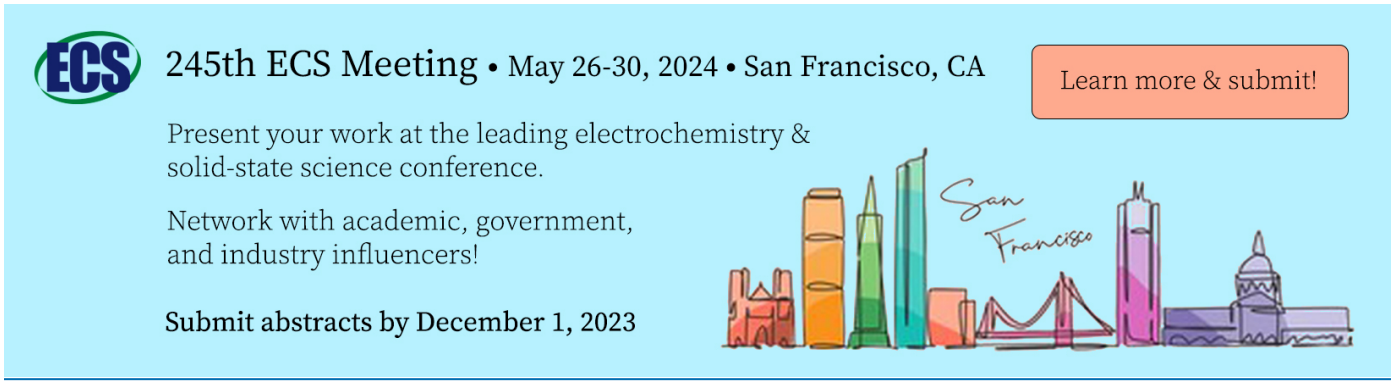
Data Cleaning of Speed Monitoring Based on Driving Behavior Characteristics for Commercial Vehicle



IOP Conf. Series: Materials Science and Engineering **392** (2018) 062156 doi:10.1088/1757-899X/392/6/062156

Data Cleaning of Speed Monitoring Based on Driving Behavior Characteristics for Commercial Vehicle

Chuan SUN*, Jianyu WANG Lian XIE Duanfeng CHU, Liqun LIU

College of Transportation, Huanggang Normal University, Huanggang, Hubei 438000, China

School Architecture and Transportation Engineering, Guilin University of Electronic Technology, Guilin, Guangxi 541004, China

Intelligent Transportation Systems Research Center, Wuhan University of Technology, Wuhan, Hubei 430063, China

Email: sunchuan33@foxmail.com

Abstract: There are lots of speed monitoring data in the commercial vehicle monitoring system, and the accuracy of the data depends on accuracy of the analysis for driving behavior characteristics. In this paper, the driving behavior characteristics for commercial vehicles is analyzed so that six common types of errors for vehicle monitoring data are obtained, such as false date for time information, outlying high/low speed values, zero-speed signal drift, false zero-speed records, outlying acceleration/deceleration values, noise jamming data. Then data cleaning rules are made. Finally, we compare data cleaning effect through typical data source case. Results show that the data cleaning method is used for fitting the data, and correct the error signal from the vehicle driving curves more correspond with the real-world driving and road conditions.

1. Introduction

With the broad application of Global Position System (GPS), Geography Information System (GIS), and wireless communication technology in the safety management of road transportation and the improvement of relevant rules and policies, that is to say, commercial vehicles must be all included in the monitoring platforms of enterprises and government. Based on technology, law and regulation, the real time monitoring of commercial vehicles has been realized in most areas, and lots of monitoring data of vehicles have been accumulated, such as the latitude and longitude coordinates, speeds and the azimuth angle of vehicles [1]. At present, most commercial vehicles monitoring systems mainly utilize GPS satellite positioning system to determine the geographical location of vehicles, realizing various functions of carrying out remote monitoring of vehicles by using the Global System for Mobile Communications (GSM) and General Packet Radio Service (GPRS) network to transmit positioning data and control commands between vehicles and monitoring centers [2].

Due to GPS satellite positioning, atmosphere, operational error and signal interruption, some errors will occur to original data collected by GPS [3]. Therefore, differences inevitably exit in analytical laws of vehicle driving behavior, which is based on the original data. It is necessary to cleanse original GPS data. The driving speed of a vehicle is an important factor influencing the driving behavior. Besides, it is the core data source of vehicle monitoring system platform. Therefore, vehicle monitoring data is chosen as the primary object. Ding et al. applies Wavelet Analysis, which is a

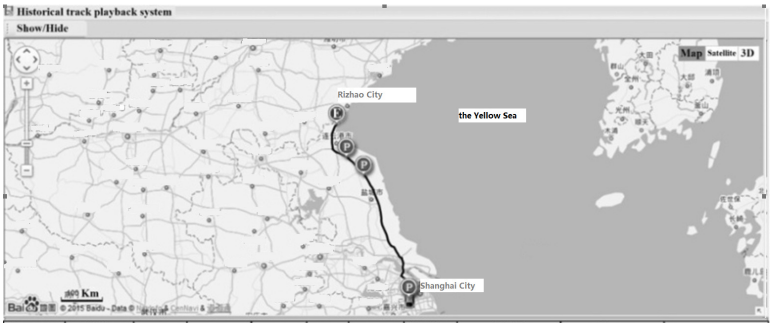
mathematical tool, to GPS data denoising treatment according to basic theory of Wavelet Analysis and GPS signal noise characteristics [4]. Liu et al. introduces multiple-telemetry equipment recording information docking method, delayed channel measurement information supplement method and other data repair technologies into GPS information measurement process [5]. Guo et al. proposes an approach of GPS supported by clock error prediction model [6]. Miura et al. makes error GPS data repaired according to adjacent GPS data characteristics and geometric form of building [7].

Few GPS data processing researches focus on driving behaviors of running vehicles and developing detailed data clearing rules. The formulate data clearing rules can improve the quality of speed monitoring data source, which will provide reliable data sources for future analysis of vehicle driving behavioral characteristics and development of driver's safety operation norms.

2. Error Analysis of Commercial Vehicle Monitoring Data

2.1. Monitoring Interface Platform for Commercial Vehicle

Vehicle monitoring data comes from operated vehicle monitoring platform which is developed by an Internet of Vehicle enterprise, which gathers the vehicle's geological position with a GPS navigation system. Through GSM and GPRS, the networks covering the whole nation, GPS data and control order are passed between vehicle and monitoring center, thereby realizing functions of vehicle remote monitoring across the country. Vehicle monitoring data is the real-time positioning data of operational truck which is equipped with vehicle terminal equipment (GPRS communication module, GPS positioning module, satellite antenna, vehicle sensor, etc.). It reports vehicle's equipment condition, operational state, geological position (latitude and longitude), instantaneous speed, operating azimuth and other real-time information at the frequency of 1Hz. The Figure 1 shows the real-time monitoring interface platform.



Sequence Number	Plate Numbers	Time	Longitude	Latitude	Speed	Angle	Position Status	Total Kilometres(KM)	Once Kilometres(KM)	Vehicle Status
20032	SU BA6372	2015/7/8 5:33:52	121.15155	31.6125	45	W.b.S. 13degree	Location	72108.3	9.4	ACC,hom,brake and so on

Fig. 1 Real-time Monitoring Interface Platform

2.2. Error Analysis of Monitoring Data

There are some errors of GPS data returning the vehicle terminal to the control center. If the data are used directly without clearing, the accuracy of vehicle driving behavior analytical results will be affected. The causes of vehicle GPS data errors mainly include GPS equipment failures, signal blocked by obstacles and human reasons. GPS equipment failure of operating vehicles mainly include: the same data or error data are returned within a period of time; stability failure of wireless communication devices and GPS receivers, leading to delay of receiving original data, etc. These failures gravely affect analytical accuracy. Data collected under failure should be deleted totally; GPS emitting device of operating vehicle are easily blocked by high-rising buildings when emitting signals, resulting in failed or delayed signal emitting, or when vehicle enters into underground parking lot and tunnel, the device is unable to emit data to the control center or signal is interrupted. These data need to be identified and cleared; human cause of GPS data abnormality is driver re-starts the equipment

within a short period of time, producing ineffective data, which cannot be used or analyzed. These data should be deleted.

Operating vehicles demonstrate distinctive driving behavior characteristics. For example, the driver has the limited speed when driving on the way; speed acceleration and deceleration has a range according to the vehicle's operating performance; besides, operating vehicles run stably since all drivers are professionals. Therefore, the characteristics of operating vehicle driving behavior can be an important principle for formulating the vehicle monitoring data clearing rules.

3. Cleaning Rule of Speed Monitoring Data for Commercial Vehicle

Data clearing rule converts "dirty data" into data satisfying data quality requirement or application requirement after analyzing the presence cause and the existence form of "dirty data" and detecting them with existing technical means and methods, thereby improving data quality of the data set [8]. Data clearing takes use of the backtracking idea, analyzing dirty data from the origin and inspecting every process of data set flow. Then, it identifies the algorithms, rules and strategies. Finally, these algorithms, rules and strategies are applied to data set to find out and clear "dirty data" (see Figure 2).

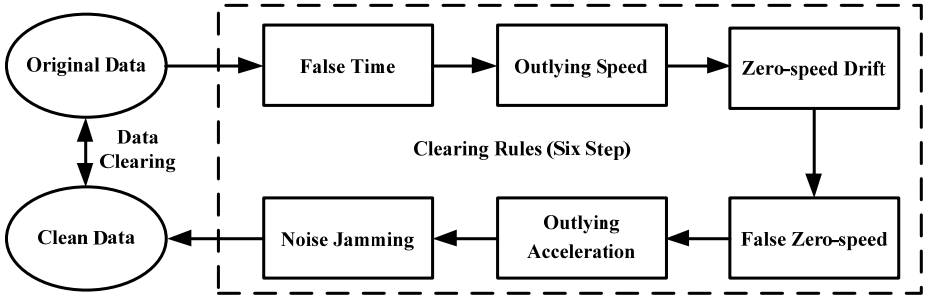


Fig. 2 Data Cleaning of Speed Monitoring Flow Chart

3.1. Data Cleaning for False Time

Clearing of time information error data is the first step for clearing the whole operating vehicle speed monitoring data. Correct sequence of time information constitutes the premise of follow-up data analysis. As shown by vehicle monitoring data collected by GPS, time information error data mainly include repetitive recording of information at certain moment and reversed time. If the time of the former data is equal to or bigger than that of the latter one, i.e., the difference of the latter data time and the former one is equal to or smaller than 0. Time information error data will cause to wrong calculation result relating to acceleration value. If speed monitoring data is repeated, the value calculated according to the acceleration formula will be infinite large. Thus, clearing repeated data information and time-reversed data is the first step for clearing operating vehicle speed monitoring data.

Clearing rules: labor confirms whether the starting time of data is right or not first, and then calculates the time difference of former and later data of data source. If the time different is equal to or smaller than 0. Then, it is considered that time information of the former data is wrong, which will be deleted or the two data will be replaced (such as Table 1).

Table 1 Data Cleaning Result of False Time (First Step)

Plate Number	Time	Longitude °	Latitude °	Speed km/h	Direction	Clearing Rule
Su BXXX	2015-03-15 09:12:20	120.49547	31.63584	30	EbS 1°	—
Su BXXX	2015-03-15 09:12:21	120.49556	31.63584	31	DE	Repetitive Recording Delete the Row Data
Su BXXX	2015-03-15 09:12:21	120.49556	31.63584	31	DE	
Su BXXX	2015-03-15 09:12:22	120.49566	31.63584	34	DE	—
Su BXXX	……	……	……	……	……	……
Su BXXX	2015-03-15 09:45:44	120.49736	31.63584	62	EbN 1°	—
Su BXXX	2015-03-15 09:45:46	120.49774	31.63585	65	EbN 2°	Time Reversal Swap Two Row Data
Su BXXX	2015-03-15 09:45:45	120.49755	31.63585	64	EbN 2°	
Su BXXX	2015-03-15 09:45:47	120.49793	31.63585	66	EbN 2°	—

3.2. Data Cleaning for Outlying Speed

The second step of clearing operating vehicle speed monitoring data can be clearing over-threshold speed data. When operating vehicles are in the running process, GPS will causes the speed at certain point of time returning to the operating vehicle monitoring system platform to deviate from actual value under impacts of some factors, always obviously exceeding the maximum speed of operating vehicles. It demonstrates distinctive characteristic of jumping. Relevant statistical analysis shows that the maximum speed of express way is 100km/h [9]; the Regulation on the Implementation of the Law of the People's Republic of China on Road Traffic Safety, the maximum speed of some operating vehicles is 100km/h; because of driving behavioral characteristics of operating vehicles, their maximum speed is generally lower than 120km/h even in the speeding state [10]. Therefore, considering that operating vehicle may runs at the speed exceeding the limit, the threshold of the operating vehicle's maximum speed is set to be 130km/h.

Clearing rules: data of speed exceeding the set threshold value in the data source is judged as error data of speed exceeding threshold and repaired with the proper interpolation method. For example, Figure 3 shows a commercial vehicle presented 2 error data of speed exceeding the threshold (130km/h) within 100s intervals (the first data clearing has been completed). Data are repaired with three Hermite interpolation values [11].

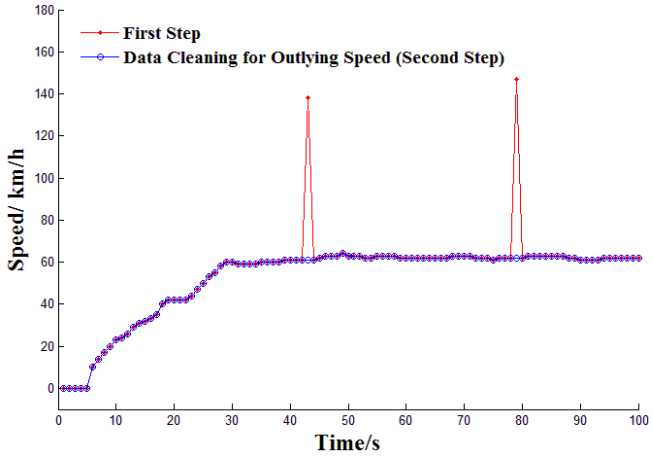


Fig. 3 Data Cleaning Result of Outlying Speed (Second Step)

3.3. Data Cleaning for Zero-speed Drift

Clearing data of zero shift of speed can be the third step of clearing operating vehicle speed monitoring data. When the operating vehicle stops transportation or is in the idle state for a long time, the speed GPS emitting to the monitoring system platform should 0 continuously. However, due to too fast response of vehicle terminal or strong interference signal outside, occasionally the speed jumps out of 0. Hence, to improve the quality of data source, it is necessary to correct vehicle speed zero drift data.

Clearing rules: non-zero data before and after which the speed is 0 is as suspected zero drift data. Then, it checks whether the latitude and longitude of the former data, latter data and suspected data are consistent or not. If consistent, the suspected data is determined as speed zero drift data and the speed data is corrected as 0. Figure 4 shows a commercial vehicle presented 3 error data of speed zero shift within 150s intervals (the second data clearing has been completed).

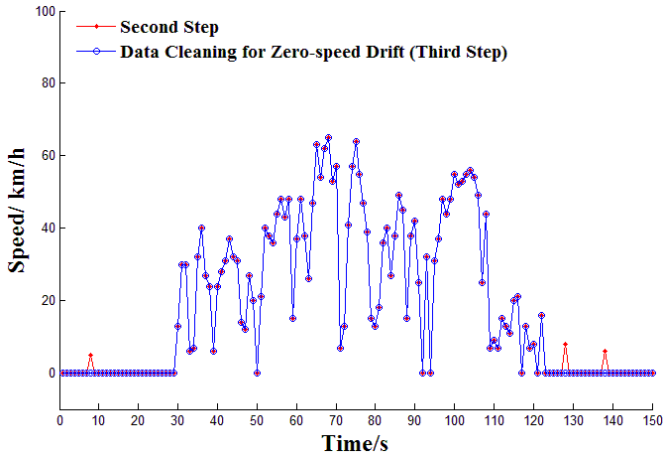


Fig. 4 Data Cleaning Result of Zero-speed Drift (Third Step)

3.4. Data Cleaning Result for False Zero-speed

Clearing speed zero error data can be taken as the forth step of clearing operating vehicle monitoring data. When the operating vehicle's GPS receives strong interference signals from outside within a short period of time, it will cause the speed emitted to the monitoring system platform jumps to 0, but the former and latter data is normal, which is called speed zero error data.

Clearing rules: define the speed of 0, before and after which the speed is 0, as zero error data. Data are repaired with three Hermite interpolation values. Figure 5 shows a commercial vehicle presented 3 speed zero error data within 100s intervals (the third data clearing has been completed).

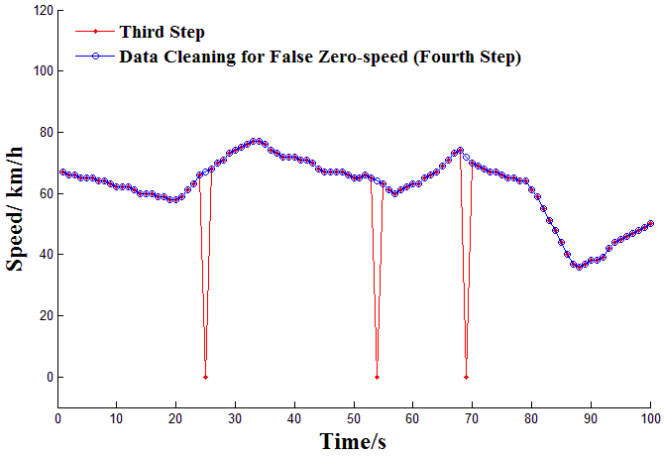


Fig. 5 Data Cleaning Result of False Zero-speed (Fourth Step)

3.5. Data Cleaning for Outlying Acceleration

Clearing data of speed exceeding threshold can be taken as the fifth step of clearing operating vehicle monitoring data. When the operating vehicle is running, the speed at certain point of time that GPS returns to the operating vehicle monitoring system platform presents greater difference with the former and latter speed, which is different from the previous types of errors. Then, it is defined as error data of speed exceeding threshold. According to the driving behavioral characteristics of operating vehicles, the acceleration of operating vehicle in typical driving condition generally does not exceed 2.5 m/s^2 [12]. On the other hand, deceleration of most operating vehicles on dry pavement does not exceed 3.5 m/s^2 [13]. Therefore, speed change rate of -3.5~2.5 m/s^2 is set as the threshold value.

Clearing rules: the data whose speed difference to the former and latter data exceeds threshold is determined as error data of speed change exceeding threshold. Proper interpolation method is used for repair. And the repaired data are further verified as well. If the value still exceeds the threshold, it is corrected as the data calculated based on the threshold. Figure 6 shows a commercial vehicle presented 2 error data of speed exceeding threshold within 100s intervals (the fourth data clearing has been completed).

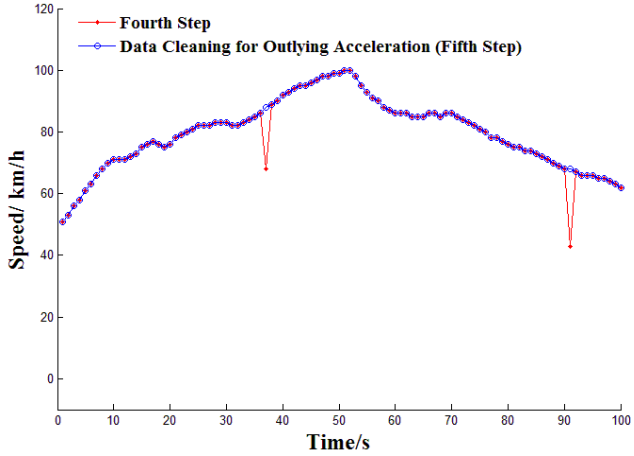


Fig. 6 Data Cleaning Result of Outlying Acceleration (Fifth Step)

3.6. Data Cleaning for Noise Jamming

Clearing noise interference data can be taken as the sixth step of clearing operating vehicle monitoring data, which is the final step. Noise interference data is generated when GPS receives interference, which causes speed data and position data to fluctuate (random error). The error value is decided by instrument performance and operating environment. Generally, the measurement noise value is much smaller than deviation values. Therefore, it can used as the last step to de-noise of the entire data source, making speed monitoring data better meet driving behaviors of operating vehicles. De-noise can be realized through Kalman filter and wavelet transform [14].

Clearing rules: after the data source is cleared by the five steps, the whole data source is de-noised with the wavelet transform. Figure 7 shows a commercial vehicle's de-noise clearing results within 100s intervals (the third data clearing has been completed).

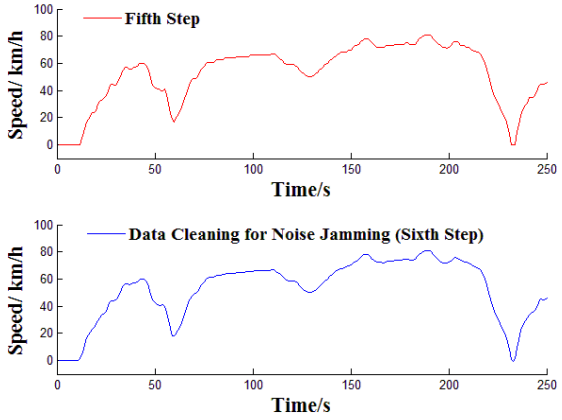


Fig. 7 Data Cleaning Result of Noise Jamming (Sixth Step)

4. Case Study for Data Cleaning

Speed monitoring data of data clearing case study comes from GPS satellite positioning system which collected from an operating truck's vehicle terminal when it was running on a national highway in Jiangsu on March 6. 2015. The operating truck is owned by a large logistics company in Shanghai. It totals about 80,000 original data. It is noted that when the operating vehicle stops transportation in a period of time, the positioning of the vehicle, i.e., latitude and longitude, may remain the same or fluctuate within a small range. Besides, the instant speed of the vehicle continues to be 0. In the process of original data processing, the research only chooses those data that does not continue to be 0. According to the specific characteristics of the data source, i.e., characteristics of the operating truck's driving behavior, clearing rules are established for the data source.

The selected data source is cleared according to the developed operating vehicle's speed monitoring data clearing rules. The distribution of error information data is counted and consideration is taken to the sixth step of clearing rules—de-noising of the entire data source, which demonstrates great randomness. Therefore, the first five steps of error information are counted. A total of 29814 groups of data in the normal transportation situation (including idle state) are chosen, including 1258 groups of error data, accounting for 4.219% of the total data source. Of the information error data source, time errors take up for about 73%, followed by speed exceeding threshold, with an proportion of 14%. For the remaining error data, the proportions are similar.

5. Conclusion

Vehicle monitoring data collected from an operating vehicle monitoring system platform are much richer. However, because of objective reasons facing collection equipments, vehicle monitoring data may have some errors. This paper sets up clearing rules for six error speed monitoring data according

to the driving behavioral characteristics of operating vehicle, i.e., time information error data, data of speed exceeding threshold, speed zero drift data, speed zero error data, speed change exceeding threshold data and noise interference data. The research selects speed monitoring data source of an operating vehicle on a national highway in Jiangsu as the typical case, summarizing distribution rules of general error information. At the same time, it compares and analyzes the effect of data clearing rules, finding that the quality of cleared data really reflects the actual operating state of the operating vehicle, meeting the driving behavioral characteristics of the operating vehicle to the largest degree. Thereby, the research lays a research foundation for further analyzing the vehicle's driving behavioral characteristics as well as formulating the corresponding driver's safety operational norms.

Acknowledgment
This study is sponsored by the National Undergraduate Training Program for Innovation and Entrepreneurship (201710514011); the Guangxi Higher Institutions Scientific Research Project (ky2016YB151); the National Natural Science Foundation of China (51675390); the Hubei Provincial Natural Science Foundation (2016CFB468) and the Fundamental Research Funds for the Central Universities (2017II50GX, 2017IVB075).

References
[1] YANG Yang, YAO Enjian, PAN Long, et al. Taxi Route Choice Behavior Modeling Based on GPS Data[J]. Journal of Transportation Systems Engineering and Information Technology, 2015,15(1):81-86. (in Chinese)
[2] Hu Jihua, YANG Guigen, LIU Zhibin. Study on the Frame Work of Commercial Vehicle Monitoring System for Road Traffic Safety[J]. China Safety Science Journal,2007,17(8):171-176. (in Chinese)
[3] KANG Silin, LI Yuqiang. Error Analysis of GPS Positioning[J]. Astronomical Research & Technology, 2010,7(3):222-230. (in Chinese)
[4] DING Wen. Study of GPS Data De-noising Method Based on Wavelet and Kalman Filtering[D]. Shenyang: Northeastern University, 2009. (in Chinese)
[5] LIU Futi, TANG Haibo, ZHANG Xiangcheng. Application of Data Repairing Technique for GPS Data Processing[J]. Journal of Telemetry,Tracking and Command, 2012,05:47-50. (in Chinese)
[6] GUO Rongchang, JIN Wenjun, FAN Duowang, et al. Reseach on Positioning Method under GPS Incomplete Constellation[J]. Journal of China Railway Society, 2015,37(3):42-45. (in Chinese)
[7] Miura S, Kamijo S. GPS error correction by multipath adaptation[J]. International Journal of Intelligent Transportation Systems Research, 2015, 13(1): 1-8.
[8] CAO Jianjun, DIAO Xinchun, CHEN Shuang, et al. Data Cleaning and its General System Framework[J]. Computer Science, 2012,39(11A):207-211. (in Chinese)
[9] PEI Yulong, CHENG Guozhu. Research on Operation Speed and Speed Limit for Freeways in China[J]. Journal of Harbin Institute of Technology, 2003,35(2):168-172. (in Chinese)
[10] LIU Jianpei, LI Sheng, GAO Jinsheng. Analysis and Optimization for Operating Speed Model on Straight Section on Highway[J]. China Journal of Highway and Transport, 2010,23(S2):83-88. (in Chinese)
[11] CHEN Zhijun, WU Chaozhong, LV Nengchao, et al. Vehicle Trajectory Reconstruction Based on Improved Cubic Hermite Interpolation[J]. Journal of Transport Information and Safety, 2013,31(6):43-46. (in Chinese)
[12] HOU Cong, WANG Hewu, OUYANG Minggao. A Comparative Study on Consumption the Evaluation Methods of PHEV Energy Between China and the U.S.[J]. Automotive Engineering, 2015,37(1):1-8. (in Chinese)
[13] Mitschke M, Wallentowitz H, CHEN Yinsan, et al. Vehicle Dynamics[M]. Fourth Edition. Beijing: Tsinghua University Press, 2009. (in Chinese)

[14] LI Honglian, FANG Hong, TANG Ju, et al. GPS Dynamic Filter Algorithm Based on Lifting Wavelet Transformation[J]. Journal of Transportation Systems Engineering and Information Technology, 2013,13(1):43-48. (in Chinese)