Traffic & Transportation

Road-Rail Intermodal Hubs Site Selection Based on Road Freight Demand Mining – A Case from Beijing-Tianjin-Hebei Region

Qichen OU[1], Mi GAN[2], Meitong AN[3], Yichen WANG[4]

Original Scientific Paper
Submitted: 20 Sep. 2023
Accepted: 4 Jan. 2024

[1] qichenou@my.swjtu.edu.cn, Department of Transportation and Logistics, Southwest Jiaotong University

[2] Corresponding author, migan@swjtu.edu.cn, Department of Transportation and Logistics, Southwest Jiaotong University

[3] Meitong@my.swjtu.edu.cn, Department of Transportation and Logistics, Southwest Jiaotong University

[4] wycw8@outlook.com, Department of Transportation and Logistics, Southwest Jiaotong University

Publisher:
Faculty of Transport and Traffic Sciences, University of Zagreb

ABSTRACT

This study introduces a holistic framework for optimising road-rail intermodal hub locations based on real regional freight data and railway station information. The primary objective is to enhance railway transportation capacity, thereby facilitating the development of a low-carbon transport system. Research begins by scrutinising the freight landscape in the region, focusing on transport volume, freight intensity, goods types and average delivery distances. Subsequently, data mining techniques, including DBSCAN clustering and frequent itemset mining, are employed to uncover freight demand hotspots across both spatial and temporal dimensions. Based on these findings, a mathematical model for hub location selection is constructed, along with criteria for goods categories suitable for rail transportation. Ultimately, using the Beijing-Tianjin-Hebei region as a case study, 12 road-rail intermodal hubs are identified, along with the main cargo types best suited for rail transport within their respective service areas. This transition is expected to result in an annual reduction of 470,000 tons of regional carbon emissions. The proposed method framework provides valuable guidance and practical insights for the optimisation of freight structures in various regions. Furthermore, it aligns with contemporary environmental and sustainability objectives, contributing to the broader goal of establishing low-carbon transport systems.

KEYWORDS
intermodal transportation; transportation site selection; carbon emission; data mining.

1. INTRODUCTION

Against the backdrop of global climate change and environmental protection, "carbon neutrality" and "carbon peaking" have become pressing issues on a global scale. In order to mitigate the serious impacts of climate change, governments and all sectors of society are actively seeking pathways and solutions to reduce carbon emissions. Road transportation, as one of the more serious sectors of carbon emissions, accounts for about 87% of the total carbon emissions from transportation and has become the main source of carbon emissions from transportation in China and the focus of emission reduction. The combined road and railway transport mode, which is jointly completed through short-barge transport and transhipment, can reduce the proportion of road transport, and give full play to the advantages of railway transport [1]. Consequently, it realises the optimisation and adjustment of the transport structure and contributes to the construction of a low-carbon transport system.

However, at present, China's railway stations are plagued by problems such as outdated facilities, low degree of standardisation and mutual competition for cargo sources. From the perspective of supply, most of the existing stations do not have the geographic and facility conditions to develop into regional intermodal hubs; from the perspective of demand, the unbalanced industrial structure and location characteristics of the region lead to an imbalance in demand for the type and quantity of goods. Therefore, integrating existing railway freight resources and planning the location and functional focus of road-rail intermodal hubs based on the characteristics of road transport is of great significance in optimising the structure of road and railway transport and reducing carbon emissions in the field of transportation.

The selection and planning of intermodal hubs represent one of the primary focal points in intermodal transportation research. This research entails pinpointing the geographical locations of hub facilities, designing the hub network, and defining the service areas and freight volume attributes within those service areas. These decisions are rooted in cost optimisation, service enhancement, and other objectives, which hold profound significance in bolstering railroad capacity, facilitating the efficient flow of goods, and fostering regional economic development. Currently, many studies on hub location problems primarily rely on a predefined set of alternative points to construct mathematical planning models for problem resolution based on the classical CAB, AP, and TR data sets [2]. These studies tend to concentrate on refining model conditions and algorithms, often without a comprehensive understanding of the intricacies of cargo transportation dynamics surrounding intermodal hubs. This limitation can hinder the development of rational and practical hub layouts and plans.

This study aims to rectify the situation by analysing the characteristics of freight hotspots by using real-world data. It adopts a data-driven combined with mathematical planning approach to identify potential hub points and further delineates the service scope and freight characteristics of these hubs, which provides a holistic set of solutions for the optimisation of intermodal transport hubs. Specifically, the study extracts freight data from a leading online trading platform in China's road trunk transportation market and takes the Beijing-Tianjin-Hebei region as an example. First of all, the collected data are used to portray the freight characteristics of Beijing-Tianjin-Hebei region; then, the machine learning method is used to realise the identification of road freight demand-intensive points, and then the intermodal hub site selection and allocation model is established on the basis of the existing railway stations selected by rule; finally, the service scope of each hub is obtained by solving the problem and the main types of goods served in the region are mined accordingly. Ultimately, the study aims to achieve rational layout and planning of intermodal hubs, develop tailored marketing strategies, attract road-to-rail freight sources, increase railway capacity, and contribute to low-carbon transportation systems.

The main contributions of this paper mainly include the following three points:

1) Using real online road freight data, the current road freight characteristics of the Beijing-Tianjin-Hebei region are portrayed through the flow of transportation, freight intensity, cargo type share, average transportation distance and carbon emission estimation.
2) A data mining method is proposed to determine the cargo aggregation area in the spatio-temporal range. Firstly, the aggregation area and centre point are determined by spatial clustering and inverse geographic coding, and then the frequent item set mining is used to find the aggregation centre in time dimension, to determine the long-term stable road freight intensive demand area.
3) Based on the road freight data and existing railway facilities, we constructed a site selection model for road-rail intermodal hubs, solved 12 intermodal transportation hubs, and determined the main types of goods suitable for railroad transportation within the service area of each intermodal hub, and estimated that the annual carbon emission could be reduced by about 470,000 tons through road-rail transfer.

2. LITERATURE REVIEW

Since its first introduction by O'Kelly in 1986 [3, 4], the hub location problem has been extensively discussed in the academic community. The objective of the classical hub location problem is to find the optimal hub locations by analysing the relationships between demand nodes and candidate hubs, in order to optimise the efficiency and service quality of the entire system [5].

It is important to note that the hub location problem is a complex combinatorial optimisation problem that requires analysing the relationships between multiple demand nodes and candidate hubs, subject to various constraints. Factors considered often include capacity constraints, allocation types, transportation network restrictions, and environmental impacts, among others. Therefore, appropriate mathematical modelling and optimisation algorithms are needed to solve this problem. Currently, the methods for hub location primarily include optimisation methods, network analysis methods, GIS and spatial analysis methods, and data-driven methods, etc. In the current research, the hub location problem is often modelled as an optimisation problem and solved using mathematical programming methods. Common optimisation techniques include linear programming [6], integer programming [7], multi-objective programming [8–11], and genetic algorithms [12]. In recent years, with the development of graph theory, network analysis methods have gained popularity in the hub location problem. This approach utilises graph theory and network analysis techniques to analyse the topological structure of transportation networks and relationships between nodes in order to determine the optimal hub locations [13]. Examples of network analysis methods include shortest path algorithms and their extensions such as the variable neighbourhood decomposition search method [14, 15], minimum spanning tree algorithms [16, 17], and network flow models. Geographic information systems (GIS) and spatial analysis techniques can be used to analyse and process geographical data in hub location problems, such as the spatial distribution of demand nodes and transportation networks [18, 19]. Data-driven approaches are also popular, utilising big data and machine learning techniques to analyse and predict the distribution of demand nodes and traffic flow [20] or extract spatiotemporal trajectories [21] to support hub location decision-making.

In typical research, the selection objective of hub locations is the primary focus, and multi-objective hub location studies are more in line with practical needs compared to single-objective approaches. Regarding the discussion on location objectives, Campbell in 1994 provided an answer by categorising hub location problems into four types based on their objectives: p-hub median, incapacitated hub location, p-hub centre, and hub covering problems [22]. Typically, the fundamental objective of hub location problems is to minimise or maximise one or multiple evaluation criteria, such as maximising service level [23], minimising total cost [12], optimising accessibility [24], and achieving balance and fairness. Meanwhile, considering the total carbon emissions of multimodal transportation networks during site selection is also an important issue to be considered [25]. In terms of research content, the hub location problem not only requires considering the optimal location of hub nodes but also the optimal allocation of demand nodes [26]. In most hub location studies, demand is assumed to be static, but this may not be wise as multimodal transportation network planning is a long-term process that needs to consider elastic demand [27]. The allocation of demand points typically includes single allocation and multiple allocation. Single allocation is suitable for situations where resources or hub capacities are limited, while multiple allocations can provide better service in different circumstances. Many scholars consider different allocation methods simultaneously in their research to achieve better results [28, 29].

In summary, existing research on road-rail multimodal transportation hubs mostly focuses on the construction and operation of mathematical models, often overlooking a deep understanding of freight demand and the selection of alternative points. This approach may lead to significant deviations between the models and the actual situations as they fail to capture the diversity and complexity of real freight transportation. Additionally, the selection of alternative points is a critical decision-making step that directly affects the construction and operational efficiency of transportation networks, but it has been rarely addressed in previous studies. In this context, this study aims to fill this gap by adopting a data-driven approach, utilising geographic information systems (GIS) and spatial analysis techniques to process geographical data and determine the spatial distribution of alternative hub nodes and demand nodes in the Beijing-Tianjin-Hebei region. At the same time, the actual conditions of freight transportation demand are also considered, analysing and predicting the distribution of demand nodes and traffic flow using big data and machine learning techniques to support hub location decision-making, thereby improving transportation efficiency and reducing carbon emissions.

3. METHOD AND DATA

3.1 Problem description

The intermodal transportation model designed in this paper is shown in *Figure 1*. Building an intermodal transportation network based on the principle of a hub-and-spoke network can effectively reduce logistics costs and improve transportation efficiency. The dispersed road freight small-volume cargo flow through the short-distance road transport concentrated on large-scale road-rail intermodal transport hub, and then use the railway for medium- and long-distance trunk transport, to make full use of the advantages of the various modes of transport, and to produce the economies of scale of transport.

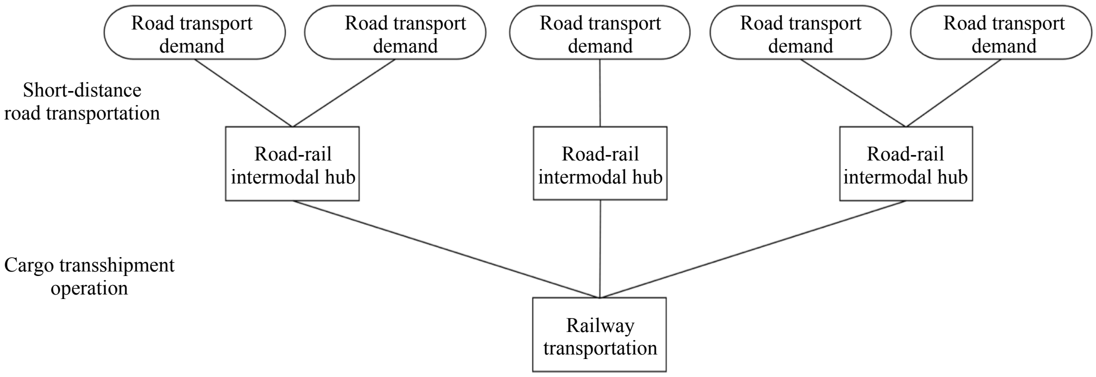


Figure 1 – Intermodal transportation model

The known conditions and problem objectives are described as follows: The latitude and longitude coordinates of the road transportation demand point and the alternative intermodal hub point, the demand of the road demand point, the fixed cost and the maximum capacity of the alternative intermodal hub point are known. The objective is to construct a site allocation model to select sites for the intermodal hub and to make decisions about the road transportation demand points served by each hub point in order to find the site allocation option that minimises the total cost (construction cost, transportation cost) and meets the regional freight demand.

3.2 The road-rail intermodal hub site selection model

The freight structure adjustment model design is shown in *Figure 2*.

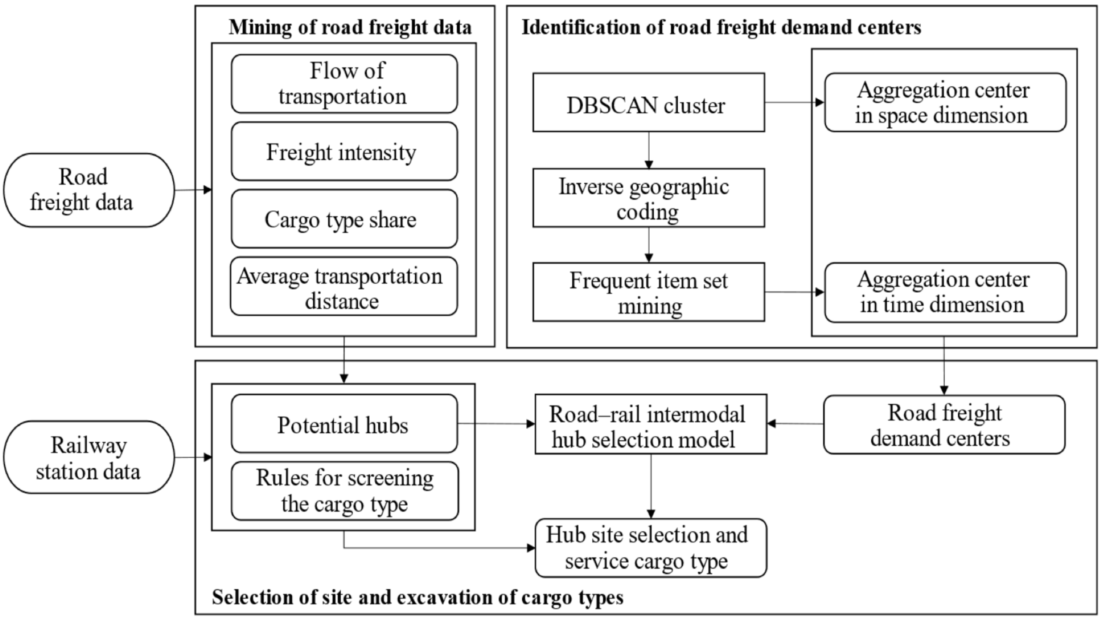


Figure 2 – The road-rail intermodal hub site selection model design

3.3 Identification of road freight demand centres

Identification of single-day road freight demand hotspot areas based on DBSCAN algorithm

DBSCAN (density-based spatial clustering of applications with noise) is a classical density-based spatial clustering algorithm. By defining the maximum set of density-connected points, the algorithm can divide a region with a certain density into a cluster and can find clusters of arbitrary shapes in noisy spatial data. The effect of the DBSCAN clustering algorithm is closely related to the scanning neighbourhood radius (eps) and the minimum number of included points in the neighbourhood (minPts), which need to be tuned according to the actual problem and clustering results of these two parameters. In the past, these parameters were determined based on the experience of the actual problem or by continuously adjusting the parameters to obtain the number of stable clusters, but the time cost of manual trial and error is high. To solve this problem, the cluster evaluation index "silhouette coefficient" was introduced in our model. By setting the step size of the input parameters in a certain range, the returned silhouette coefficients are constantly compared until the input parameter corresponding to the maximum silhouette coefficient is found, and the automatic selection of DBSCAN clustering parameters is realised. The silhouette coefficient of the defined samples is calculated as follows:

$$s_i = \frac{b_i - a_i}{\max\{a_1, b_1\}} \tag{1}$$

$$b_i = \min(b_{i1}, b_{i2}, \ldots, b_{ij}) \tag{2}$$

where a_1 is the average distance from sample i to other samples in the cluster, the smaller the value indicates that the greater the degree of similarity within the cluster, the greater the degree of sample i belongs to the cluster; b_i is the minimum average distance from sample i to all samples in other clusters, the smaller the value indicates that the greater the degree of similarity between the clusters, then sample i may belong to the other clusters; the range of silhouette coefficient s_i is (-1,1).

The $\overline{s}_i$ is averaged over all the sample silhouette coefficient within a cluster to obtain the silhouette coefficient score for that cluster. The closer $\overline{s}_i$ is to -1 the worse the clustering result is, and the closer $\overline{s}_i$ is to 1 the more reasonable the sample clustering result is. In addition, due to the error in the direct calculation of Euclidean distance by latitude and longitude, this paper uses the spherical distance formula (Haversine Formula) to calculate the distance between latitude and longitude.

Identification of road freight demand centres based on frequent item set mining

To ensure the sustained stability of freight demand points over an extended timeframe depends on principles and methodologies of frequent pattern mining, subsequent concept definitions are offered:

- *Definition 1*: Transaction database: a transaction is a subset of a global item, and a collection of transactions becomes a transaction database. The set of demand centre addresses for a particular day in this paper can be called a transaction;
- *Definition 2*: Frequent item set: a collection of items in the dataset that appear frequently and simultaneously, i.e., a collection of freight demand points that appear frequently within a day within the time covered by the data;
- *Definition 3*: Support threshold: the percentage of occurrences of a particular freight demand point set in the dataset. When the support threshold of a freight demand point set D is greater than the pre-set minimum support threshold, the set D is said to be a frequent item set.

Commonly employed algorithms for mining frequent item sets include Apriori and FP-growth. In contrast to Apriori, the FP-growth algorithm achieves superior execution efficiency by employing a Frequent Pattern Tree (FPTree) for data storage and employing recursive techniques to extract all frequent item sets from the FPTree.

The pseudocode of the FP-growth algorithm

```
Input: $D$, a transaction database; $min\_sup$, the minimum support count threshold.
Output: The complete set of frequent patterns.
Method:
```

1. Construction of FP_tree.
(a) Scan the entire dataset to identify frequent 1-itemsets (containing only one geographical location) and sort them in decreasing order of support until the support is no less than the minimum support threshold.
(b) Scan the dataset again, sorting sets of geographical locations based on support as computed in Step (a). Then, sequentially insert these items into a tree with a null root node to construct the FP_tree.
1. Mining of FP-tree.

```
procedure FP-growth ($Tree, \alpha$)
if $Tree$ contains a single path $P$ then
    for each combination (denoted as $\beta$) of nodes in the path $P$
        generate pattern $\beta$ with $support\_count$ = $minimum\ support\ count\ of\ nodes\ in\ \beta$;
else for each $a_i$ in the header of $Tree${
    generate pattern $\beta=a_i \cup \alpha$ with $support\_count = a_i.support\_count$;
    construct $\beta$'s conditional pattern based and then $\beta$'s conditional FP_tree $Tree_\beta$;
    if $Tree_\beta=\emptyset$ then
        Call FP_growth ($Tree_\beta, \beta$)}
```

3.4 Site selection and cargo type excavation for Road-rail intermodal hubs

Potential hub point selection

Stations that have the conditions for construction and operational capacity of road-rail intermodal hubs are screened according to the following rules:

- *Station grade*. According to the *Approval Method of Railway Station Grade*, Station grade is determined through three key indicators: the average daily count of passengers boarding, alighting, and transferring, the quantity of arriving and departing transhipment packages, and the average daily count of loaded cars, unloaded cars, and cars used for transhipment operations. The station's grade directly reflects its transportation capacity. Only first-class or premium stations are selected as alternate stations.
- *Geographic location of stations*. Some of the stations are built in the main urban areas due to passenger demand and do not have the conditions for building hubs. This part of the passenger station will be removed from the potential hub points.
- *Station operation scope*. According to the public data on the website of China Railway 95306, some stations have restrictions on the handling of transportation business, for example, Shijingshan Station only handles steel transportation business. Stations with limited-service coverage will be removed too.

Construction of hub node location model for hub-spoke road-rail intermodal

The specific model symbols and parameters are defined in *Table 1*.

Table 1 – Symbols and parameters used in this problem

Type	Notations	Detailed definition
Sets	N_i	Set of demand points for road transportation, $N_i=\{1,2,3,\dots,i\}$
	N_j	Set of potential hub points for road-rail intermodal hubs, $N_j=\{1,2,3,\dots,j\}$
Parameters	v_i	Total volume transported at road demand point i, in tons
	f	Transit handling costs per unit of cargo volume at the potential hub point
	h_j	Maximum transport capacity of potential hub point j, in tons
	C	Unit transportation cost, unit: yuan per ton per kilometre
	e_j	Fixed cost of hub construction, unit: yuan
	d_{ij}	Distance from point i to point j, in kilometres
	q_{ij}	Volume of goods transported from point i to point j, in tons
	p_j	Volume of goods transported to hub point j, in tons
Variables	$x_j=\{0,1\}, j\in N_j$	Whether the potential hub point is selected
	$y_{ij}=\{0,1\}, i\in N_i, j\in N_j$	Whether hub point j provides intermodal services to demand point i

The following assumptions are made to facilitate the modelling solution and simplify the process of site allocation for actual road-rail intermodal transportation:

- The short-haul road transport distance is measured using road network data, but the excess transport distance due to road conditions and driver behaviour is not considered;
- Each road transportation demand point's needs can be met by only one hub site, and there is no behaviour of transportation to multiple hub sites dispersed for intermodal transportation;
- Only the fixed cost of building intermodal hubs and the cost of transit loading and unloading are considered, not other operating costs.

$$\min \sum_{j\in N_j}(e_j+fp_j)x_j+\sum_{i\in N_i}\sum_{j\in N_j}q_{ij}y_{ij}d_{ij}C \tag{3}$$

s.t.

$$\sum_{j\in N_j}x_j\le S \tag{4}$$

$$\sum_{j\in N_j}y_{ij}=1,\quad \forall i\in N_i \tag{5}$$

$$q_{ij}\le v_i y_{ij},\quad \forall i\in N_i, \forall j\in N_j \tag{6}$$

$$\sum_{j\in N_j}q_{ij}\le v_i,\quad \forall i\in N_i \tag{7}$$

$$p_j=\sum_{i\in N_i}q_{ij},\quad \forall j\in N_j \tag{8}$$

$$p_j\le h_j x_j,\quad \forall i\in N_i, \forall j\in N_j \tag{9}$$

$$y_{ij}=\{0,1\},\quad i\in N_i,\quad j\in N_j \tag{10}$$

$$x_j=\{0,1\},\quad j\in N_j \tag{11}$$

$$q_{ij}\ge 0, p_j\ge 0 \tag{12}$$

The objective function 3 is constructed with the objective of the most optimal road-rail intermodal hub construction, operation cost and transportation cost. *Equations 4–9* are the constraints related to site selection and transportation, in which *Equation 4* indicates that the maximum number of hub points to be constructed is *S*; *Equation 5* indicates that the demand of each road transportation demand point can only be satisfied by one intermodal hub; *Equation 6* indicates that the volume of transportation at a demand point cannot exceed the total demand at that demand point; *Equation 7* indicates that all the transportation demands at a demand point must be satisfied; *Equation 8* indicates that the transportation volume consolidated at the intermodal hub point should be equal to the sum of the volumes shipped from each demand point to the hub point; and *Equation 9* indicates that if the intermodal hub alternative point *j* is selected, the total consolidated volume at that node does not exceed its maximum transportation capacity. *Equations 10 and 11* are 0-1 decision variables, respectively, and *Equation 12* is a constraint to ensure the logical soundness of the model.

Goods type mining within the scope of hub service

Upon the identification of key hubs and their service areas, it becomes imperative to delve into the freight transport characteristics within the regional scope. This information will enable the railway department to enhance the planning and design of road-rail intermodal transport hubs, as well as formulate tailored marketing strategies to attract more road cargo sources for transfer to rail transport. Of primary concern to the railway department is the determination of whether bulk cargo transportation exhibits stable, high-volume characteristics. Furthermore, given that the economies of scale in railway transportation are not effectively realised in short-distance hauls. The criteria for screening cargo types suitable for intermodal transportation operations at the hub points, by combining the transportation volume and average transportation distance of sub-categories, are outlined as follows:

- Identify cargo categories with daily transportation volumes exceeding 10% of the total volume;
- In accordance with research on the optimal rail transportation distance, it is established that when the

average distance exceeds 500 km, and this cargo category's average distance surpasses its median distance, it will be considered as an alternative cargo type. This implies that the majority of goods within this category have distances exceeding 500 km, making them suitable for rail transportation.

3.5 Data

The data used in this paper mainly includes road freight data and railway station data.

The road freight data used in this paper is collected from an online trading platform with the highest market share in China's road transportation market, which currently has millions of active shippers and active drivers, with massive freight data and good big data storage conditions. The main data used in this paper is the 24 consecutive days of source data based on the shipment logs intercepted from the shipper's end with the Beijing-Tianjin-Hebei region as the research object. The data format is shown in *Table 2*.

Railway station data come from the station directory of the *2020 yearbook of China Railway Beijing Group Co. Ltd*, a total of 471 existing stations under the jurisdiction of the Beijing Bureau are obtained. These stations include various types, such as first-class stations and premium stations. This extensive network covers the entire Beijing-Tianjin-Hebei region, indicating significant differences in terms of transportation capacity, service operation scope, and primary responsibilities among these stations. Some stations may have greater responsibilities in passenger services, while others are primarily used for freight and other purposes. These stations will be selected as potential intermodal hub points according to the above-mentioned screening criteria.

Table 2 – Data format display

Item	Type	Note	Item	Type	Note
create_time	Date	Create Time	first_cargo_name	String	First-class cargo
timestamp	Timestamp	Creation Timestamp	Second_cargo_name	String	Second-class Cargo
start_prov	String	Departure Province	weight	Float	Weight
start_city	String	Departure City	capacity	Float	Volume
start_county	String	Departure District	o_lat	Float	Departure Latitude
end_prov	String	Arrival Province	o_lon	Float	Departure Longitude
end_city	String	Arrival City	d_lat	Float	Arrival Latitude
end_county	String	Arrival County	d_lon	Float	Arrival Longitude

4. RESULTS

4.1 Mining of road freight data

Before digging commences, we will initiate it, first-class goods types that mainly transport in the Beijing-Tianjin-Hebei region are classified into three distinct categories:

- Agricultural products and raw Materials: Corn, timber nursery, agricultural materials, food grains, fruits, vegetables, cash crops, live poultry and livestock, as well as agricultural and aquatic products.
- Industrial products and raw Materials: Building materials, coal and minerals, machinery and equipment, waste materials, accessories and parts, metal and steel, vehicles, paper, chemicals and plastics, clothing and textiles, and leather.
- Living goods: Food and beverages, office and education supplies, sports supplies, furniture, express deliveries, digital home appliances, fast-moving consumer goods, medicine, printing materials, and household paper.

Regional freight intensity measures the quantity of goods requiring transportation within a specific region and is commonly assessed using regional freight generation and attraction metrics. Subsequently, the 13 cities within the jurisdiction of Beijing-Tianjin-Hebei are delineated based on administrative boundaries, and the average of the outbound and inbound freight volumes for each city is calculated to create a regional freight transport intensity map, as illustrated in *Figure 3*. The average volume of goods attracted exceeds the volume of goods originating in all 13 cities. This suggests a prevalent focus on receiving goods in the region during the specified period. Notably, cities such as Shijiazhuang, Xingtai and Handan exhibit distinct char-

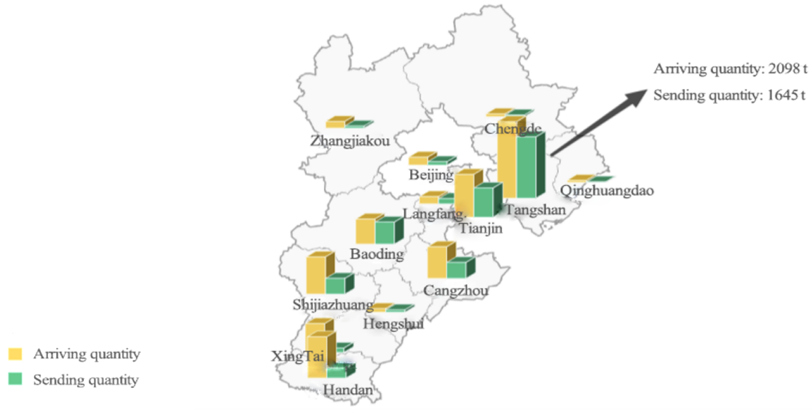
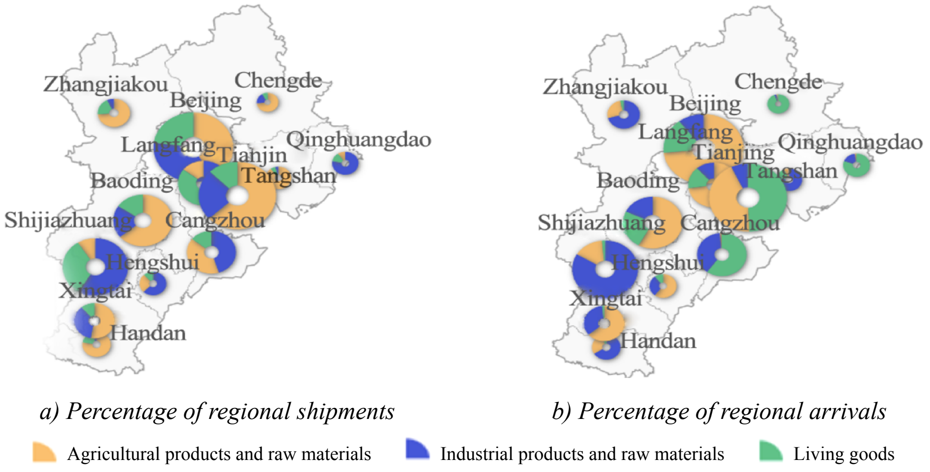


Figure 3 – Schematic diagram of freight transport intensity in the Beijing-Tianjin-Hebei region

acteristics favouring inbound shipments, while the other cities maintain a more balanced freight flow. Furthermore, excluding Tianjin and Beijing, when assessing the sum of goods attracted and goods originating from all cities and arranging them in descending order, a clear correlation emerges with the respective GDP of each city in 2020. This observation indicates that the freight transportation intensity of prefecture-level cities in the Beijing-Tianjin-Hebei region is closely tied to the local GDP level.

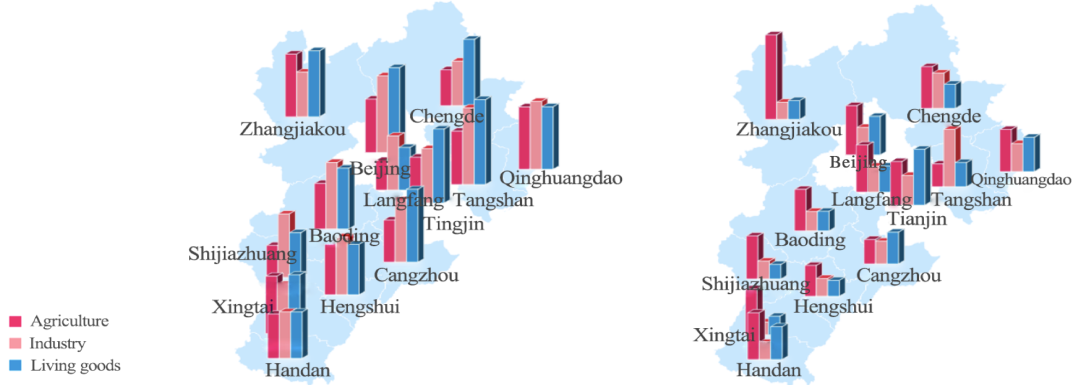
The results depicting the proportions of the three types of goods, both incoming and outgoing, in each sub-region of the Beijing-Tianjin-Hebei region are presented in *Figure 4*. By comparing the inbound and outbound volumes, an insight into the region's industrial structure can be gleaned through the utilisation of these proportions to depict the industrial relationships within each sub-region. For instance, taking Tangshan as an illustration, it is notable that the incoming goods consist mainly of industrial finished products, raw materials, and daily necessities, while the outgoing goods primarily comprise agricultural finished products and raw materials. This suggests that Tangshan's primary and secondary industries are well-developed, whereas the tertiary industry remains relatively underdeveloped.



a) Percentage of regional shipments b) Percentage of regional arrivals

Figure 4 – Beijing-Tianjin-Hebei regional goods types distribution diagram

Transport distances of the same cargo type within the same region are averaged, with the results presented in *Figure 5*. The average transportation distance for the three types of goods in the Beijing-Tianjin-Hebei region exceeds 300 kilometres, with some journeys stretching over 800 kilometres. It can be inferred that in the Beijing-Tianjin-Hebei region, where road transport constitutes a substantial portion of the total freight volume. In light of this, the railway sector should expedite the integration of road freight volume, aiming to enhance the overall efficiency of freight transport in the region. Furthermore, the transportation distances for various types of goods can reveal disparities in industrial structures between neighbouring regions. For instance, the extensive transportation of industrial products and raw materials arriving in the city of Tangshan signifies the resemblance of its industrial structure to neighbouring regions, primarily centred around iron and steel, mining, and metallurgy. As a result, it relies on longer-distance transportation to support the development of these industries.



a) Percentage of regional shipments *b) Percentage of regional arrivals*

Figure 5 – Average transportation distances for three types of goods in cities in the Beijing-Tianjin-Hebei region

When considering the distribution of regional goods categories and the average distances travelled, it becomes apparent that there remains a substantial volume of goods well-suited for railway transportation. This underscores the substantial potential for the development of "road-to-rail transportation". In such a context, the layout and construction of road-rail intermodal hubs can effectively optimise the transportation structure of the Beijing-Tianjin-Hebei region while simultaneously reducing pollutant emissions.

4.2 Demand centres of road freight

In this paper, the data from a total of 24 days from 25 June to 18 July 2021 are selected for cluster analysis to get the cluster class labels of each road freight demand point, in order to find out the demand hotspot area more clearly and accurately. The results of the first 4 days are shown in *Figure 6*. It can be found that a part of the hotspot area exists a large number of stable demands for cargo transportation; at the same time, there also exists a part of the hotspot area, whose demand for cargo transportation appears intermittently, presenting the characteristics of instability and low demand. The railway department needs to find the area with a stable and large amount of cargo source to layout the intermodal hub, so it needs to find the hotspot area which appears frequently in the time dimension in the above gathering area.

The transaction database is constructed according to the freight demand address, with days as a unit as a transaction, and considering the stability of freight demand, larger support degrees of 0.8, 0.9 and 1 are set

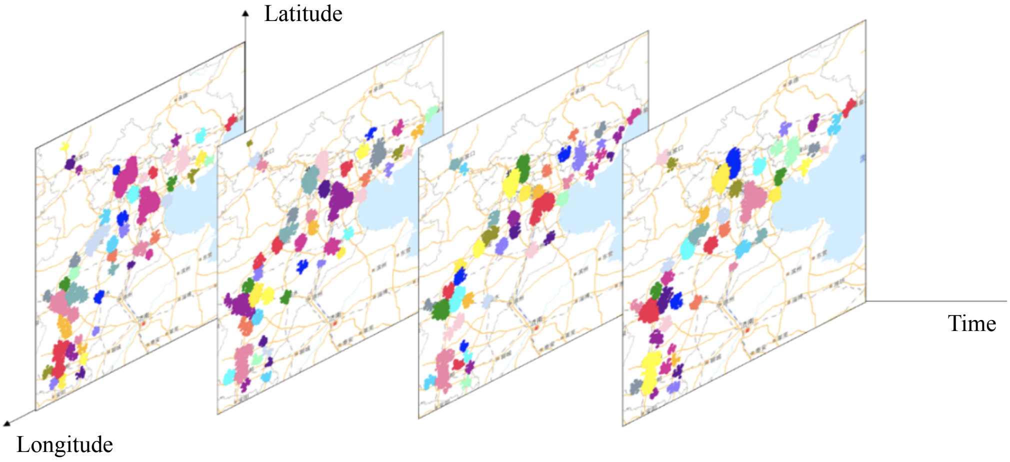


Figure 6 – Visualisation of road freight demand hotspots in the Beijing-Tianjin-Hebei region

to mine the frequent item sets. The results are shown in *Table 4*. When the support degree is 1, there is only one frequent item [Guangyang district of Langfang City, Renqiu district of Cangzhou City, Tang county of Baoding City, Binhai New District of Tianjin City, Xinhua district of Cangzhou City, Yutian county of Tangshan City, and Huanghua City, Cangzhou City, Wuan City, Handan City, Ninghe District, Tianjin City, Qian'an City, Tangshan City, Wen'an County, Langfang City], which means that all the sub-addresses in these 52 frequent item sets can be used as the centre point of the road freight demand intensive region 100% of the time, and they are correlated with each other. And there are 12 frequent item sets when the support is set to 0.9, and 52 frequent item sets when the support is set to 0.8.

4.3 Road-rail intermodal hub site selection and service goods type excavation

Following the guidelines established in the previous section, a total of 14 stations have been selected for the final screening. These stations include Fengtai West, Zhangjiakou, Baoding, Langfang, Nanchang, Tanggu, Cangzhou, Guye, Qinhuangdao South, Tangshan South, Shijiazhuang South, Handan, Hengshui, and Xingtai.

The data collection process involves gathering information on known conditions. This includes using the map API interface to retrieve navigation distances from each demand point to the alternative road-rail intermodal hubs. Additionally, the demand volume of road freight demand points is searched based on the geographical cluster to which they belong. The average daily total freight volume is calculated by summing the sending and arriving volumes within the cluster. It is assumed that each hub handles 20% of the road freight volume for the region to which the demand point belongs, recorded as the average daily freight demand in that area. Station capacity data is obtained from the *2019 yearbook of China Railway Beijing Group Co. Ltd*, where the average total daily traffic volume is calculated and recorded as the station's capacity. For stations with missing data, capacity is estimated using the average daily number of loading and unloading vehicles, with an average load of 60 tons per vehicle. The fixed cost of the hub and transhipment cost are computed at an average daily cost of RMB 200 yuan and RMB 8.6 yuan per ton, respectively. The price of road transportation is approximately RMB 0.5 yuan per ton-kilometre.

The problem involves 27 demand points and 14 alternative points when the support threshold is set at 0.8. The MATLAB platform is utilised to implement the modelling process. The results of the site allocation and the primary goods types scheme are presented in *Table 3*, with the first-class goods types listed in order of volume share. The spatial extent of the clustered cluster is used to define the service area of the road-rail intermodal hubs, as illustrated in *Figure 7*.

Figure 7 illustrates that the 12 road-rail intermodal hubs effectively cover the entire Beijing-Tianjin-Hebei region with a concentration of freight demand. There is a notable clustering of stations in the vicinity of Tianjin, Tangshan, and Qinhuangdao, each with a relatively small service area. In contrast, the stations in Shijiazhuang, Hengshui, Xingtai, and Handan, located in the southern region of Hebei, are spread over a wider area with more extensive service capabilities. Due to the dense presence of ports and industrial and mining enterprises, the freight demand density and transportation network density in Tangshan, Tianjin, and Qinhuangdao areas are higher compared to the southern Hebei region. Consequently, goods originating from Shijiazhuang, Hengshui, Xingtai, Handan, and other areas may require longer road transportation distances to facilitate intermodal transportation. This extended road transportation can result in increased costs, potentially diminishing customers' willingness to switch from road to rail. To address this challenge, it is advisable to consider increasing the number of alternative points for intermodal hubs or implementing a multi-level hub model for road-rail intermodal transportation hubs.

The types of goods that occupy the head position of road transportation volume in the Beijing-Tianjin-Hebei region are coal and minerals, grain and cereals, building materials, metal and steel, and vegetables. The following recommendations are made concerning the characteristics of regional development:

- Langfang and Baoding are near Beijing and Daxing Airport. They can serve a significant role in meeting the demand for relocating non-core functions from Beijing. These areas should consider introducing regular shuttle trains to attract express parcels, food and vegetables, as well as essential goods and materials

Table 3 – Scheme of hub siting, service area allocation and main service goods

Station name	**Service district**	**First-class goods types**
Baoding	Renqiu City, Tang County, Gaobeidian City, Dingzhou City, Lianchi District	Building materials, grain and cereals
Handan	Wuan City, Chengan County, Yongnian District, Da Ming County	Coal and Minerals, Metal and Steel, Grain and Grains, Building Materials
Langfang North	Guangyang District, Wen'an County, Bazhou City, Baodi District	Building Materials, Grain Grain, Lumber Nursery, Metal Steel
Tangshan South	Yutian County, Luannan County, Caofeidian District	Coal Minerals, Metal Steel, Grain and Grain Products, Building Materials
Hengshui	Xinji City, Taocheng District	Grain and Grain, Agricultural Materials, Coal Minerals, Building Materials
Tanggu	Binhai New Area, Ninghe District	Coal Minerals, Building Materials, Metal Steel, Agricultural Materials, Grain Grain
Qinhuangdao South	Changli County, Harbor District	Grain grain hulls, vegetables, metal steel, building materials
Xingtai	Baixiang County	Grain grain hulls, building materials, coal minerals, food and beverage
Shijiazhuang South	Xinhua District	Coal Minerals, Building Materials, Grain Grain, Metal Steel
Cangzhou	Huanghua City	Building Materials, Coal Minerals, Machinery and Equipment
Guye	Qian'an City	Coal and Minerals, Metal and Steel, Scrap and Waste, Building Materials
Zhangjiakou	Xuanhua District	Vegetables, coal minerals, metal steel, food grains

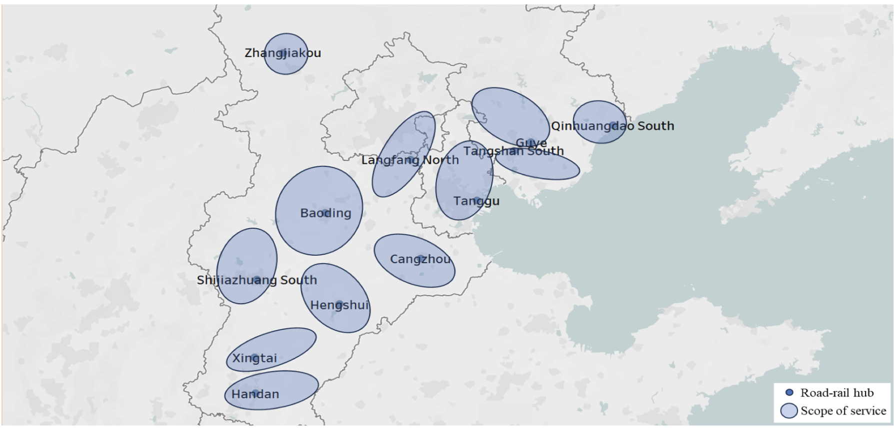


Figure 7 – Schematic diagram of the service area of the hubs

currently transported by road freight. Furthermore, with the development of the airspace economic zone, there's an opportunity to enhance rail transportation capabilities for high-end equipment, large machinery, and value-added finished goods.

- Tangshan, Tanggu, Qinhuangdao, and Cangzhou are situated in the Bohai Sea Economic Zone, which forms the most densely interconnected segment of the transportation network in the Beijing-Tianjin-Hebei region. It is also home to a multitude of industrial and mining enterprises. The railway department should leverage the strategic location of these cities to excel in facilitating the transportation of raw materials and finished products for the industrial sector.
- Shijiazhuang, Xingtai, Hengshui, and Handan should be strategically aligned with regional industrial development and transformation policies. It is crucial to engage in proactive planning and capacity allocation for various cargo categories, including but not limited to iron and steel, chemical finished products, building materials, textiles, garments, and foodstuffs. Furthermore, customised operational guidelines need to be established for enhancing the freight transportation capacity of the railway.

4.4 Estimation of freight attraction and carbon emission reduction

In this section, road freight volume that can be attracted by establishing a road-rail intermodal transport hub at an appropriate railway station is calculated based on real data. The section also estimates the carbon emission reduction resulting from the transfer of this freight volume from road to railway using the carbon emission reduction coefficient method [30].

According to promote the development of multimodal transport optimisation and adjustment of transport structure work program (2021–2025) issued by the General Office of the State Council, 2025 national railway freight volume relative to the 2020 growth target of 10%, calculated that the railway freight needs to increase 42.344 million tons to reach the target. According to the road goods data used in this paper estimates that the Beijing-Tianjin-Hebei region, the platform's annual goods volume of about 118,534,000 tons, 42,344,000 tons of the current platform annual goods volume of the Beijing-Tianjin-Hebei region of about 35.7%. Therefore, 35.7% of the current platform's goods volume as the road-rail intermodal hub can attract the amount of road freight, respectively, calculate the carbon emissions from road transportation of this part of the goods and the carbon emissions generated by railway, and the difference between the two is the carbon emission reduction of the intermodal hub area road to railway, results is shown in *Table 4*.

Under the premise of establishing road-rail intermodal hubs and each hub attracting 35.7 per cent of the surrounding freight, carbon emissions will be reduced by about 469,000 tons per year. Therefore, rationalising the layout of intermodal transport hubs and continuously increasing the proportion of road-to-rail transport volume can effectively reduce regional carbon emissions.

Table 4 – Estimated annual carbon emission reductions from road-to-rail transfers in each intermodal hub region (unit: tons)

Station name	Annual carbon emission reductions	Station name	Annual carbon emission reductions
Baoding	29660.95	Qinhuangdao South	7774.739
Handan	68000.51	Xingtai	50305.33
Langfang North	45495.46	Shijiazhuang South	64390.28
Tangshan South	17781.13	Cangzhou	34037.54
Hengshui	20490.36	Guye	17199.82
Tanggu	96304.46	Zhangjiakou	17209.11

4.5 Discussion

Freight intensity in the Beijing-Tianjin-Hebei region is positively correlated with urban economic development. Specifically, in Shijiazhuang, Xingtai, and Handan City, the volume of arriving freight significantly surpasses the volume of outgoing freight, while in other areas, the arrival and delivery volumes are roughly equal. The predominant goods types responsible for the high volume of road transport in the Beijing-Tianjin-Hebei region are primarily agricultural products and daily necessities, with an average transportation distance exceeding 500 kilometres. This issue is particularly pronounced in Tangshan, Tianjin, Baoding, Handan, and other locations. Establishing strategically positioned road-rail intermodal transport hubs has the potential to enhance the capacity of rail transport and reduce pollutant emissions.

Within the Beijing-Tianjin-Hebei region, there exists a set of highly stable freight transportation needs in terms of both timing and location. These freight needs are primarily concentrated in the Bohai Sea Economic Zone (Tangshan, Tianjin, Langfang, Cangzhou) and the southern urban agglomeration (Shijiazhuang, Hengshui, Xingtai, Handan). Furthermore, these regions host a substantial volume of goods suitable for rail transport, indicating significant untapped potential in rail transport.

The site selection scheme proposed in this paper identifies 12 multimodal transport hubs and specifies the primary source types. This initiative is expected to result in a reduction of approximately 469,000 tons of carbon emissions annually. These hubs effectively cover the most prominent road freight demand hotspots in the region. Notably, Tangshan and Tianjin have a dense distribution of multimodal transport stations, whereas Shijiazhuang, Handan, and Xingtai in the southern city cluster have a sparser distribution. This phenomenon is related to the difference in regional railway resources.

5. CONCLUSION

Existing hub selection studies often rely on idealised models, lacking substantial empirical data support. On the other hand, most freight data mining investigations focus on a single mode of transportation or a specific region, with limited integration across different transportation modes. Therefore, this paper starts from the perspective of real data, and three main studies are carried out in this paper. First of all, an analysis of the road-to-rail transfer in the Beijing-Tianjin-Hebei region was conducted, utilising goods source data to outline the freight transportation patterns within the region. Subsequently, clustering techniques and frequent itemset mining methods were employed to identify clusters of road freight demand aggregation areas. Building upon this foundation, mathematical planning models were used to determine the distribution of road-rail intermodal transport nodes and their corresponding service areas, as well as the types of goods most suitable for rail transportation within the region were identified.

However, the method framework has some limitations that should be noted. Firstly, an in-depth analysis of goods source data was not conducted in this paper. It would be beneficial to delve deeper into the analysis by examining different types of goods arrivals and shipment volume trends in future research. This analysis could be complemented by considering various regional economic indicators, changes in the industrial structure of the area, and correlations with freight demand. Secondly, when analysing the source of goods transitioning from road to rail, this paper did not account for the characteristics of shipping enterprises. It would be more meaningful to select the locations for intermodal hubs by identifying the unique characteristics of these shipping enterprises. In addition, the specific details of model construction for different modes of transportation are not considered due to the research focus. Incorporating the actual operational aspects of transportation in model construction can enhance the realism of the results. Subsequent studies can enhance siting needs analysis by combining source data with map points of interest (POI), and enhance siting realism and practicability by adding more traffic model details.

ACKNOWLEDGEMENT

This work was supported by the Fundamental Research Funds for the Central Universities (2682023JX007, 2682023ZTPY029); Science & Technology Department of Sichuan Province international innovation cooperation between Hong Kong, Macao and Taiwan (24GJHZ0331).

REFERENCES

[1] Wang R, Yang K, Yang L, Gao Z. Modeling and optimization of a road–rail intermodal transport system under uncertain information. *Engineering Applications of Artificial Intelligence*. 2018;72(1):423-436. DOI: 10.1007/978-0-387-30440-3_562.

[2] Basallo-Triana MJ, Vidal-Holguín CJ, Bravo-Bastidas JJ. Planning and design of intermodal hub networks: A literature review. *Computers & Operations Research.* 2021;136(105469). https://www.sciencedirect.com/science/article/pii/S0305054821002203 [Accessed 1st July 2021].

[3] O'Kelly ME. Activity levels at hub facilities in interacting networks. *Geographical Analysis.* 1986;18(4):343–356. DOI: 10.1111/j.1538-4632.1986.tb00106.x.

[4] O'Kelly ME. The location of interacting hub facilities. *Transportation Science.* 1986;20(2):92–106. DOI: 10.1287/trsc.20.2.92.

[5] Alumur SA, et al. Perspectives on modeling hub location problems. *European Journal of Operational Research.* 2021;291(1):1–17. DOI: 10.1016/j.ejor.2020.09.039.

[6] Wang S, Chen Z, Liu T. Distributionally robust hub location. *Transportation Science.* 2020;54(5):1189–1210. DOI: 10.1287/trsc.2019.0948.

[7] Yang X, Bostel N, Dejax P. A MILP model and memetic algorithm for the hub location and routing problem with distinct collection and delivery tours. *Computers & Industrial Engineering.* 2019;135:105–119. DOI: 10.1016/j.cie.2019.05.038.

[8] Soylu B, Katip H. A multiobjective hub-airport location problem for an airline network design. *European Journal of Operational Research.* 2019;277(2):412–425. DOI: 10.1016/j.ejor.2019.02.056.

[9] Kaveh F, et al. A new bi-objective model of the urban public transportation hub network design under uncertainty. *Ann Oper Res.* 2021;296(1–2):131–162. DOI: 10.3233/ifs-151709.

[10] Zargini B. Multiobjective location problems with variable domination structures and an application to select a new hub airport. *Logistics.* 2022;6(2):24. https://www.mdpi.com/2305-6290/6/2/24 [Accessed 14th Mar. 2022].

[11] Stoilova SD, Martinov SV. Selecting a location for establishing a rail-road intermodal terminal by using a hybrid SWOT/MCDM model. *IOP Conf. Ser.: Mater. Sci. Eng.* 2019;618(1):012060.

[12] Mohammadi M, Jula P, Tavakkoli-Moghaddam R. Reliable single-allocation hub location problem with disruptions. *Transportation Research Part E: Logistics and Transportation Review.* 2019;123:90–120. DOI: 10.1016/j.tre.2019.01.008.

[13] Kolář J, Stopka O, Krile S. Specification of a location for opted port logistics center in a given area using the method of operational research. *Transport Problems.* 2019;13(4):56–63. DOI: 10.20858/tp.2018.13.4.6.

[14] Shang X, et al. A credibility-based fuzzy programming model for the hierarchical multimodal hub location problem with time uncertainty in cargo delivery systems. *Int. J. Mach. Learn. & Cyber.* 2021;12(5):1413–1426. DOI: 10.1007/s13042-020-01239-8.

[15] Fontes FFDC, Goncalves G. A variable neighbourhood decomposition search approach applied to a global liner shipping network using a hub-and-spoke with sub-hub structure. *International Journal of Production Research.* 2021;59(1):30–46. DOI: 10.1080/00207543.2019.1693652.

[16] Sitek P, Wikarek J. Capacitated vehicle routing problem with pick-up and alternative delivery (CVRPPAD): Model and implementation using hybrid approach. *Ann Oper Res.* 2019;273(1–2):257–277. DOI: 10.1007/S10479-017-2722-X.

[17] Contreras I, Fernández E, Marín A. The tree of hubs location problem. *European Journal of Operational Research.* 2010;202(2):390–400. DOI: 10.1016/j.ejor.2009.05.044.

[18] Wu Z, Zhang Y. Integrated network design and demand forecast for on-demand urban air mobility. *Engineering.* 2021;7(4):473–487. DOI: 10.1016/j.eng.2020.11.007.

[19] Shahparvari S, et al. A GIS-LP integrated approach for the logistics hub location problem. *Computers & Industrial Engineering.* 2020;146(106488). https://www.sciencedirect.com/science/article/pii/S036 0835220302229 [Accessed 14th Apr. 2022].

[20] Huang D, et al. A static bike repositioning model in a hub-and-spoke network framework. *Transportation Research Part E: Logistics and Transportation Review.* 2020;141(102031). https://www.sciencedirect.com/science/article/pii/S1366554520306827 [Accessed 5th July 2020].

[21] Wen R, Yan W. Vessel crowd movement pattern mining for maritime traffic management. *LOGI – Scientific Journal on Transport and Logistics.* 2019;10(2):105–115. DOI: 10.2478/logi-2019-0020.

[22] Campbell JF. Integer programming formulations of discrete hub location problems. *European Journal of Operational Research.* 1994;72(2):387–405. DOI: 10.1016/0377-2217(94)90318-2.

[23] Mahmoodjanloo M, Tavakkoli-Moghaddam R, Baboli A, Jamiri A. A multi-modal competitive hub location pricing problem with customer loyalty and elastic demand. *Computers & Operations Research.* 2020;123(105048). https://www.sciencedirect.com/science/article/pii/S030505482030165 9 [Accessed 30th June 2020].

[24] Khodemani-Yazdi M, Tavakkoli-Moghaddam R, Bashiri M, Rahimi Y. Solving a new bi-objective hierarchical hub location problem with an M/M/c queuing framework. *Engineering Applications of Artificial Intelligence.* 2019;78:53–70. DOI: 10.1016/j.engappai.2018.10.004.

[25] Dukkanci O, Peker M, Kara BY. Green hub location problem. *Transportation Research Part E: Logistics and Transportation Review.* 2019;125:116–139. DOI: 10.1016/j.tre.2019.03.005.

[26] Alumur SA, Kara BY, Karasan OE. Multimodal hub location and hub network design. *Omega.* 2012;40(6):927–939. DOI: 10.1016/j.omega.2012.02.005.

[27] Cobeña B, Contreras I, Martínez-Merino LI, Rodríguez-Chía AM. The profit-oriented hub line location problem with elastic demand. *Computers & Operations Research.* 2023;159(106335). https://www.sciencedirect.com/science/article/pii/S0305054823001995 [Accessed 22th June 2023].

[28] Alumur SA, Nickel S, Saldanha-da-Gama F, Seçerdin Y. Multi-period hub network design problems with modular

capacities. *Ann Oper Res*. 2016;246(1–2):289–312. DOI: 10.1007/s10479-015-1805-9.

[29] Taherkhani G, Alumur SA. Profit maximizing hub location problems. *Omega*. 2019;86:1–15. DOI: 10.1016/j.omega.2018.05.016.

[30] Yao Z, Gan M, Li X, Liu X. Strategic plan for China’s air high-speed rail express freight network and its carbon reduction potential. *Environ Sci Pollut Res*. 2022;30(11):29110–29124. DOI: 10.1007/s11356-022-24244-6.

欧启晨，甘蜜（通讯作者），安媚童，王宜琛

基于公路货运需求挖掘的公铁联运枢纽选址——以京津冀地区为例

摘要:

为了实现公铁联运枢纽的合理布局选址，充分发挥铁路运输能力，助力低碳交通系统的构建，本文基于区域真实货运数据及铁路站点数据，提出了一整套完整的联运枢纽选址方案框架。首先，研究从运输流、货运强度、运输货类、平均运距角度分析挖掘了京津冀地区的货运现状；然后使用DBSCAN聚类和频繁项集挖掘方法对货运需求热点进行了空间与时间纬度的挖掘；最后根据以上结果，构建了数学选址模型和服务货类筛选规则。以中国京津冀地区为例，最终确定了12个多式联运枢纽及其服务范围内适合铁路运输的主要货物类型，经过测算预计每年可为该区域减少47万吨的碳排放。所提出的方法能够为各区域货运结构的优化提供有价值的指导。此外，它符合当代环境和可持续发展目标，有助于区域低碳交通系统的构建。

关键词:

多式联运; 枢纽选址; 碳排放; 数据挖掘