Journal of Engineering Science and Technology
Vol. 15, No. 2 (2020) 1429 - 1440

GPS DATA PROCESSING FOR DRIVING CYCLE DEVELOPMENT IN HANOI, VIETNAM

YEN-LIEN T. NGUYEN[1], NGOC DUNG BUI[2,*], TRUNG-DUNG NGHIEM[3], ANH-TUAN LE[4]

[1]Faculty of Transport Safety and Environment, University of Transport and Communications, Hanoi, Vietnam
[2]Faculty of Information Technology, University of Transport and Communications, Hanoi, Vietnam
[3]School of Environmental Science and Technology, Hanoi University of Science and Technology, Hanoi, Vietnam
[4]School of Transportation Engineering, Hanoi University of Science and Technology, Hanoi, Vietnam
*Corresponding Author: dnbui@utc.edu.vn.

Abstract

Real-world driving data collected by the global positioning system (GPS) technology normally contains many errors, especially outlying data points and signal gaps. Consequently, the typical driving cycle, which is constructed based on this data, is badly affected due to the misleading information. In addition, the real-world driving data of vehicles under the densely populated areas tends to contain more error points because of high-rise buildings and other interfering sources. This paper proposes a method for detecting and repairing such errors focusing on buses in the over-crowded city of Hanoi. The GPS data processing procedure consisting of nine steps was designed to normalize the dataset and to minimize errors related to sudden signal loss, abnormally data points, velocity signal drift, and signal white noise for the purpose of typical driving cycle development. Erroneous data points were detected and removed by creating gaps at the appropriate positions in the dataset. Then, the missing data estimation algorithm developed in the literature was used to fill up the missing values in these gaps. In addition, a modified Kalman filter was used as the last filtering step to further denoise and smooth data points. The ratios of errors related to speed drifting and sudden signal loss are high, 3% and 2.5%, respectively.

Keywords: Denoise, Driving cycle, GPS, Least square, Markov, Modified Kalman filter.

1. Introduction

The real-world driving data of vehicles plays a very important role in the local driving cycle development, based on which, country-specific emission factors (CSEFs) can be determined using the emission measurement of vehicles under controlled conditions in laboratories (dynamometer tests). Using this method, CSEFs can be more precisely developed in order to improve the quality of emission inventory in the transport sector [1].

Three main methods used to collect the real-world driving data of vehicles included car chase method, on-board measurement, and global positioning system (GPS) technology [2]. Car chase and on-board measurement came first, but in recent years, the GPS technology has gained more attention. The reason is that GPS allows continuous monitoring of vehicle speed and position [2, 3]. Other benefits of GPS include the ease of use and cost effectiveness. Therefore, it has become the highly appreciated technique for vehicle real-world driving data collection [4]. In Vietnam, as traffic jams occur frequently, especially within the inner city, and thus the use of the GPS device is the most suitable.

GPS technology, however, also has its limitations. As the GPS technology is used to capture the vehicle real-world driving characteristics, some errors often occur on the raw dataset. Typical errors in GPS data include sudden interrupted signal, abnormal data points, velocity signal drift, and signal white noise. These errors result in the lower quality of data, which reduces the accuracy of reconstitution of realistic driving characteristics in the developed typical driving cycle [4]. It would make sense to have this raw data processed to minimize errors while preserving its integrity before using it in the driving cycle development. In this paper, we proposed a GPS filtering process consisting of nine steps, which is used to process the GPS data for the purpose of the typical driving cycle development for bus system in Vietnam. The detecting and repairing of errors in the GPS data are developed in MATLAB, and a modified Kalman filter is used as the last step of the filtration process to de-noise and smooth final signals.

2. Design of GPS data processing

This study is part of our overall research to development of CSEFs for buses in Vietnam. In our study, a GPS device (Garmin etrex vista HCx) with the 1Hz position update rate was used to collect real-world driving data on the fifteen bus routes in the urban areas of Hanoi. The real-world driving data collection was described in detail in our previous study [5]. In this paper, we only focus on the GPS data processing to achieve our overall study purpose.

Through the analysis of common errors in the collected GPS data and the requirement of data used in the driving cycle development, we proposed a filtration process including nine steps as shown in Fig. 1.

2.1. Generating trip segments

The driving cycle is synthesized from the trip segments, in which each trip segment is defined as a movement between the starting and finishing point of each bus route. According to the data collection method described above, the collected GPS dataset consists of many single trips. Thus, in the first step, the collected GPS dataset must be divided into trip segments.

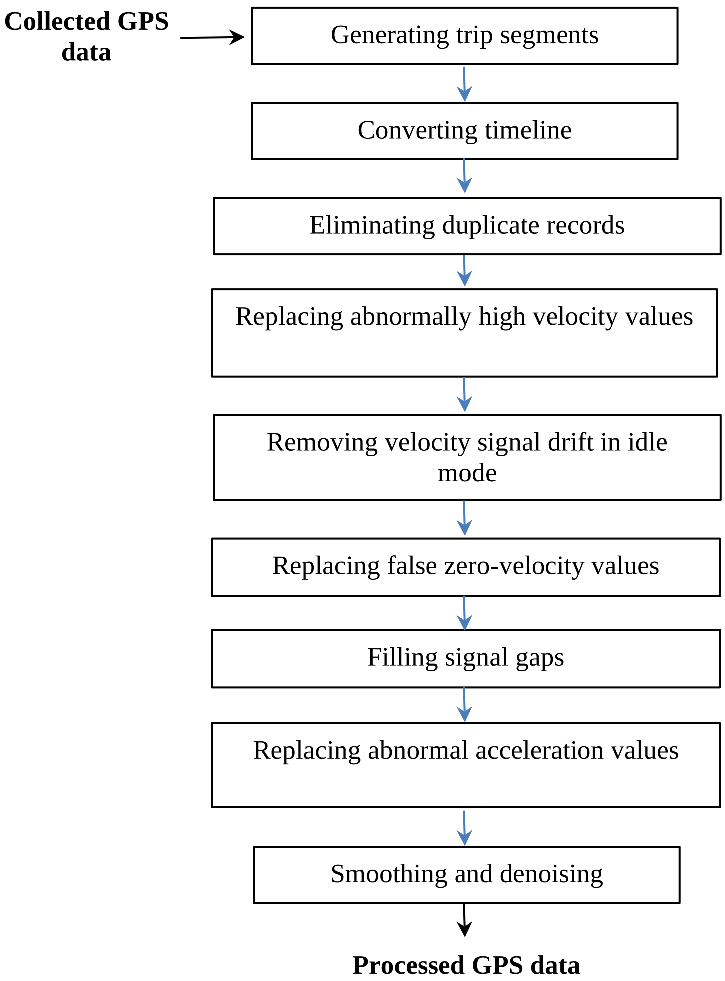


Fig. 1. Overview of steps in GPS data filtration.

2.2. Converting timeline

The data was extracted from the GPS device including of the profile of instantaneous velocity versus time, in which time data was recorded in the form of HH:MM:SS. Therefore, for the purpose of driving cycle development, the GPS data must be converted to timeline of zero at the starting point of the data series. Conversion of the time moment into second was performed first. The time value in second at the first data point was converted equally to zero, after that, the time value at the other data points were determined by the difference between the time value at those data points and the first data point.

2.3. Eliminating duplicate records

To avoid failure in overall next processing steps, the GPS data must be processed to eliminate any data points having duplicate time values. To remove these points, the filtration process first calculated the time steps between successive data points in the raw dataset and removed any point with time step equal to zero [4].

2.4. Replacing abnormally high velocity values

In this step, any data points with abnormally high velocity values are replaced by new values. The highest velocity limit of vehicles in Vietnamese urban areas, i.e., 60 kph, is used to detect points with abnormally high velocity values. The filter was designed to find any points with the velocity value higher than 60kph, and after that deleted them to create gaps. These gaps were amended by new velocity values using the missing data estimation algorithm by Selesnick [6].

2.5. Removing velocity signal drift in idle mode

In the collected velocity profile, it often appears that, velocity values are not zero even if the vehicle is in its idling state, but the engine is still in operation. These velocity values are called the drift of zero-velocity. According to [4], the velocity values in this case are only 0.1 or 0.2 mph. Taking these values in consideration, the filter was designed to first determine microtrips, in which a microtrip is defined as the segment of instantaneous velocity data between two velocity values equal to zero consecutively. After that, the filter checks all velocity values in these microtrips, if all these velocity values are greater than or equal to zero but smaller than 0.2 mph (~ 0.4 km/h), they would be replaced by zero.

2.6. Replacing false zero-velocity values

If a velocity value is equal to zero while the two adjacent velocity values are nonzero, this zero-velocity point is deleted to create a gap in the dataset. After that, this gap is amended according to the similar method as the replacing abnormally high velocity values as mentioned above.

2.7. Filling signal gaps

In the process of signal recording by GPS device, signals are sometimes interrupted due to urban canyon effects; therefore, gaps in the collected GPS data may appear. In this step, the filter was designed to detect and fill the gaps in the collected GPS data. In our overall study, the typical driving cycle was constructed using the Markov theory, an ideal approach for the development of typical driving cycle in recent years. According to this approach, the velocity profile of vehicles should be discretized with the time step of 1s to ensure that the real-world driving data and the velocity profile in driving cycle have the Markov attribute [7, 8]. Therefore, in this study, the filter was designed to detect the time step between two adjacent data points. If it is greater than one second, this position is considered to be a signal gap. After that, the missing data estimation algorithm developed by Selesnick [6] is used to fill gaps in the GPS data.

2.8. Replacing abnormally acceleration values

In the next step of filter process, we repaired the remaining random errors in the dataset by checking the variability of each velocity value over time before performing data denoising and smoothing. The acceleration was used to estimate this variability, in which it was calculated based on the velocity change in each interval of time as in Eq. (1) [7, 9].

$$a(t) = \frac{v_i - v_{i-1}}{3.6Ts} \quad \text{, m/s}^2 \text{ with } \forall t \in [t_{i-1}, t_i) \tag{1}$$

where: Ts is the time step (s). In this study, Ts = 1 s.

The filter was designed to check the acceleration dataset, called the secondary dataset, to ensure that the collected data is in conformity with the efficient operation of the vehicle. The acceleration limits in accordance with vehicle performance were used to find abnormal acceleration data points in the secondary dataset as presented in Table 1. During checking the secondary dataset, if an acceleration data point is found to be outside the chosen limits, the filter would delete the velocity value at the position corresponding with that acceleration position to create a gap. After that, the filter adds a new velocity value, which is estimated based on the algorithm of Selesnick [6], into this gap.

Table 1. Outlying acceleration filtration limits [4].

Vehicle category	Acceleration limits
Light duty vehicle	-17.5 to +17.5 mph/s
Heavy duty vehicle	-8.8 to +8.8 mph/s

(Note: 1 mph/s ≈ 0.45 m/s^2)

All the above steps are repeated until all acceleration values in the secondary dataset fall within the limits as shown in Table 1.

2.9. Smoothing and denoising

In this study, the Kalman filter was used as the last step to smooth and denoise the signals. The Kalman filter is a popular signal smoothing and denoising technique [10-12]. The Kalman filter smooths and denoises data by reforming error values recursively. According to Jun et al. [11], when using the GPS device with frequency of 1Hz to collect the real-world velocity profile, the equations of Kalman filter is abridged as following:

Prediction stage:

$$\hat{x}_k^- = \hat{x}_{k-1} \tag{2}$$

$$P_k^- = P_{k-1} + W \tag{3}$$

Update stage:

$$K_k = P_k^- (P_k^- + V)^{-1} \tag{4}$$

$$\hat{x}_k = \hat{x}_k^- + K_k (z_k - \hat{x}_k^-) \tag{5}$$

$$P_k = (I - K_k) P_k^- \tag{6}$$

where: k denotes the time step, $\hat{X}_k^-$ denotes a prior state estimate vector, $\hat{X}_k$ denotes a posterior state estimate vector, z is the measured data vector, P_k^- is the a prior state estimate covariance matrix, P_k is posterior state estimate covariance matrix, K is the optimal Kalman gain matrix, W is the process noise variance matrix, V is the measurement noise variance matrix. In the theory of Kalman filter, it is assumed that the process and measurement noise have the normal probability distributions: $p(w) \sim N(0,Q)$, and $p(v) \sim N(0,R)$. Therefore, to apply the Kalman filter, it is necessary to determine the measurement noise (R) and the process noise (Q). According to Jun et al. [11], when the rate of data capture is 1 Hz, both of the process noise and the measurement noise ought to be 0.25. In this case, the Kalman filter is called the modified Kalman filter.

3. Evaluation of the designed filter

All the collected GPS data was put into the filter developed in the MATLAB software. After step 1, the whole collected GPS data was splitted into 317 trip segments. This dataset was passed through next steps of the filter to minimize errors, smooth and denoise the signals.

3.1. Ratio of errors in GPS data

The real-world driving data of fifteen bus routes in Hanoi was used to evaluate the efficiency of designed filter as described in Fig. 2. As can be seen from Fig. 2, the errors related to velocity signal drift are the highest, at approximately 3.1%. The second-high position is the errors related to interrupted signals, approximately 2.5%. Total of random errors in the GPS data, which was processed through the steps 4 - 8 of the filter, is approximately 7%. In other words, the filter was well designed to be able to identify random errors in the GPS data before repairing them.

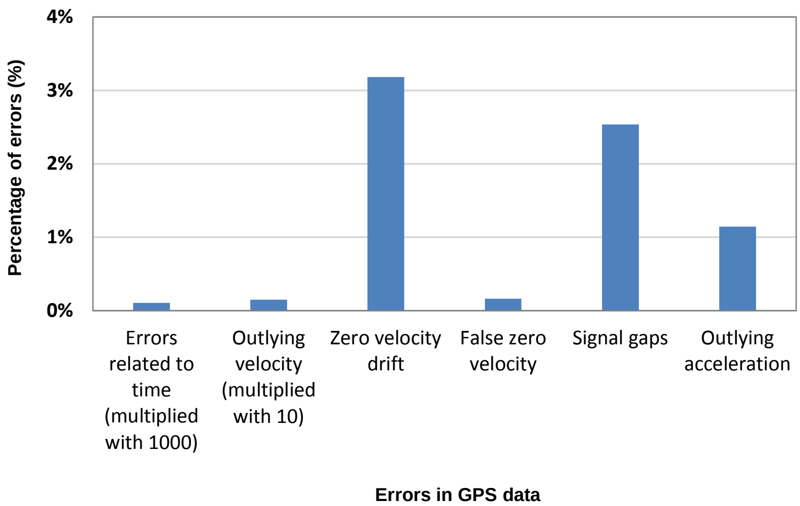
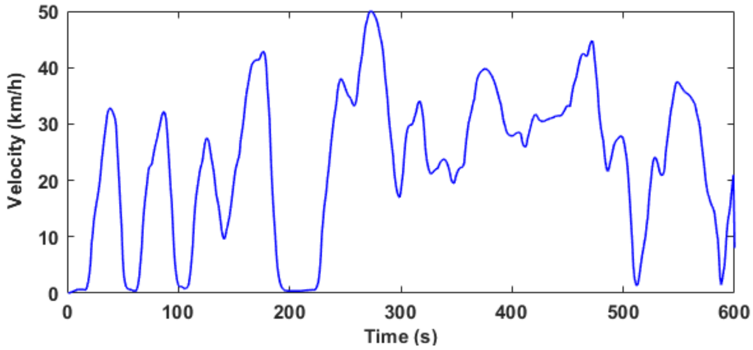


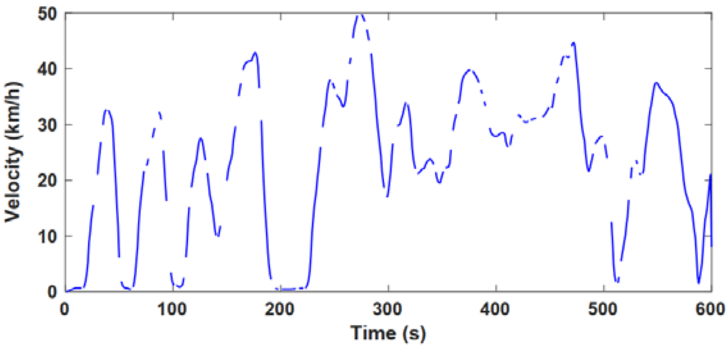
Fig. 2. Percentage of random errors in the GPS data.

3.2. Reliability evaluation of the method for filling data gaps

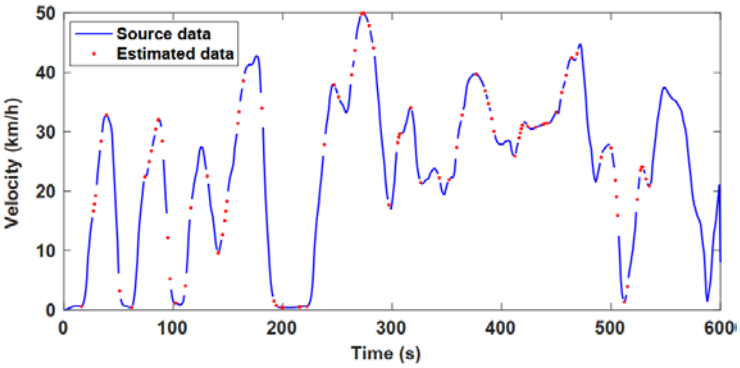
As presented above, error data points detected in the steps 4, 5, 7 and 8 were processed by deleting them to create data gaps. These data gaps were amended using new values obtained by the missing data estimation algorithm developed by Selesnick [6]. This approach is different from the study of Duran and Earleywine [4], in which new data points are obtained using a cubic spline interpolation drawn from neighbouring data points. The reliability of this approach was evaluated by testing on the standard dataset, the velocity-time profile in European Transient Cycle (ETC) - part 1 [13]. One hundred data points of the ETC-part 1 were randomly deleted to create data gaps. Three different methods for data gap amending were used to find the values at 100 gaps above, including: the cubic spline interpolation – the same method as the study of Duran and Earleywine [4], the missing data estimation algorithm developed by Selesnick [6], and the fillgap function in MATLAB. All the performance steps are demonstrated in Fig. 3. The standard deviation of the original data and the estimated data based on three different methods are determined as shown in Fig. 3(d). The root mean square errors (RMSE) between the true data and the estimated data according to three methods are 0.4, 0.2, and 0.6 km/h, respectively. It can be seen that the RMSE determined according to the algorithm of Selesnick [6] is smallest. Therefore, it can be concluded that the algorithm developed by Selesnick to estimate the missing data is successful in amending the GPS data gaps.



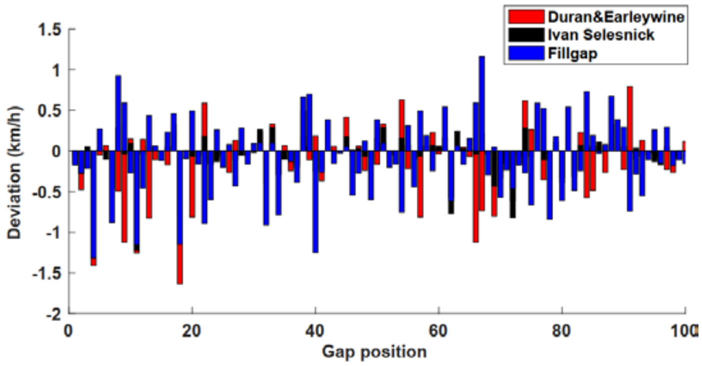
(a) The velocity-time profile of ETC-part 1 (the standard signal).



(b) Creating random gaps.



(c) Amending gaps.



(d) Comparison of different methods.

Fig. 3. Comparison of the data gaps amending methods.

3.3. Evaluation of the Kalman filter

The effectiveness of the Kalman filter, which was used in the final step in the filtration process, was estimated based on statistical parameters for two cases: not using and using the Kalman filter. The evaluation results are presented in Table 2 and Fig. 4.

Table 2. Some statistical parameters for the two cases.

Parameters	Raw data	Treated data	
		Not using the Kalman filter	Using the Kalman filter
Standard deviation of velocity (km/h)	10.94	10.86	10.71
95th percentile of velocity (km/h)	37.0	37.0	36.4

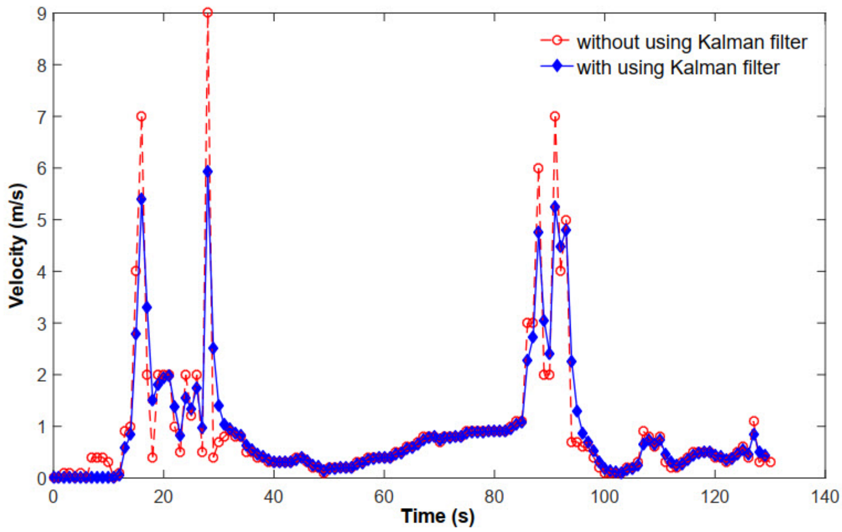


Fig. 4. The smoothing and denoising results of the Kalman filter.

As can be seen in Fig. 4, the data, which was processed through steps 1 to 8 without using the Kalman filter, is not smoothing while the velocity profile in the actual condition is smoothing. Therefore, this data must be smoothed before being used in the development of typical driving cycle. In addition, results in Table 2 and Fig. 4 show that the Kalman filter plays a very important role as the final step to smooth and denoise the data. For both cases (not using and using the Kalman filter), the standard deviations of velocity are smaller than that for the raw data. However, the use of Kalman filter leads to better results and the standard deviation of velocity is smallest. The 95th percentile of velocity in this case is quite similar.

3.4. Evaluation of filter effectiveness

The effectiveness of the filter (consisting of nine steps as described in Fig. 1) was evaluated through the comparison between the raw data and the processed data (see in Table 3 and Fig. 5).

As presented in Table 3, the standard deviations of all driving cycle parameters in the raw data are higher than those in the processed data because there are many random errors in the raw data. However, the value ranges of all parameters in the raw data and processed data are quite similar. In addition, the large difference between raw data and filtered data is detected for the acceleration-related parameter group and the time-related parameter group. This result is suitable for the error ratio as presented on Fig. 2. According to Fig. 2, the outlying velocity-related error is smallest, therefore, the velocity difference between raw data and filtered data is unnoticeable.

The velocity-acceleration frequency distribution of the raw data and the processed data is demonstrated in Fig. 5 with the resolutions of velocity and acceleration of 5 km/h and 0.5 m/s^2, respectively. Results in Fig. 5 show that the velocity-acceleration frequency distribution of the raw data and processed data is quite broad. Therefore, the filter has been well designed to process raw data, as it does not distort the shape of the root data while adding missing data, replacing outlying data, smoothing and denoising the signals in the raw data.

Table 3. Comparison of the driving cycle parameters between the raw data and processed data.

Driving cycle parameters	Raw data			Processed data		
	Average	Median	Standard deviation	Average	Median	Standard deviation
Maximum velocity (km/h)	50.64	49.00	6.88	46.97	46.30	4.09
Average velocity (km/h)	16.72	16.86	1.73	16.67	16.82	1.71
Average driving velocity (km/h)	17.19	17.35	1.91	17.38	17.75	1.83
Maximum acceleration (m/s^2)	12.35	11.67	2.18	3.85	3.50	1.07
Minimum acceleration (m/s^2)	-10.19	-9.17	2.29	-3.39	-3.08	1.22
Average positive acceleration (m/s^2)	0.85	0.85	0.10	0.46	0.45	0.06
Average negative acceleration (m/s^2)	-0.84	-0.84	0.09	-0.45	-0.44	0.05
Time portion of acceleration (%)	40.75	40.97	1.73	46.90	47.03	1.41
Time portion of deceleration (%)	41.12	40.99	1.92	48.79	48.83	1.82
Time portion of cruising (%)	18.13	17.79	3.31	4.31	3.60	2.73
Time portion of idling (%)	2.80	0.54	3.63	4.05	3.37	2.62

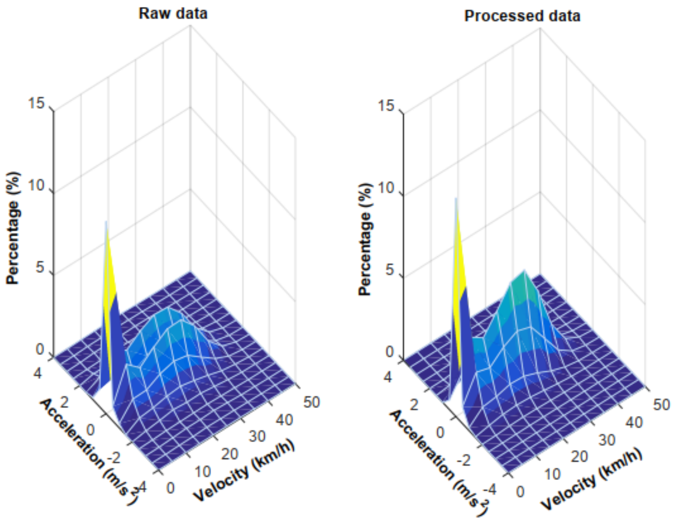


Fig. 5. Comparison of the velocity-acceleration frequency distribution.

4. Conclusions

A filter consisting of consecutive nine steps was designed to normalize dataset and to minimize errors in the GPS data for the development of the typical driving cycle, which is based on the Markov theory. Errors related to abnormal data points, zero-velocity signal drift, signal gaps in the GPS data have been detected and repaired from step 4 to step 8 by creating gaps at appropriate positions. The missing data estimation algorithm developed by Selesnick [6] was used to find new values to fill up the gaps. The reliability of this algorithm was verified by the comparison of RMSE between the true data and the estimated data obtained by three different methods. Results show that RMSE of the missing data estimation algorithm of Selesnick is smallest, 0.2 km/h. This is a new approach in the GPS data processing. The ratio of errors detected and repaired from step 4 to step 8 according to above approach is approximately 7%. Two common kinds of errors in the GPS data were detected, i.e., zero-velocity drift and signal gap, with the error ratios of 3.1% and 2.5%, respectively. The modified Kalman filter was used as the last step in the overall filtration process to minimize the effects of any remaining errors after undergoing through the prior filtration steps and to reduce the white noise in the signals. It is found that the use of Kalman filter can smooth GPS data well while the standard deviation of velocity is kept smallest. Therefore, the filter was well designed to improve the GPS data quality while conserving the integrity of the original data.

Acknowledgement

This research was supported by University of Transport and Communications under grant number T2019-CN-013 TĐ.

References

1. Franco, V.; Kousoulidou, M.; Muntean, M.; Ntziachristos, L.; Hausberger, S.; and Dilara, P. (2013). Road vehicle emission factors development: A review. *Atmospheric Environment,* 70, 84-97.
2. Niemeier, D.A.; Limanond, T.; and Morey, E.J. (1999). *Data collection for driving cycle development: evaluation of data collection protocols*. Institute of Transportation Studies, University of California at Davis.
3. Tong, H.Y.; and Hung, W.T. (2010). A framework for developing driving cycles with on-road driving data. *Transport Reviews: A Transnational Transdisciplinary Journal*, 30(5), 589-615.
4. Duran, A.; and Earleywine, M. (2012). *GPS data filtration method for drive cycle analysis applications*. SAE 2012 World Congress.
5. Nguyen, Y.-L. T.; Nghiem, T.-D.; Le, A.-T.; and Bui, N.-D. (2019). Development of the typical driving cycle for buses in Hanoi, Vietnam. *Journal of the Air & Waste Management Association*, 69(4), 423-437.
6. Selesnick, I. (2013). *Least squares with examples in signal processing*. Electrical and Computer Engineering Department at Polytechnic Institute of NYU. Retrieved April 21, 2017, from https://cnx.org/contents/XRPKcVgh@1/Least-Squares-with-Examples-in-Signal-Processing.
7. Shi, S.; Lin, N.; Zhang, Y.; Cheng, J.; Huang, C.; Liu, L.; and Lu, B. (2016). Research on Markov property analysis of driving cycles and its application. *Transportation Research Part D,* 47, 171-181.
8. Lien, N.T.Y.; and Dung, N.T. (2018). Markov property analysis of the real-world driving data and its application. *Journal of Science & Technology Technical Universities,* 126, 54-58.

9. Torp, E.; and Önnegren, P. (2013). *Driving cycle generation using statistical analysis and Markov chain.* Department of Electrical Engineering, Likopings University, Sweden.
10. Lavagnini, I.; Pastore, P.; and Magno, F. (1990). Comparison of the Kalman filter and dedicated algorithms for processing data for solution equilibria, noise smoothing and calibration. *Analytica Chimica Acta*, 239, 95-106.
11. Jun, J.; Guensler, R.; and Ogle, J.H. (2006). Smoothing methods designed to minimize the impact of GPS random error on travel distance, speed, and acceleration profile estimates. *Transportation Research Record*, 1972, 141-150.
12. Kim, Y.; and Bang, H. (2018). *Introduction to Kalman filter and its applications.* Retrieved January 10, 2017, from http://dx.doi.org/10.5772/intechopen.80600.
13. DieselNet (2018), *Emission test cycles*. Retrieved May 10, 2018, from https://www.dieselnet.com/standards/cycles/bac.php.