HOW DOES WIM DATA PROVIDE VALUE TO ITS INDUSTRY?

Ping Wang
Weighing Products Division of Vanjee Co., Ltd.
Chang'an University, China
Obtained M.Sc in vehicle Master of Engineering, 2011

Zhongyuan Chen
Weighing Products Division of Vanjee Co., Ltd.
Beijing University of Chemical Technology, China
Obtained M.Sc in Communication Engineering, 2012

Abstract

With the development of emerging technologies such as Internet of Things (IoT) and big data, the new data-driven development model has become the key in ITS construction. Traditional WIM generates value for vehicle enforcement, road planning, road and bridge design through data from a single station, and big data technology provides the possibility for WIM to generate greater value. In this paper, data from multiple WIM stations is fetched, remarkable applications have been achieved in overlimit enforcement, WIM accuracy tracking, road damage evaluation, vehicle profiling, etc., by data sampling, data fusion, data mining and data processing. In the following sections, the related experience will be shared; the establishment and application of WIM data warehouse will be discussed and how these applications create value.

Keywords: data-driven, data mining, WIM data, warehouse, ITS application

1. Background

WIM technology has been applied in China for more than 30 years and is a technology frequently used in the field of road management, which is widely used in applications such as vehicle enforcement, road planning, road and bridge design, and weight-based tolling. Relying on these application scenarios, the government has built a large number of WIM systems, but the value of WIM systems in these applications is realized through a single station data analysis. A large number of WIM systems generate rich data about vehicle weight, which is not integrated for create greater value.

WIM data platform brings together all WIM system data in a certain region into one data warehouse, adding two dimensions of spatial and temporal information to WIM data through data aggregation and accumulation. By tracking and mining the changes of WIM data in time and space, it optimizes the traditional processes of overlimit enforcement, WIM accuracy assessment and vehicle service, providing a new pattern of WIM data application.

Taking a city in China as an example, its local WIM data platform has been built in 2021. In the first year, 46 direct enforcement stations, 10 fixed overlimit control stations (fixed static weighing system), and 20 source enterprise stations (static weighing system at the factory exit to control all the freight vehicle weight from source) have been accessed; In the next five years, a total of 200 direct enforcement stations, 20 fixed overlimit control stations, and 180 source enterprise stations are planned to be accessed. The authors participated in the construction of this data warehouse and carried out data mining and application on the basis of this data warehouse.

2. WIM Big Data Architecture

Figure 1 shows the architecture of WIM data platform. It consists of five layers from bottom to top:

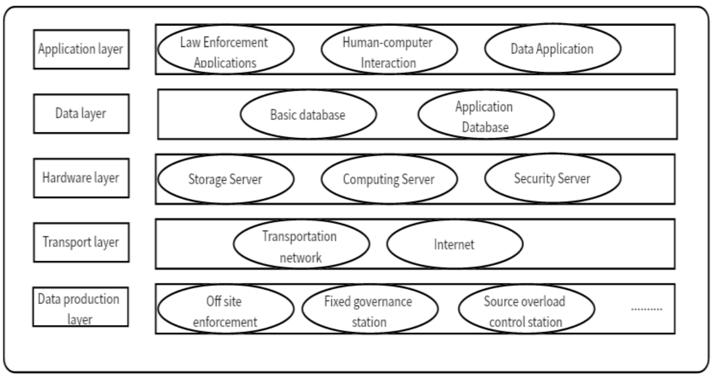


Figure 1. WIM data platform architecture

The 1st layer is the data generation layer, which is the WIM in vehicle law enforcement, road planning, road/bridge design, weighing and other application scenarios to complete the intended function, while uploading data to the data center via the data network;

The 2nd layer is the transmission layer, which uses multiple type of network transmission including private network, VPN and the Internet, etc., to complete the data transmission from each level of stations to the data center;

The 3rd layer is the hardware layer, which possesses computing and storage capability and provides hardware and software support while building a security system to ensure that the system is protected from network attacks and viruses;

The 4th layer is the data layer, which is built with various databases such as basic database and business application database. The data layer also undertakes data sharing and exchange inner this system as well as with other pre-completed systems;

The 5th layer is the application layer, which provides various business processing and business services for the application of WIM data and complete human-computer interaction.

3. Introduction of Data Construction

3.1. Project Overview

The WIM data platform was fully implemented for use in 14 districts and counties of this city. It has accessed to 46 sets of direct enforcement stations, 13 sets of fixed overlimit control stations and 48 sets of source enterprises stations. All the uploaded data of these stations include vehicle weight, license plate number, license plate picture and other information. It should be noted that the vehicle GPS data is also be accessed from the vehicle supervision platform of the province for enriching the basic data.

WIM data platform is a type of B/S architecture service platform. Its main users are law enforcement officers and supervisors. It achieves the supervision and assessment of the case processing process and processing efficiency in each district and county, also forms an online work pattern of overlimit controlling by means of business supervision, monitoring and management, command and dispatch, electric fence deployment and vehicle warning.

At the time of February 28, 2023, the access rate of the city outfield equipment of the overlimit control platform has reached 96% (a total of 106 access points, 102 of which have been accessed). The number of detected vehicles in the city is 9,270,370, of which 244,138 are detected at highway overlimit detection stations. 8,908,705 vehicles are detected by the non-stop overlimit detection system, and 117,527 vehicles are detected by freight source enterprises stations. There are 14 traffic law enforcement organizations in the city. And the WIM data platform has been accessed more than 1,200,000 times in total, with a total of 204 users, 121 of which are law enforcement officers for managing daily overlimit control cases, reviewing 42,864 overlimit data and handling 3,328 overlimit cases.

Figure 2. Data analysis platform

3.2. WIM System Basic Information

3.2.1. Verification Accuracy

There are 102 lanes in 46 sets of existing WIM stations, all of the lanes are built complying with class-5 accuracy of OIML R134, and have passed the metrology department's inspection in actual use.

To simplify the long-term accuracy comparison with in-service accuracy, standard deviations are calculated as an assessment mean of weighing performance in each station.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{m}\left(\frac{TMV_i - TMV_{ref}}{TMV_{ref}}\right)^2}{n}} \tag{1}$$

Table 1. Standard deviation statistics of verification among each station

Station number	Lane number	Average std	Minimum std	Maximum std
46	102	0.99%	0.63%	1.68%

3.2.2. In-service Accuracy

Among the 46 sets of WIM stations, data of 8 stations can be matched with adjacent control instruments installed at fixed overlimit control stations or source overlimit control stations by license plate number. Also, according to Equation 1, the weight of vehicle measured by control instrument as the conventional true value, the actual in-service accuracy of 8 stations are determined, and its comparison with the verification accuracy is shown in Table 2.

Table 2. The comparison std of between in-service and verification data of 8 stations

Station	Actual in-service		Verification	
	Number of vehicles	std (%)	Number of tests	std (%)
A	502	2.88	270	1.16
B	288	2.82	180	1.04
C	584	1.94	180	1.04
D	288	3.60	180	0.81
E	378	2.18	180	0.91
F	443	4.78	180	0.94
G	567	3.49	270	1.11
H	574	4.03	180	0.93

In Table 2, there is a large difference between in-service accuracy and verification accuracy, which results from the following reasons:

1. Verification only simulates vehicle weighing under the limited condition. But in actual use, the following factors have an impact on the weighing results. The external environment changes (temperature, weather, etc.), the changes in vehicle speed, and the changes in driving behaviour (acceleration, deceleration and lane- switching), etc.;
2. A single lane is as a unit during verification, but there are two or three lanes in one site. The data difference existing in each lane causes the difference between in-service accuracy and verification accuracy;
3. During verification, only WIM data accuracy is focused, but the data output from a station also includes vehicle weight and license plate number, incorrect matching these two types of data will bring serious errors;
4. Some of the WIM station equipment malfunctions in service.

Therefore, if the conditions allow, long-term data monitoring should be used by the metrology department as the basis for assessing the accuracy of WIM; even if the conditions for long-term data monitoring are not available, the measurement department should ask the equipment manufacturer to adjust the data output of a station as consistently as possible.

4. Data Value in Applications

4.1. Enforcement Assistance

4.1.1. Online Enforcement

In a region where there are multiple fixed overlimit inspection stations and direct enforcement stations, the overlimit vehicles may be punished for multiple times during one-time transportation. The WIM data platform achieves the definition of one-time overlimit transportation behavior by combining "algorithm + manual".

The algorithm of "one-time transportation behavior" groups the city overlimit detection data, and the staff will check and implement the education and execute penalty for the responsible subject in the process of overlimit reviewing, and share the review and punishment results to the 14 districts and counties in the city to realize the sharing of law enforcement information and avoid vehicles being repeatedly punished by the administrative department in one-time overlimit transportation behavior.

Figure 3. One-time transportation behavior judgement

4.1.2. Mobile Inspection

Mobile inspection as a necessary means of law enforcement, the need for law enforcement officers to perform on-site inspection tasks. The online strategy in this platform is to statistically analyze everyday data with more than 50% overlimit rate, and dynamically assess the passing vehicle overlimit distribution over time, as follows in Figure 4, 9:00-16:00 is the high prevalence time for overlimit at this station.

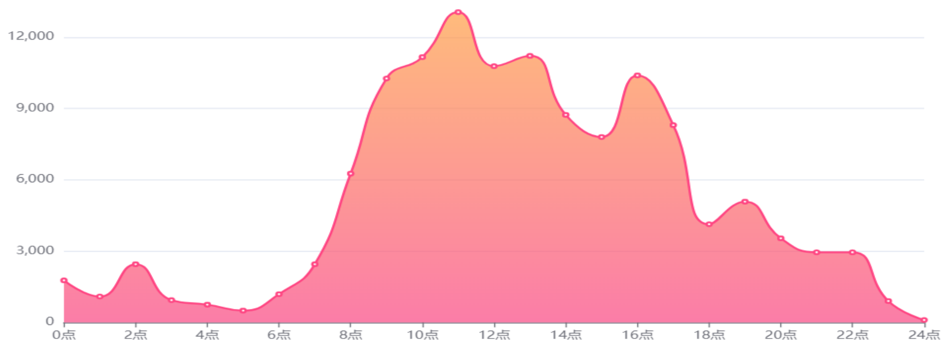


Figure 4. Overlimit behavior time with more than 50% overlimit rate over

Law enforcement officers accordingly make targeted deployment strategies, focus on breaking through key issues, achieve precise interception of repeatedly over-limited vehicles on the road, take mobile inspection measures for the phenomenon of maliciously blocking license plates, reasonably dispatch mobile law enforcement officers to key road sections for mobile inspection between 9:00 and 16:00 through the command hall. By combining online and offline, targeted law enforcement forces on key road sections, and a total of 135 intercepted vehicles have been seized so far.

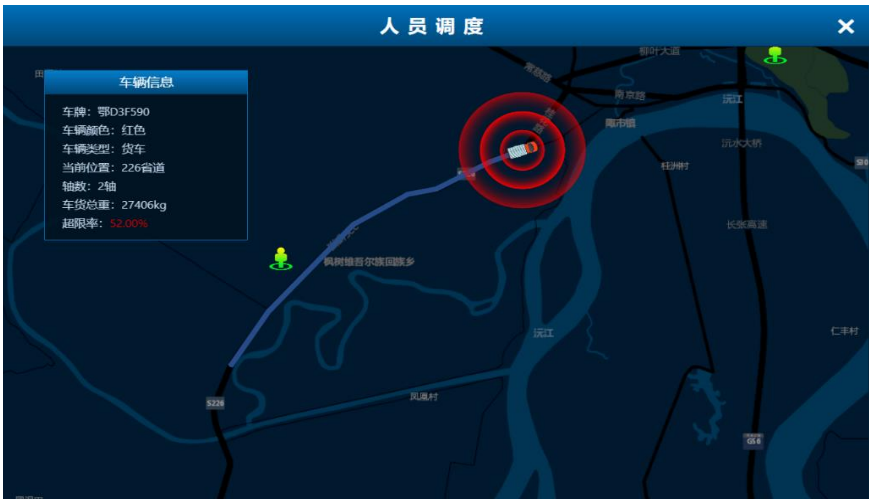


Figure 5. Law enforcement officer dispatch

4.1.3. Obscured License Plate Search

For direct enforcement obscured number plate violations, the WIM big data platform can extract vehicle-feature data from the passing vehicle images by AI image recognition algorithm and image feature extraction algorithm, etc. The vehicle images of obscured license plate are intelligently matched with historical data, to provide evidence chain for enforcement, so that make direct enforcement more credible, controllable and feasible.

Figure 6. Search for obscured license plate

4.1.4 Source Enterprise Supervision

As mentioned in 3.1, the WIM data platform accesses weighing data from multiple source enterprise stations. Source enterprise stations means important production enterprises and logistics center have their own weighing equipment (some are WIM some are static scales) at entrances and exits. If the vehicle exceeds the limit when entering the enterprise, in addition

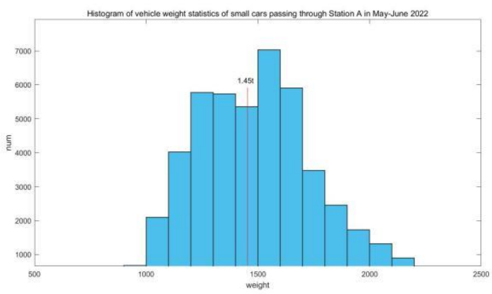
to penalizing the vehicle and driver, the government will also order the enterprise to manage the vehicle, and track where the vehicle exceeds the limit when loading, and pursue the source enterprise and freight enterprise for this overloading; if the vehicle exceeds the limit at the exit, the vehicle needs to be unloaded to a reasonable load before it can leave the factory.
The data from WIM and fixed overlimit control stations can monitor whether the source enterprises strictly implement vehicle loading management, and realize company profiling to control overlimit from the source.

4.2 WIM accuracy tracking

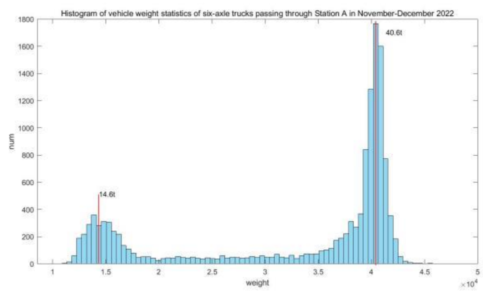
4.2.1. Analysis of individual WIM station history

By analyzing and tracking the long-term accumulated data from a single station, some problems in WIM usage can be identified and recognized.

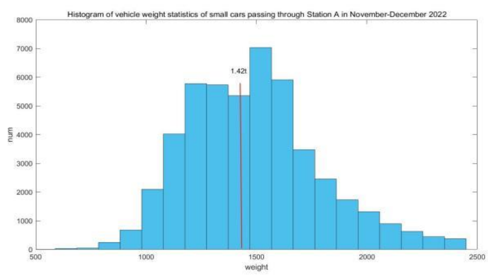
The histograms of weighing data of vehicle passing through station A for two time periods are shown in Figure 8, where *a* and *b* are the weight statistics histograms of small cars and six-axle trucks passing through station A in May and June 2022, respectively; *c* and *d* are the statistical histograms of the weight of small cars and six-axle trucks passing through Station A in November and December 2022 in Figure 8, respectively.



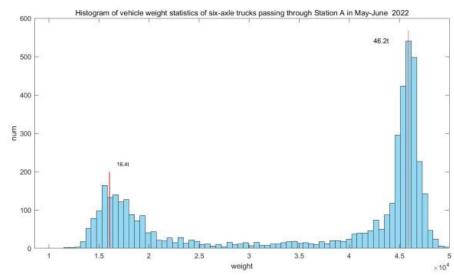
a. two-axle vehicles (May-June)



b. six-axle trucks (Nov.-Dec.)



c. two-axle vehicles (Nov.-Dec.)



d. six-axle trucks (May-June)

Figure 7. Histogram of vehicle weight data for two time periods at Station A

As can be seen from the figure:

1. In the two time periods, the vehicle weight of small cars and six-axle trucks appeared obvious data sets, of which small cars had one data set peak, and six-axle trucks had two data set peaks for empty cars and heavy trucks;

2. The average value of vehicle weight distribution in the two time periods has changed, their changes are in Table 3.

It can be clearly seen that the data of Station A has changed dramatically. It can be tell from these data that Station A needs further calibration. WIM data platform tracks the changes of the peaks of different types of vehicle data on a daily basis to determine whether the current Station is working properly. It should be noted that all data comes from Station A, and Average vehicle weight in May and June 2022, abbreviated as $\overline{TMV_1}(t)$ Average vehicle weight in November and December 2022, abbreviated as $\overline{TMV_2}(t)$

Table 3. Average vehicle weight statistics for two periods at Station A

	$\overline{TMV_1}$ (t)	$\overline{TMV_2}$ (t)	$\overline{TMV_1}/\overline{TMV_2}$
six-axle empty vehicle	16.4	14.6	0.89
six-axle truck	46.2	40.6	0.88
small vehicle	1.45	1.32	0.91

4.2.2. Multiple WIM stations data analysis

The data processing method mentioned in 3.2.2 is only applicable to some of the stations: the data from these stations can be matched to the neighboring control instruments (installed at fixed station or Source Enterprise Supervision station) by license plate number. For the stations that do not meet the conditions, we designed the following evaluation method based on standard deviation.

There are four WIM stations A, B, C, D. Based on the license plate number, weighing data of the vehicle passes through two WIM stations can be matched:

$$E_{AB} = \frac{W_A - W_B}{W_B} \tag{2}$$

where W_A and W_B are the weights of WIM station A and B,

$E_A = \frac{W_A - GVW}{GVW}$, $E_B = \frac{W_B - GVW}{GVW}$, plug into Equation 2,

where GVW stands for Gross Vehicle Weight. So that,

$$E_{AB} = \frac{W_A - W_B}{W_B} = \frac{GVW*(1+E_A) - GVW*(1+E_B)}{GVW*(1+E_B)} = \frac{E_A - E_B}{1+E_B} \tag{3}$$

If E_B << 1, we get

$$E_{AB} \approx E_A - E_B \tag{4}$$

So σ_{AB}, the standard deviation of E_{AB}, can be expressed as:

$$\sigma_{AB} \approx \sqrt{\frac{\sum_{i=1}^{n}(E_A - E_B)_i^2}{n}} = \sqrt{\frac{\sum_{i=1}^{n}(E_A^2 + E_B^2 - 2*E_A*E_B)_i}{n}} = \sqrt{\frac{\sum_{i=1}^{n}(E_A^2 + E_B^2 + 2*E_A*E_B)_i}{n}}$$
$$= \sqrt{\frac{\sum_{i=1}^{n}E_{A_i}^2 + \sum_{i=1}^{n}E_{B_i}^2 - 2*\sum_{i=1}^{n}(E_A*E_B)_i}{n}} \tag{5}$$

Here E_A and E_B are normally distributed data with a mean of 0, then

$$\sum_{i=1}^{n}(E_A*E_B)_i = 0 \tag{6}$$

$$\sigma_{AB} \approx \sqrt{\frac{\sum_{i=1}^{n}E_{A_i}^2 + \sum_{i=1}^{n}E_{B_i}^2}{n}} = \sqrt{\frac{\sum_{i=1}^{n}E_{A_i}^2}{n} + \frac{\sum_{i=1}^{n}E_{B_i}^2}{n}} = \sqrt{\sigma_A^2 + \sigma_B^2} \tag{7}$$

Re-write the formula as,

$$\sigma_{AB}^2 \approx \sigma_A^2 + \sigma_B^2 \tag{8}$$

Apply Equation 8 to 4 stations, we have an equation set:

$$\begin{cases} \sigma_{AB}^2 \approx \sigma_A^2 + \sigma_B^2 \\ \sigma_{BC}^2 \approx \sigma_B^2 + \sigma_C^2 \\ \sigma_{CD}^2 \approx \sigma_C^2 + \sigma_D^2 \\ \sigma_{AC}^2 \approx \sigma_A^2 + \sigma_C^2 \\ \sigma_{AD}^2 \approx \sigma_A^2 + \sigma_D^2 \\ \sigma_{BD}^2 \approx \sigma_B^2 + \sigma_D^2 \end{cases} \tag{9}$$

In the Equation 9, $\sigma_{AB}^2, \sigma_{BC}^2 \ldots\ldots \sigma_{BD}^2$, all of them can be calculated by matching the data between WIM stations A, B, C, and D. Plug their values into the equation set, and solve the over-determined equation by the least squares method. Once get $\sigma_A^2, \sigma_B^2, \sigma_C^2, \sigma_D^2$. The standard deviation of the weighing error of the four WIM stations can be obtained.

The data from the above 8 stations mentioned in 3.2.2 are substituted into the above formula. The CV of the model and the Comparison Value Effect (CVE) with the control weighing instrument are shown in Table 5:

Table 5. Standard deviations comparison of control instrument and multi-station data models

CVE (%)	2.878	2.818	1.935	3.596	2.175	4.781	3.494	4.026
CV (%)	2.71	2.72	1.68	3.51	2.01	4.57	3.33	3.91

Compared to the CV of standard deviation in Table 5, The calculated standard deviation (CV) is very close to the standard deviation by using the control instrument. Therefore, calculated standard deviation can be used to evaluate the accuracy of WIM.

By using the methods in 4.2.1 and 4.2.2, the usage of online WIM devices can be evaluated, in order to quickly and accurately identify WIM usage issues and guide subsequent maintenance

4.3. Road Damage Pattern

The Chinese government, in order to monitor the operation of freight vehicles, has installed positioning terminals on trucks to obtain vehicle driving information, including GIS, speed, time, angle and corresponding vehicle license plates, etc. The above information of freight vehicles in this province is aggregated to the provincial vehicle supervision platform. As mentioned in 3.1, the vehicle GIS data of the provincial vehicle supervision platform is connected to the WIM data platform. On the platform, the GIS trajectory of all vehicles entering the city is collected, as well as overlimit information of the vehicles are matched. By calculating the equivalent axle data, the vehicle damage impact on the road surface is obtained.

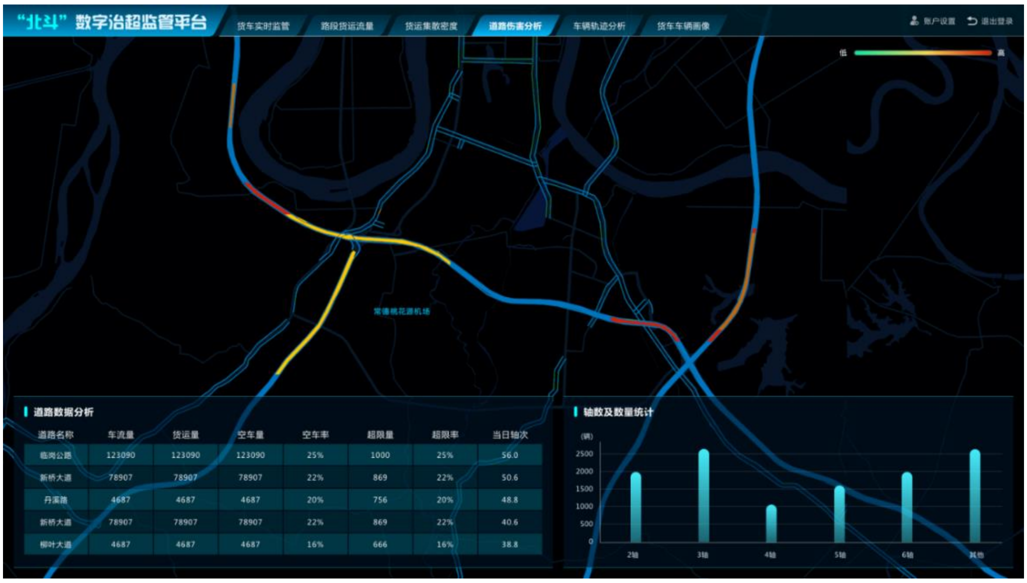


Figure 8 Vehicle Communication Heat Map

GIS and WIM data combination of road damage pattern can give road damage information of each precise small road section in the range. For some sections without the deployment of overlimit equipment, the road damage condition can also be measured by a reasonable trajectory matching, in order to provide effective evidence for accurate and intelligent road maintenance decisions. Subsequently, the pattern can be integrated with additional information according to different road surface materials, road age, and so on to offer targeted road damage patterns of different road sections.

4.5. Truck Distribution Density

By combining vehicle GIS data, vehicle WIM data and other information in the target administrative area, the overlimit control platform marks the collection and distribution hotspot area, while monitoring the circulation of trucks in each district within the range.

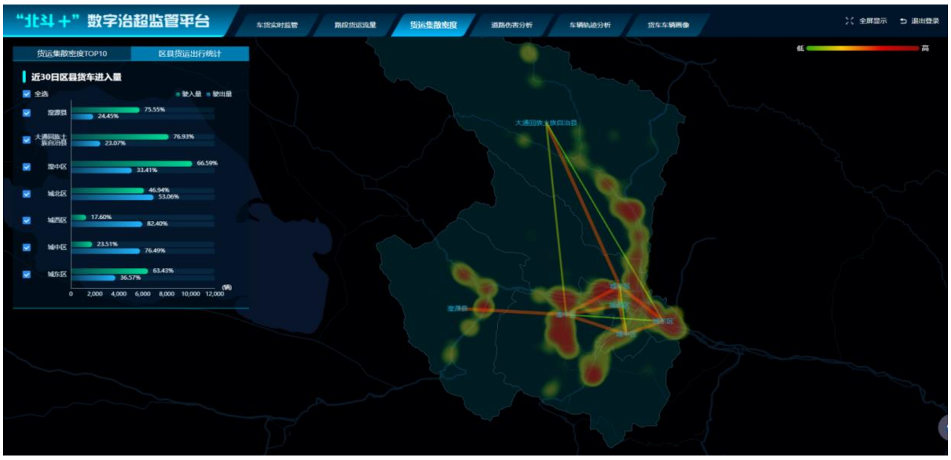


Figure 9 Schematic diagram of truck distribution density

The platform uses GIS data to slice and dice the trajectory of trucks and determine the location of parking hotspots. When combined with the license plate, time, and other details, the platform can match the WIM data generated by the corresponding trip of the vehicle. WIM data provides the weight records of trucks in different trajectory segments, and combined with GIS data can form the vehicle trajectory with a single time of transportation. It can keep track of how many vehicles are moving through various zones while also accurately determining the distributing center according to the change of truck weight at the parking point.

The accurate determination of the truck distributing center helps the overlimit management department to carry out targeted supervision and control inspection for the key areas, as well as monitor the cargo flow in each area, and discover and track the abnormal cargo condition in time.

5. Conclusion

We believe that the intelligence and digitization of traffic system has always been a crucial direction of traffic development. The digitization of roads and vehicle information on the roads is the foundation of intelligent transportation technology development. Vehicle weight as an important indicator to describe the condition of the vehicle, its digitization is indispensable in the intelligent transportation system. With the advancement of ITS technology and big data technology, an increasing number of application scenarios require vehicle weight information. In the foreseeable future, the WIM system will be fully integrated with big data technology and ITS technology to develop enhanced functionalities and systems that are not confined to the current existing application scenarios.