2022 5th International Conference on Data Science and Information Technology (DSIT) | 978-1-6654-9868-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/DSIT55514.2022.9943937

Spatio-temporal trajectory big data analysis based on HNCORS

Yubing Zhang
Hunan Institute of Geomatics Science and Technology
Changsha, China
280584476@qq.com

Xiangqiang Zeng
Hunan Institute of Geomatics Science and Technology
Changsha, China

Minsi Ao
Hunan Institute of Geomatics Science and Technology
Changsha, China

Wenping Jin*
Hunan Institute of Geomatics Science and Technology
Changsha, China
404140619@qq.com

Abstract*—The analysis based on spatio-temporal trajectory big data is of great significance for mining user operation rules, predicting behavior trends, and exploring application expansion. Hunan continuously operating reference station network, which provides the official real-time kinematic service to the public, is one of the most important regional geospatial datum infrastructures in Hunan province, China. Using the location based big data of users, the multi-scale and multi-time characteristics of trajectory operation data and fixed solution ratio are analyzed by preprocessing, interpolation, kernel density analysis and other methods. The results show that the distribution characteristics of trajectory data in Hunan province are related to the working habit in field surveying and mapping, and the overall performance is reduced from the Middle East to the West, from economically developed areas to economically poor areas, and from plain hilly areas to mountain / lake areas. Among them, the suburbs of towns are relatively dense areas of trajectory data. The amount of trajectory operations is also significantly correlated with the local economy. By analyzing the spatio-temporal sequence diagram of the trajectory data, the evolution process is divided into six stages, which is manifested in the continuous growth and decline around Changsha-Zhuzhou-Xiangtan region and the eleven urban areas in a natural day.

***Keywords*—Trajectory big data, real-time kinematic, HNCORS, kernel density analysis, spatial-temporal analysis**

I. Introduction

Trajectory data plays a significant role in the field of data mining due to the wide applications. The rapid development of satellite navigation, Internet, wireless communication technology and the popularity of various mobile terminals have spawned the generation of massive mobile trajectory data. The analysis and mining of trajectory data and spatio-temporal feature information has become one of the most significant research topics.

Researchers summarized the characteristics and processing of trajectory data, and analyzed the technical development of a large number of trajectory data in the era of big data[1]. Research on trajectory spatio-temporal data based on people and vehicles, especially taxis, is one of the hotspots in big data research around the world. The acquisition and research of large amounts of spatio-temporal trajectory data of people and vehicles can break the traditional spatial boundaries. Some researchers have studied the travel trajectory and surrounding environmental characteristics of social work and life, and excavated the social activity of individuals or groups to predict the travel behavior characteristics and spatio-temporal distribution of individuals or groups[4][5][6]. By using the trajectory data collected by tens of thousands of taxis, researchers extract the structural information of 189 intersections of urban roads, the accuracy rate is 96.1%[7][11][8]. Paper analyzes the dynamics and 'heat' of urban space in Lanzhou and Wuhan in different periods based on the urban space-time data such as vehicle trajectory data and online social check-in, studies the functional characteristics and interaction modes of urban areas, understands traffic and urban planning problems, and provides support for urban construction[9][10][11][12][13][14]. In the literature, a method for channel boundary extraction based on large-scale ship trajectory data is proposed, which can effectively solve the problem of cavity in traditional boundary extraction[15]. Reference [16] proposed a road data acquisition method suitable for large data processing with high precision and easy realization based on the mass source trajectory data, which can reduce the economic cost and time cost of road data extraction.

The Hunan continuously operating reference station network (HNCORS) provides official service for users in Hunan province. In 2011, HNCORS completed the first phase of construction and entered the trial operation stage, providing real-time kinematic (RTK) services to the public. HNCORS completed the second phase of the project and the upgrading construction of compatible Beidou in 2016. On the basis of the original, the data storage format is improved to store RTK position big data, which provides data support for the research and analysis of trajectory big data. Since 2011, HNCORS has provided high-precision real-time satellite positioning services to the society, covering more than 600 units in more than 20 industries such as natural resources, meteorological monitoring, water conservancy and hydropower, and traffic navigation, with an average daily number of 1.2 million visits. Based on the trajectory data generated by RTK service in 2018, this study analyzes the characteristics by spatio-temporal analysis method, and provides reference for the follow-up improvement and upgrading of HNCORS.

supported by the National Natural Science Foundation (No.42074016), The Science and Technology Platform and Talent Program of Hunan Science and Technology Administration (No.2018TP2040), the Changsha City Natural Science Foundation (No.kq2014161) and the Science and Technology Research Program of Hunan Provincial Natural Resources Department (No.2021-08)., Hunan Engineering Research Center of High Precision Satellite Navigation and Location Based Service of BDS

II. Data Resource

The user data obtained by GNSS measurement can reflect the user's location and behavior preferences, which are characterized by large volume, fast update and information fragmentation. In this paper, a total of 250 million original position data with a total of 25 GB in 2018 are used for extraction and analysis. The original indexes such as time and position (geodetic coordinates of the 2000 national geodetic coordinate system) are selected for extraction and transformation. Three indexes, namely time, position and fixed proportion, are condensed for feature analysis, and the distribution law and characteristics of user spatio-temporal trajectory data are excavated.

III. Processing Method

Data processing includes preprocessing, time analysis, spatial analysis and spatio-temporal evolution analysis.

A. Preprocessing

Like all kinds of big data, trajectory data have a series of data quality problems, such as inaccurate data caused by positioning terminal or physical environment, data missing caused by equipment, transmission problems or operation errors, data redundancy and so on. Therefore, the original data cannot be directly analyzed and mined. Firstly, data screening and conversion need to be carried out, which mainly includes data cleaning, thinning, merging, trajectory segmentation, and machine training. Through data cleaning to eliminate incomplete structure or obvious error data, such as file format storage error records. Then, the high-frequency records sent by the same user are diluted, and the thinning interval is generally set to 1 min. Then, the data after cleaning and thinning are segmented into trajectories, and the long-term trajectories are segmented in the unit of day. Each segmented record represents a trajectory and is stored in a single file. Finally, through the spatial homing of all trajectories, the distribution map on the background of the corresponding regional geographic map is formed for subsequent spatial and temporal characteristics analysis. A total of 235958 trajectory data records were extracted after preprocessing

B. Spatio-temporal analysis

Time feature analysis Statistics are carried out from the monthly, intra-week and 24-hour scales to mine and analyze the time distribution characteristics of the number of user trajectory operations and the proportion of fixed solutions.

Spatial characteristics are analyzed from the provincial, municipal and county levels, according to job density, fixed solution ratio. Based on the theory of data mining, the kernel density analysis method of cluster analysis is used to estimate the user trajectory job density in a certain range. The search range of kernel density analysis is set, and then the spatial characteristics are analyzed by thermal diagram. The spatial distribution characteristics of provincial fixed solution proportion are directly weighted and interpolated according to the corresponding geographical location. The spatial characteristics of cities and counties are analyzed by using the preprocessed data, combined with the trajectory operation intensity and the fixed solution proportion, and the correlation between the surveying and mapping activities in the region and the surrounding environment and the geographic information management of surveying and mapping are studied and judged.

Spatio-temporal characteristics analysis is based on the province ' s trajectory distribution map according to the 24-hour sequence, and then the preprocessed data set is divided into each hour for kernel density analysis, drawing the corresponding trajectory data distribution map, and analyzing the spatial and temporal variation characteristics of each hour.

IV. Results and Analysis

A. Analysis of time characteristics

The trajectory data of users in different periods reflect the changes in the number of users in the periods. Figure 1 (1), (2) and (3) are the characteristic curves of user trajectory data at monthly, intra-week and hourly scales.

It can be speculated from the above figure that the number of user operation trajectories should be related to the working and rest habits of field surveying. On the monthly scale, the number of user trajectory operations is significantly decreased in January and February when the temperature is low and the Spring Festival is located. In the warmer August and September,

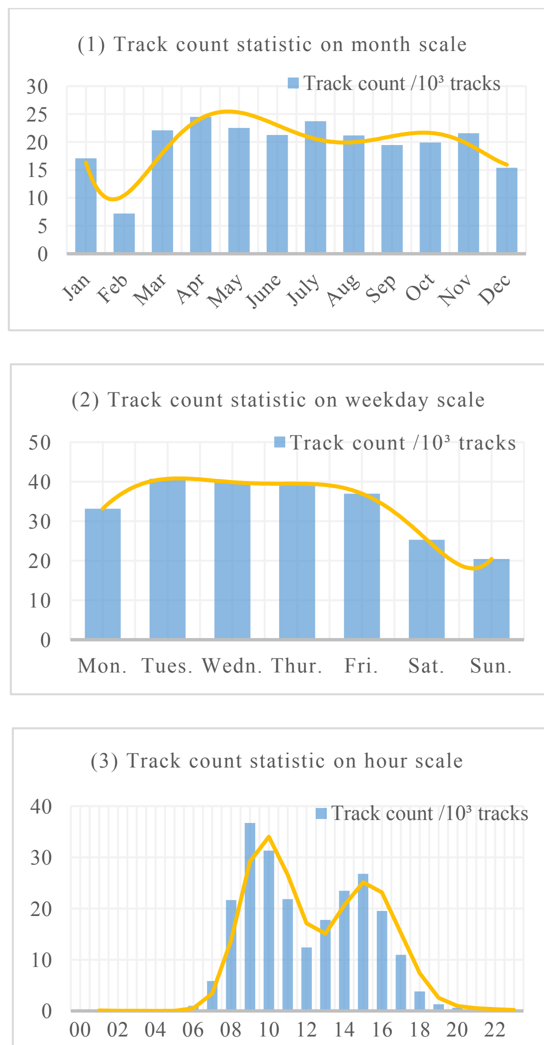

Fig. 1. Characteristic curve of user trajectory data in each period

the number of users also decreased slightly. On the week scale, the minimum number of count is the rest day. In addition, the user trajectory characteristics are significantly related to the working time, showing two peaks at 09a.m. and 15p.m. respectively. During lunch and rest around 12:00, the number of track operations is in the trough. From 18a.m. to 8a.m. the next day, the number of users fell sharply to almost zero. The distribution characteristics of trajectory data are basically the same as the frequency distribution characteristics of point operations[19]. Based on the above situation, the normalized benchmark service operation, maintenance and debugging should be arranged as far as possible in the period of less user operations.

The time characteristics of fixed solution proportion reflect the change of fixed solution proportion in each period. The higher the fixed proportion is, the larger the proportion of users reaching centimeter-level accuracy level is than that of users at decimeter-level or meter-level accuracy level. Fig. 2 shows the characteristic curves of monthly and hourly user fixed solution ratios.

It can be seen from fig.2 that the proportion of fixed solutions is relatively stable at the monthly scale, but there is a certain fluctuation at the hourly scale. On the monthly scale, the maximum, minimum and mean fixed solution ratios are 69.30 %, 62.72% and 65.48%, respectively. On the hourly scale, the characteristics of point-following operation are slightly different[19].Combined with the analysis of trajectory workload, the trajectory workload from 07:00 to 19:00 accounted for 98.87 % of the total number of trajectories, and the proportion of fixed solution showed the characteristics of similar two peaks, with the peaks appearing at 9:00 and 14:00, respectively, and the proportion of fixed solution reached 66.91% and 66.55%, respectively. When the trough appeared at 12:00, the fixed proportion was 59.97%. User's observation conditions, communication networks, benchmark service systems and other aspects may cause changes in the proportion of fixed solutions. The characteristics of the hourly scale of the fixed solution are reflected in three aspects. Firstly, due to the time habit of field work, the morning and afternoon are the high-efficiency periods of operation, respectively. When the rest period in the mid-noon, GNSS measurement equipment is no longer placed in the open area that meets the GNSS observation conditions, but is placed under the tree and in the room, which is more suitable for rest, resulting in a decrease in the proportion of fixed solutions. Secondly, ionospheric disturbance is also a factor affecting the fixed proportion. The low-latitude region in Hunan is more vulnerable to the influence of ionosphere, which is also one of the reasons for the change of fixed proportion. At the same time, the daily ionospheric monitoring results show that the ionospheric disturbance and I95 index reach a peak at about 12 o ' clock. The period with the lowest fixed proportion appeared at 23:00, and the fixed proportion was only 44.00%.According to the analysis of trajectory number, there is almost no trajectory operation at 23:00, which may be caused by project catch-up, equipment testing and other reasons.

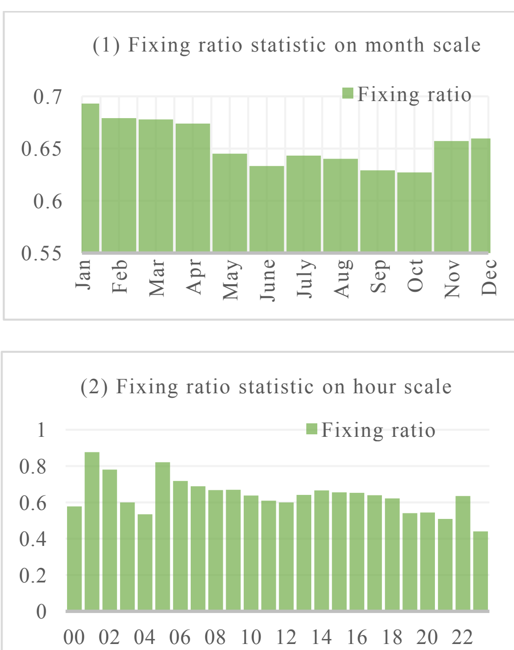

Fig. 2. Fixed proportion characteristic curve of each time period

B. *Analysis of the spatial characteristics*

The spatial characteristics of different scales reflect the trajectory operation and fixed proportion distribution in HNCORS coverage area. Figure 3 (1) and (2) are the distribution maps of the number of trajectory operations and the fixed proportion in the coverage area.

As shown in Fig.3 (1), the distribution characteristics of trajectory data in Hunan province is reduced from the Middle East to the West. The specific analysis shows that it is mainly manifested in three aspects. Firstly, the trajectory data in developed areas is denser than that in backward areas. The Changsha-Zhuzhou-Xiangtan urban agglomeration with relatively developed economy and convenient transportation, the surrounding area of Dongting Lake and the area adjacent to Guangdong Province in the south have higher data intensity. In contrast, less developed areas in western Hunan have lower data intensity. Secondly, there are more trajectory operations in plain hills and less trajectory operations in mountain / lake areas. In Fig. 3 (1), the areas with the smallest trajectory intensity are located in the central Dongting Lake area, Huping Mountain area, Wuling Mountain area, northern Luoxiao Mountain area, southern Xuefeng Mountain area, Wanyang Mountain and Dongjiang Lake area, and Jiuyi Mountain area, respectively. Only due to the prohibition of development zones or nature reserves in mountainous and lake areas, the demand for engineering construction is small, and the mapping work is also less carried out. Thirdly, the intensive track operation areas are mainly urban suburbs. Combined with the urban distribution in Figure 3 (1), it can be seen that the trajectory operation intensive area is located in the suburban area with the fastest urbanization process. In addition to the Changsha-Zhuzhou-Xiangtan area has already formed the urban group effect, Yueyang, Loudi, Zhangjiajie, Huaihua, Yongzhou, Chenzhou track operation hot spots are basically located in the western urban area, Hengyang, Shaoyang and Yiyang are located in the southwest suburbs. According to the spatial distribution of trajectory operation intensity, planning and layout should be carried out. In recent years, it should focus on promoting real-time surveying and

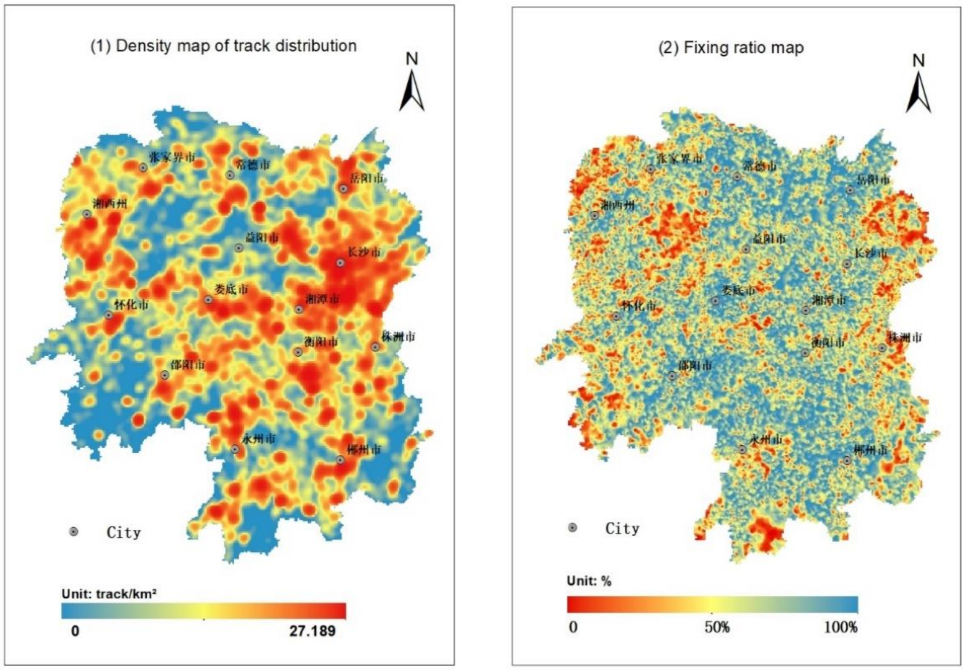

Fig. 3. Overall Track Data and Fixed Proportion Spatial Distribution in the Province

mapping reference service in the suburbs of cities and states. It can be seen from Fig. 3 (2) that the proportion distribution of fixed solution in the whole province is generally similar to that of trajectory operation, and the most obvious characteristic is that the proportion of fixed solution in mountainous areas is low. In most areas of Hunan Province, the fixed solution ratio is about 65%. However, the proportion of fixed solutions in the northwest, northeast, southwest and south is significantly low, and the reasons are mainly two aspects. Firstly, these areas are mostly mountainous areas, and the complex terrain and lush trees in mountainous areas are not suitable for GNSS observation, which cannot ensure sufficient GNSS signal positioning. Secondly, the communication network infrastructure in mountainous areas is also weaker, and users lack effective mobile communication networks to transmit and obtain differential observation data to achieve high-precision positioning.

The number of trajectory operations is divided by the city and county scale, the results are as follows.

It can be seen from Fig.4 (1) that among all the cities, Changsha is the most and Zhangjiajie is the least. On the whole, cities and states in the developed eastern region have the largest demand for high-precision surveying and mapping reference services, so their trajectory operation intensity is also high on the whole. As the pillar industry of Zhangjiajie City, the construction quantity related to surveying and mapping reference service is obviously less. In addition, the small area (9534.59km^2) is one of the reasons for the small number of trajectory operations in Zhangjiajie. The trajectory operation per unit area of Zhangjiajie City is only 0.712 times/ km^2. Compared with Huaihua City with the largest area (27569.24km^2), its trajectory operation intensity is only 0.54 times/km2, which is the lowest in 14 cities. Fig. 4 (3) is the results of the analysis of gross domestic product combined with data intensity in Hunan Province in 2018 [20] .The GDP data are from the sum of the GDP of each district and county in the city of Hunan Statistical Yearbook in 2019. The blue line in Fig. 4(3) is the logarithmic model to fit GDP and the number of trajectory operations, and the orange line is ± 2 times the median error. Fig. 4(2) further reflects the distribution of the number of county-scale trajectory operations. The distribution characteristics of trajectory operation data at city and county levels are more obvious. The number of trajectory operations in Changsha-Zhuzhou-Xiangtan region is the largest, and the region with the smallest number of trajectory operations is mainly located in Jingzhou Huitong in Huaihua, Suining Chengbu counties in Shaoyang. The reasons for the low density of trajectory operations in these counties are the constraints of economic development and the existence of single base stations and other mapping reference services or methods. Based on the spatial distribution characteristics of the trajectory operation density at the city and county levels, it can be seen that there is a significant relationship between the trajectory operation density and the local production economic index, especially for the areas with rapid urbanization development dominated by the first and second industries.

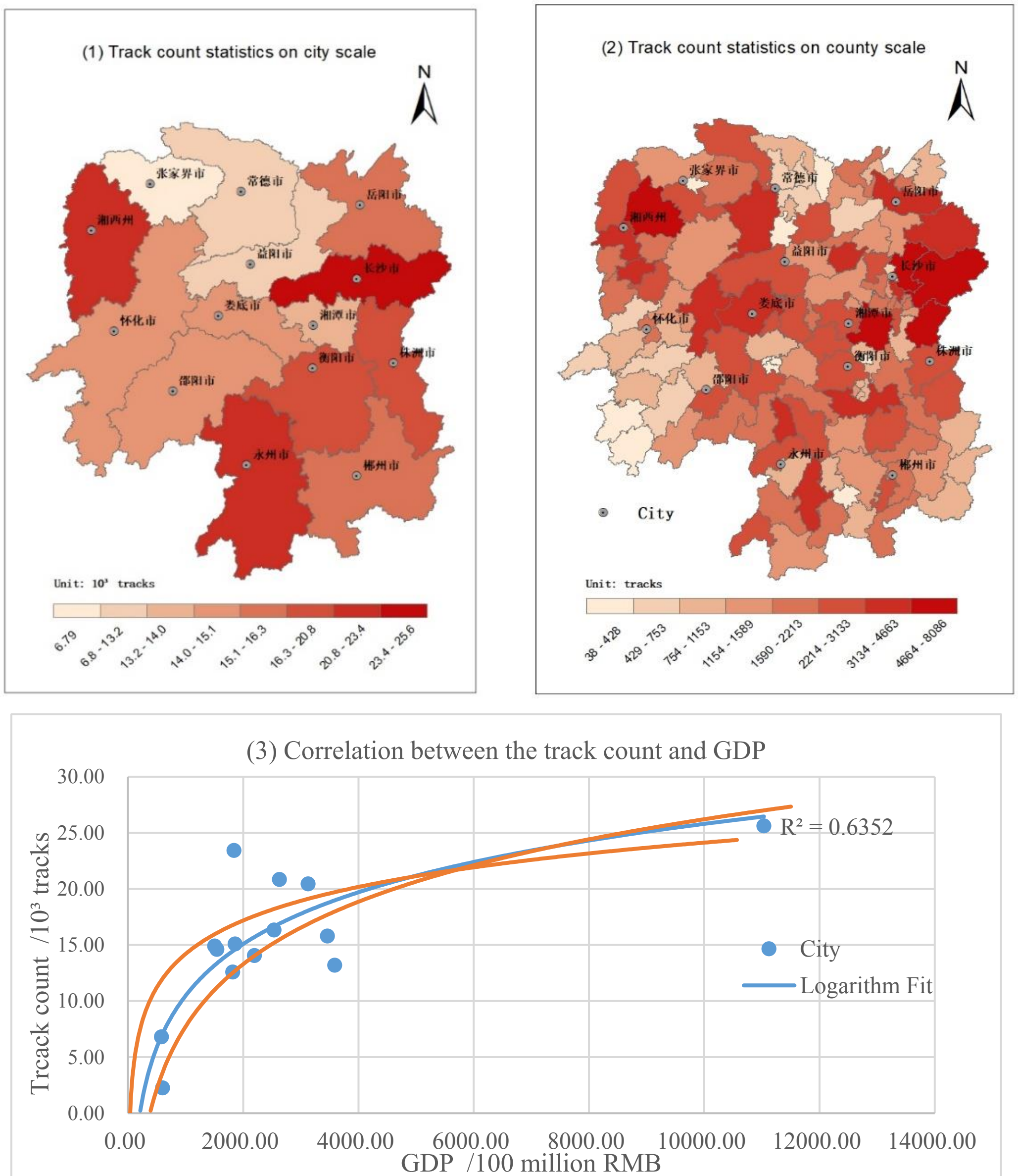

Fig. 4. Correlation curves between urban / county scale spatial distribution and trajectory number and economic development

C. Analysis of spatial-temporal change

The analysis of spatial and temporal evolution characteristics is reflected in the 24 hours of a day, the evolution of the hot areas of mapping reference service trajectory in Hunan Province is as shown in Figure 5.

According to the evolution characteristics of user activity in Figure 5, 24 hours are divided into six stages. The first phase is [0 o ' clock, 6 o ' clock). During this period, there ‘s basically no use of mapping reference service system users in the province. The second stage is [6 o ' clock, 8 o ' clock). At this time, sporadic users begin to perform trajectory operations. Sporadic active areas are also concentrated in the urban areas of Changzhutan and other cities. At this stage, users are mostly in equipment testing, planning, control point calibration and other preparatory work. The third stage is [8 o ' clock, 12 o ' clock). In this stage, the trajectory operation density ushered in the first outbreak, and the operation density point area expanded from the center of the previous stage to the surrounding. From 09 : 00 to 11 : 00, the dense area reached the first peak. In the peak period, the location comparison between dense and non-dense areas is quite obvious. The former is concentrated around the city, and the latter is concentrated in the mountain / lake area. At 11 : 00, the concentration point gradually subsided and gradually entered the next stage. The fourth stage is [12 o ' clock, 14 o ' clock), the trajectory intensive range of this stage reaches a relatively stable state, which is the outbreak preheating of the fifth stage. The fifth phase is [14 o ' clock, 18 o ' clock). Compared with the first outbreak, the density and scope of the second outbreak were weakened. It is the growth period of the second outbreak from 14 : 00 to 15 : 00, and it will reach the second peak in a day at the end of 15 : 00. The decrease of the second peak will continue until the end of 17 : 00. The number of trajectory operations in two bursts (8 hours) accounted for 81.47 % of the total number of trajectory operations in 24 hours. The sixth stage is [18 o ' clock, 24 o ' clock]. During this period, the intensity and scope gradually weakened and transited to the first stage of the second day cycle.

Conclusion

This paper uses the trajectory data extracted by HNCORS service in 2018 to extract and analyze the user trajectory characteristics. The results show that user trajectory operation and fixed solution proportion index are significantly related to the life law of field surveying and mapping. The overall distribution of trajectory operation in the whole province is characterized by the decrease from the middle east to the west, the decrease from the economically developed areas to the economically poor areas, and the decrease from the plain hilly areas to the mountain / lake areas. Among them, the suburbs of towns are relatively dense areas of trajectory operation. The construction of surveying and mapping related infrastructure and the improvement of fixed solution ratio in mountainous areas such as Xuefeng Mountain and Wuling Mountain are

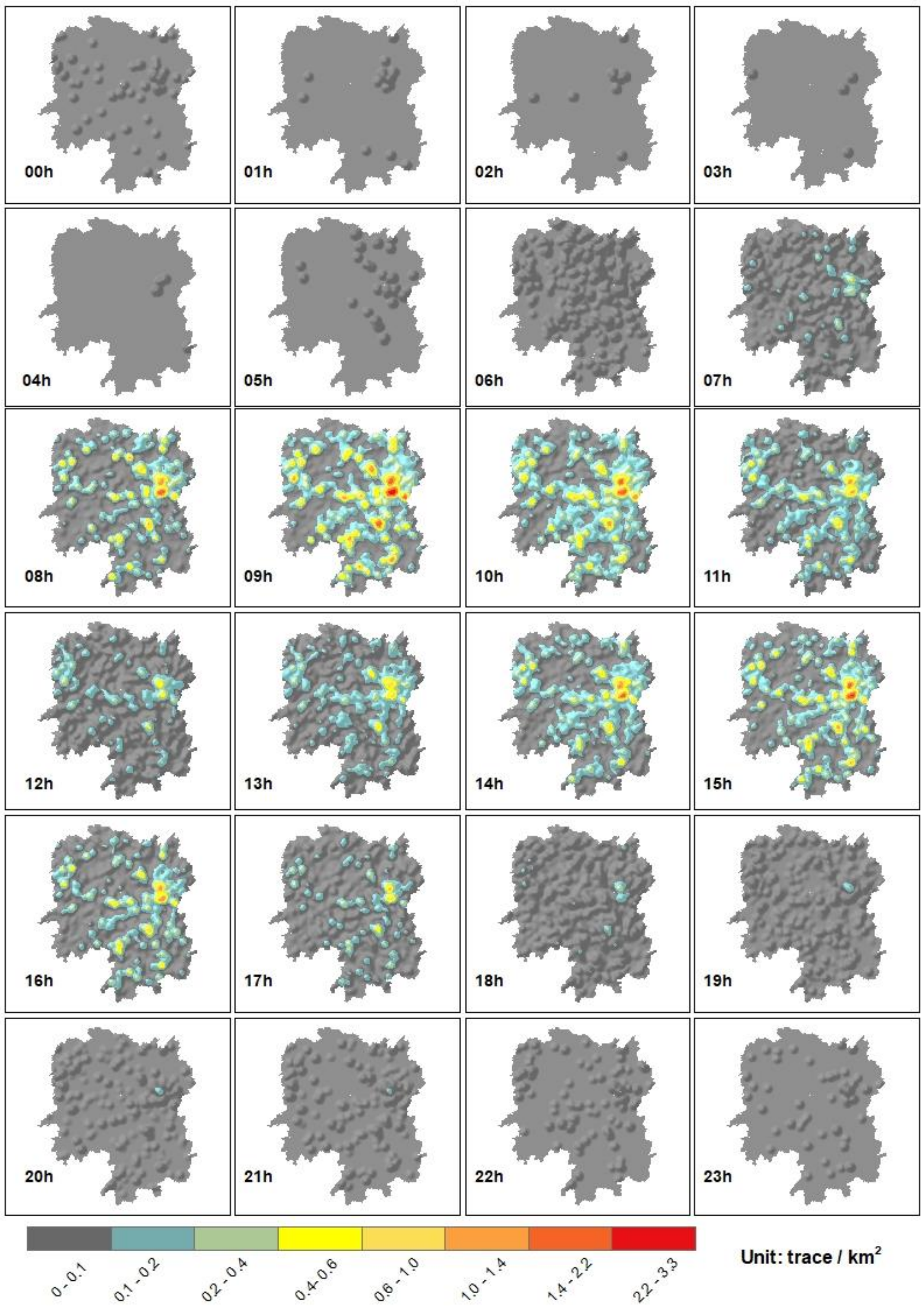

Fig. 5. The change diagram of user spatial distribution hotspots at 24 - hour scale

important directions for the next planning. From the city and county scale analysis, it can be seen that there is a significant correlation between user trajectory operation and local economic indicators. The southern Huaihua and southern Shaoyang are the main directions of application promotion. By analyzing the spatio-temporal sequence diagram of user trajectory operation data, the evolution process is divided into six stages, which is manifested in the continuous growth and decline around Changsha-Zhuzhou-Xiangtan region and 11 urban areas in any day. The analysis and research based on trajectory operation data can provide reference for HNCORS optimization and improvement. At the same time, it is of great significance to further use its characteristics of large amount of data, strong foundation and early stage, and combine with the application in related fields, such as remote sensing change pattern, social and economic development index, and engineering project progress data for prediction and simulation.

REFERENCES

[1] Zhao Z J, Ji G L. 2017. Research progress of spatial-temporal trajectory classification[J]. Journal of Geo-information Science, 19(3):289-297, in press.

[2] Gao Q, Zhang FL, Wang RJ, Zhou F. Trajectory big data: a review of key technologies in data processing[J]. Ruan Jian Xue Bao/Journal of Software, 2017,28(4):959−992, in press.

[3] XU Jia-jie, ZHENG Kai, CHI Ming-min, Zhu Yangyong, Yu Xiaohui, Zhou Xiaofang. Trajectory big data: data, applications and techniques [J]. Journal on Communications, 2015, 36(12):97-105, in press.

[4] ZHANG Jian-qin, QIU Pei-yuan, DU Ming-yi. Mining method of travel characteristics based on spatio-temporal trajectory data[J]. Journal of Transportation Systems Engineering and Information Technology, 2014, 14(6): 72-78, in press.

[5] Domingo-Ferrer J, Trujillo-Rasua R. Microaggregation- and permutation-based anonymization of movement data[J]. Information ences, 2012, 208(none):55–80, in press.

[6] KONG Yangxin, JIN Cheqing, WANG Xiaoling. Population flow analysis based on cellphone trajectory data [J]. Journal of Computer Applications, 2016, 036(001):44-51, in press.

[7] TANG Luliang. NIU Le. YANG Xue, X. Zhang. Urban intersection recognition and construction based on big trace Data[J]. Acta Geodaetica et Cartographica Sinica. 2017,46(6): 770-779, in press.

[8] FENG Hui-fang, BAI Feng-shan, XU You-ji. Urban traffic perception and critical node identification of road network based on trajectory big data[J]. Journal of Transportation Systems Engineering and Information Technology, 2018, 18(3): 42-47, in press.

[9] TANG Luliang, ZHENG Wenbing, WANG Zhiqiang, Xu Hong, Hong Jun, Dong Kun. Space time analysis on the pick-up and drop-off of taxi passengers based on GPS big data.[J]. Journal of Geo-Information Science, 2015, 17(10):1179-1186, in press.

[10] FENG Huifang, YANG Wenliang. Identification of urban functional areas based on GPS trajectory and POI data with association rules [J]. Journal of Geomatics Science and Technology，2020. 37(4)：414-420, in press.

[11] YANG Wenjun, MENG Fandong, LI You, Tang Mengmeng. Research on passenger travel characteristics based on OD data analysis of vehicle trajectory[J]. GEOMATICS & SPATIAL INFORMATION TECHNOLOGY, 2020, 43(S1):144-147，150, in press.

[12] WU Qunyong, ZHANG Liangpan, WU Zufei，Identifying city functional areas using taxi trajectory data [J]. Journal of Geomatics Science and Technology，2018，35(4) : 413-417, in press.

[13] Zhaowei Q, Xin Wang, Xianmin Song, Zhaotian Pan, Haitao Li. Location optimization for urban taxi stands based on taxi GPS trajectory big data[J]. IEEE Access, 2019, PP(99):1-1, in press.

[14] ZHU Tingting, TU Wei, YUE Yang, Zhong Chen, Zhao Tianhong, Li Aiuping, et al. Sensing urban vibrancy using geo-tagged data[J]. Acta Geodaetica et Cartographica Sinica, 2020,49(3): 365-374, in press.

[15] XU Yao, LI Zhuoran, MENG Jinlong, Zhao Lipo, Wen Jianxin, Wang Guilin. Extraction method of marine lane boundary from exploiting trajectory big data[J]. Journal of Computer Applications，2019, 39(1):105-112, in press.

[16] YANG Wei, AI Ting-hua. Road centerline extraction from crowdsourcing trajectory data[J]. Geography and Geo-Information Science, 2016, 32(3):1-7, in press.

[17] HUA Liangchun, XUAN Yong, YIN Haohua, Ao Minsi, Chen Chunhua. Tests and analysis on upgraded BDS-compatible HNCORS network[J]. Journal of Navigation and Positioning, 2017, 5(3):62-66,71, in press.

[18] Li C, Zhang Y , Ao M , Liu Q, Tang C. Integrated services management and statistic system for HNCORS[C]. 2017 IEEE 2nd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA). IEEE, 2017.

[19] Mingxu Dong, Minsi Ao, Xiangqiang Zeng, Chunhua Chen, Chenxi Li. Big data analysis based on HNCORS RTK services [C]. The 9th China Satellite Navigation Conference. The organizing committee of China Satellite Navigation Conference, 2019.

[20] Hunan Provincial Bureau of Statistics. Hunan statistical yearbook[M]. China Statistics Press, 2019.

[21] AN Yan-hui, WANG Yong. Correlation analysis on CORS user frequency and local economy[J]. Science of Surveying and Mapping, 2015, 40(15):115-118, in press.