(08): 47-51. DOI: 10.13714/j.cnki.1002-3100.2025.08.010.