Received: 17 November 2023 | Revised: 2 September 2024 | Accepted: 27 September 2024

IET Intelligent Transport Systems

DOI: 10.1049/itr2.12574

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH

# Validity of GPS data in driving cycles

Harry Smith II | Suhail Akhtar | Brian Caulfield | Margaret O'Mahony

Civil, Structural, and Environmental Engineering, The University of Dublin Trinity College, Dublin, Ireland

**Correspondence**
Harry Smith, Civil, Structural, and Environmental Engineering, The University of Dublin Trinity College, Dublin, Ireland.
Email: smithiih@tcd.ie

**Funding information**
HORIZON EUROPE European Innovation Council, Grant/Award Number: HORIZON-CL5-2021-D5-01

**Abstract**

There is continuous research into driving cycles (DCs) as researchers across the globe seek to define driving characteristics, energy consumption, and emissions in a local context. For decades, data collection for the development of DCs has been conducted in three ways: chase car, instrumented vehicle, or a combination of both. Many studies have moved on to cheap and easily available global positioning system (GPS) technology, while others record vehicle data directly through the on-board diagnostics (OBD) port. However, there are major limitations to GPS data collection such as frequent inaccuracies and loss of coverage in urban environments. For this reason, both OBD and GPS vehicle speed data have been collected. Then, the recorded data has been analysed to capture any differences in sampling rates and dropping data. Finally, basic DCs were created from smoothed GPS and OBD data and compared. DCs were developed with a microtrip-based method, and a relative error term was used to compare candidate DCs to the collected data. DCs were compared based on kinematic characteristic parameters that are most used in the field. The results of this study could be used to assess the validity of GPS-based DCs compared to OBD cycles using low-cost devices.

## 1 | INTRODUCTION

Driving cycles or drive cycles (DCs) are used across the world as standard load profiles to compare emissions, fuel efficiency, and overall optimization of vehicles and their parts. The goal of any DC is to characterise the driving behaviour of a particular region, vehicle class, or road type through a single speed-time profile. Therefore, the representativeness of the final DC is crucial for all purposes of the DC.

However, there is no standardisation of the methods and manner for constructing DCs. The lack of standardisation has led to a wide but shallow pool of literature on the subject of DC creation, with many studies developing localised DCs based on their own judgement and only loose references to the knowledge base. There is a lack of research that compares the common state-of-the-art methods for each step of the creation process.

There are many steps in the creation of a DC, but a paramount step is data collection. Historically, either the chase car or instrumented vehicle method has been used for the vast majority of cycles. Recent upgrades in technology have shifted the state-of-the-art DCs toward the instrumented vehicle method. Still, DC developers have great freedom when deciding the manner of instrumentation. Most commonly, the data is collected through either GPS (global positioning system) or OBD (on-board diagnostics) technology. This study compares the accuracy of GPS data compared to OBD data to assess the validity of the growing number of studies that attempt to measure driving behaviour characteristics with GPS technology.

## 2 | LITERATURE REVIEW

DCs are continuously developed globally for purposes such as fuel emissions testing and vehicle optimisation [1]. Certain DCs act as legal standards to ensure vehicles meet emissions standards [2]. Lin et al. (2002) concluded that regional differences in driving behaviour are great enough to have distinct influences on emissions, leading to hundreds of localised DCs that attempt to capture the unique patterns of a region. Both legislative and non-legislative cycles have a major impact on the development of new vehicles and reaching climate goals.

There are four steps to create a DC. The first step, route selection, involves selecting one or more routes that are typical

**TABLE 1** Characteristic parameters.

| Category | Parameter |
|---|---|
| Speed-based | Average speed |
| | Average moving speed (>0) |
| | Std speed |
| Acceleration based | Max acceleration and deceleration |
| | Average acceleration and deceleration |
| | Std acceleration and deceleration |
| Drive mode percentage | % time in idle |
| | % time in cruise |
| | % time acceleration |
| | % time deceleration |

of the region for data collection. Quantitative measures such as traffic volumes may be used when determining a preferred route [3, 4]. With long-term data collection, route selection is not necessary as typical driving behaviour can be recorded through weeks or months of regular operation [5]. Because this study does not attempt to represent a region, there were no predetermined trip routes.

The next step is data collection. As previously mentioned, chase car and instrumented vehicle are the two major methods of data collection [1, 6]. In the chase car method, a study vehicle follows a car unrelated to the study on a determined route, matches the car's speed as well as possible, and records the speed of the study vehicle. This method has major flaws as shown in Morey et al. [7], so this method is not widely used today. The more common method is the instrumented vehicle where a test vehicle is studied using some data recording device installed in the vehicle. Most commonly either OBD or GPS speed data is collected from the vehicle [2].

After data collection is the construction of one or more candidate DCs. The most common methods are the microtrip (MT) method [1] and the Markov chain method (Wang et al., 2019). Amirjamshidi and Roorda [8] outlines the steps of the MT method, and this study opts for a random MT method due to ease. A MT is any section of a trip between two idle periods. In other words, a single MT is comprised of the entire speed-time profile of a vehicle between two periods of zero speed. The final DC is composed of MTs from several trips spliced together to represent the average statistics recorded in data collection. The Markov chain method stochastically builds a new speed-time profile based on probabilities from the entire dataset [9].

Finally, a candidate DC must be validated by comparing its statistics to the average statistics recorded in the dataset. A necessary step in the validation of a DC is selecting characteristic parameters (CPs) that are used to describe driving behaviour and validate the DC [10]. Most studies opt for CPs based only on the speed-time profile. There are three major characters of these kinematic parameters: speed-based, acceleration-based, and drive mode percentages. The thirteen CPs used in this study, shown in Table 1, are based on the parameters used in Komorska et al. [11].


**FIGURE 1** OBDLink MX+ Dongle.

The CPs chosen for this study are among the most common in the literature. As with every step of creating a DC, there is a large amount of freedom when selecting CPs that represent the data. For example, a common parameter in the literature is an addition drive mode called creep time [3, 12, 13]. This parameter measures the amount of time when the vehicle is moving at low speeds. This option was not chosen because low speeds in the recorded data were generally parts of larger accelerating or decelerating movements. A DC is generally validated using a relative error term that compares the CPs of the candidate cycle to the DC average [12]. A candidate cycle may be chosen once the relative error term is below a certain threshold [10]. Alternatively, a predetermined number of candidate cycles may be iteratively created, and the final cycle is chosen as the candidate cycle that has the least relative error [13].

After a thorough literature review, it was discovered that there is often little to no justification for the methods used to create a DC. This study of Dublin seeks to discover if cheap GPS technology available in smartphones provides adequate speed data to produce a representative DC. This study uses state-of-the-art methodology to produce DCs from both GPS and OBD data to compare the two common data collection methods.

## 3 | METHODOLOGY

Data was collected from a Hyundai Kona in Dublin, Ireland. The data was recorded through the OBD port of the vehicle. An OBDLink MX+ dongle [14], seen in Figure 1, was attached to the vehicle and connected to the Car Scanner app [15] on a smartphone to receive and export the collected data. The app receives OBD data from the dongle while simultaneously recording GPS location from the smartphone and calculating instant speed based on the updated coordinates. This method does not allow for a regular sampling rate, but the average rate of 2 Hz is adequate to create smooth MTs after data processing [4].

A total of ten trips were recorded with this setup. The final database was composed of 78 MTs (13,678 s) across the nine trips after processing. One trip was removed from the database

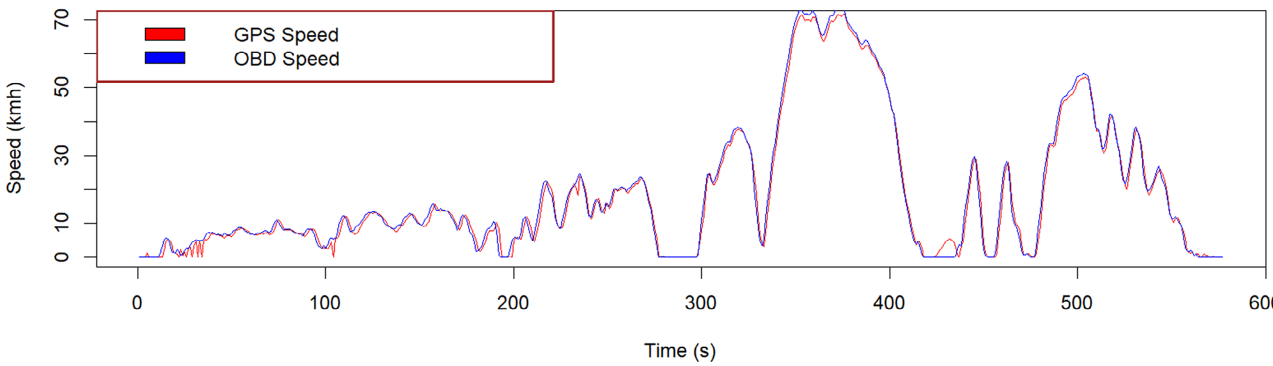


**FIGURE 2** GPS and OBD speed data from sample trip.

for having unusually high amounts of missing OBD and GPS speeds. The trips were made across both urban and suburban areas. As seen in Figure 2, the GPS and OBD speed readings were mostly similar during a typical trip. The GPS data is slightly less consistent in the form of missing values and spikes of noise, especially at lower speeds.

Data pre-processing was done first in *R*, where the irregular sampling size of the data was transformed into a smoothed data frame with a single value per second. Speed readings were aggregated to each second, and any missing value was filled in with interpolated speeds from the nearest readings done via cubic spline interpolation. The OBD data was then fit for processing.

To create a smoother speed profile, a second set of GPS data was formed from the original GPS data by creating a rolling average over the three most recent datapoints. This mimics what a study might do while building a DC using only GPS data. A period of three was chosen so that steep acceleration and decelerations could be accounted for while still filtering noise. Then, the GPS speeds were scrubbed manually to delete "phantom" MTs caused by random spikes of GPS speeds during idle periods.

All three sets of data (OBD, GPS, and GPS smoothed with rolling average) were analysed to ensure that missing datapoints filled by cubic spline interpolation would not affect the data quality. Across the over 13,000 s of moving speeds, only 3.03% of OBD and 2.51% of GPS data were missing. On the basis of such low levels of missing data and the spline interpolation matching the curved nature of trip data well, analysis of the dataset proceeded.

The construction method used in this study is the random MT method as used to create the Korean DMZ cycle among other global DCs [16]. This method relies on a simple algorithm that pieces together random MTs from the dataset until a desired length is reached. The resultant candidate cycle is then evaluated. Because of the random nature of generating a DC, the best way to compare GPS and OBD data is to iterate an equal number of candidate cycles and select the best one as the final DC. Then, comparisons can be made across the entire population of DCs as well as the candidate cycle closest to the real-world CPs.

The desired length of the DC was set to 1500 s, slightly less than the average trip duration of 1723 s to allow for more variability between candidate cycles. This means that once a randomly selected idle period brought the duration of the candidate cycle to 1500 s or more, the candidate cycle would be complete, and its statistics would be calculated. One thousand cycles were iteratively generated for each of the three datasets.

Cycles were created with the same method for each of the three datasets: OBD, GPS, and GPS smoothed. Acceleration was calculated by the change in speed between each second, which was then converted to meters per second squared. Then, drive mode was found by comparing the speed and acceleration at each second. Cruising was defined as maintaining a speed within 0.5 km/h of the previous second at speeds over 0. From there, the remaining CPs were calculated.

A relative error term was introduced as a method of evaluating candidate cycles. Relative error was calculated by the following equation, which produces the sum of relative errors of all CPs as a single term:

$$\mathrm{RE} = \sum_{i=1}^{n} \left| \frac{|x_i - y_i|}{y_i} \right|$$

Where $x_i$ is the value of a CP in the candidate cycle, $y_i$ is the average value of the same CP in the dataset. The relative error is found by totalling the errors of all 13 CPs. This summed relative error term is referred to as simply relative error for the remainder or this paper. There was no weighting of parameters or relative error threshold set to accept a candidate cycle as the final cycle. Instead, the DC with the lowest relative error from the 1000 candidates was selected.

## 4 | RESULTS

The crux of this research occurs during data collection. Therefore, it is necessary to examine the CPs collected during the four trips that attempt to characterise the recorded driving behaviour. The twelve CPs of the various datasets are presented in Table 2.

The GPS and OBD technologies recorded similar characteristics overall. There are no major differences across the speed characteristics in any dataset. Stark differences show in the acceleration-based parameters. The inaccuracies inherent to

**TABLE 2** Characteristic parameters of the dataset.

| | Dataset | | |
|---|---|---|---|
| **Parameter** | **OBD** | **GPS** | **GPS smoothed** |
| Mean speed (km/h) | 42.01 | 41.08 | 41.08 |
| Mean speed >0 (km/h) | 50.06 | 50.82 | 50.19 |
| Std speed (km/h) | 34.56 | 33.79 | 33.75 |
| Mean accel (m/s$^2$) | 0.37 | 0.50 | 0.38 |
| Max accel (m/s$^2$) | 4.23 | 5.86 | 3.62 |
| Std accel (m/s$^2$) | 0.45 | 0.52 | 0.42 |
| Mean decel (m/s$^2$) | −0.38 | −0.50 | −0.39 |
| Max decel (m/s$^2$) | −3.67 | −5.80 | −3.22 |
| Std decel (m/s$^2$) | 0.47 | 0.51 | 0.42 |
| Idle time (%) | 16.08 | 19.17 | 18.16 |
| Cruise time (%) | 32.34 | 26.02 | 33.15 |
| Accel time (%) | 26.02 | 27.41 | 24.34 |
| Decel time (%) | 25.56 | 27.40 | 24.35 |

GPS create noise instead of the smooth cruising behaviour captured by the OBD. The smoothed dataset does a good job of bringing average acceleration and deceleration back to the OBD-recorded levels but falls short in capturing maximum accelerations due to the averaging technique.

More differences appear in the quantitative statistics that describe drive mode percentages. Unsurprisingly, the raw GPS data underestimates the cruise time due to the volatile measurements at high and low speeds. The smoothed GPS data once again succeeds in bringing this parameter closer to reality. Somewhat more interestingly, both GPS sets overestimate the duration of idle time during trips. This is most likely due to delay in speed recordings when the vehicle starts from idle.

While the datasets show varying levels of differences across the CPs, the ultimate measure is the performance of each dataset when building a DC. Figures 3–5 show the final DC selected as the best candidate cycle by having the least relative error for each dataset.

Importantly, the final DCs in Figures 3–5 are visually similar, showing that the construction process was effective. Even still, the raw GPS data in Figure 4 noticeably lacks smooth cruising and acceleration periods, instead appearing more piecewise. This can be best seen in the final MT of Figure 4, which is also the final MT of Figure 3. This juxtaposition emphasises how the same trip data is recorded differently by the OBD and GPS technologies. The smoothed GPS cycle in Figure 5 lacks some of the sharp acceleration spikes that feature prominently in the cycles of Figures 3 and 4. Further differences between these final DCs, as well as the average of all 1000 candidate cycles, can be seen quantitatively in Table 3 below.

An important finding is that the average relative error term of all candidate cycles is slightly higher for the raw GPS data. This suggests that there is more variation across MTs in the original GPS dataset, so the actual driving behaviour may not be captured in many candidate cycles. A lower average relative error could be of importance with larger datasets and computational time if the final cycle is selected by an error threshold rather than an iteration limit as seen here. The threshold would most likely be reached faster using the OBD or smoothed GPS dataset.

Interestingly, the final DC of the GPS and GPS smoothed data have lower relative errors than the OBD cycle. However, this coincidence tells very little about the validity of GPS data. All three cycles are built to replicate the trip data recorded using that respective method. A more important measurement is to

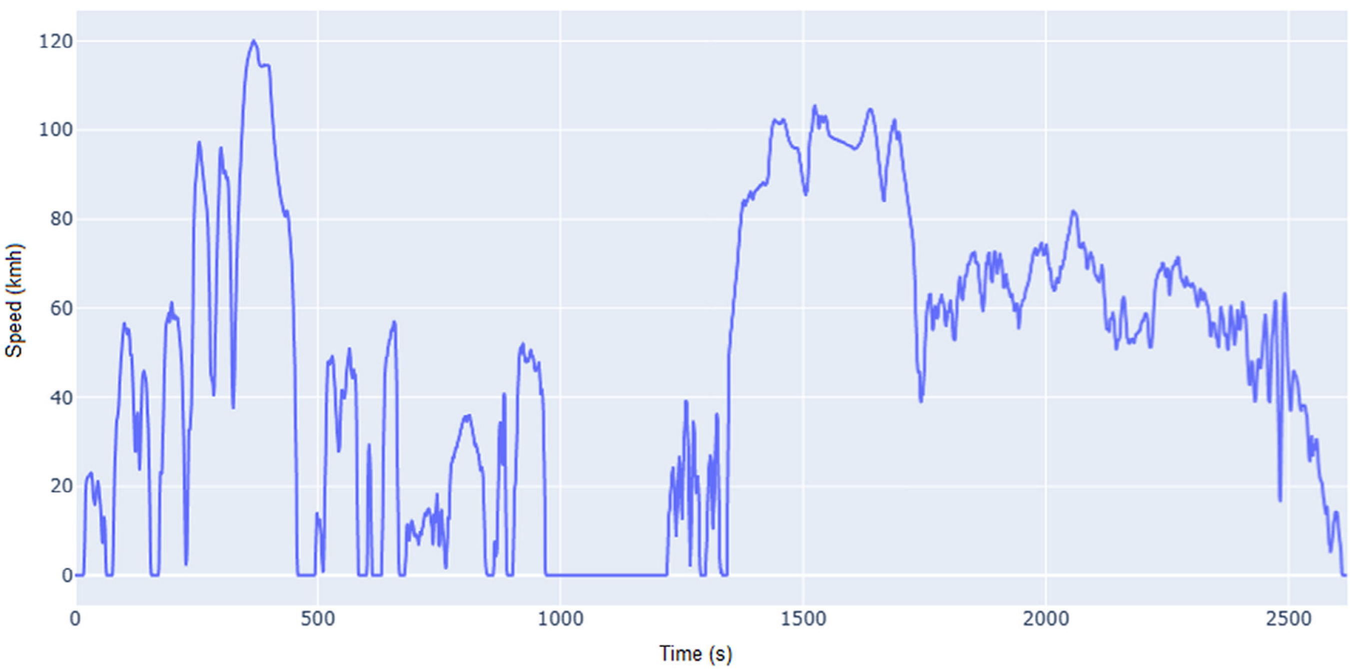


**FIGURE 3** OBD driving cycle.

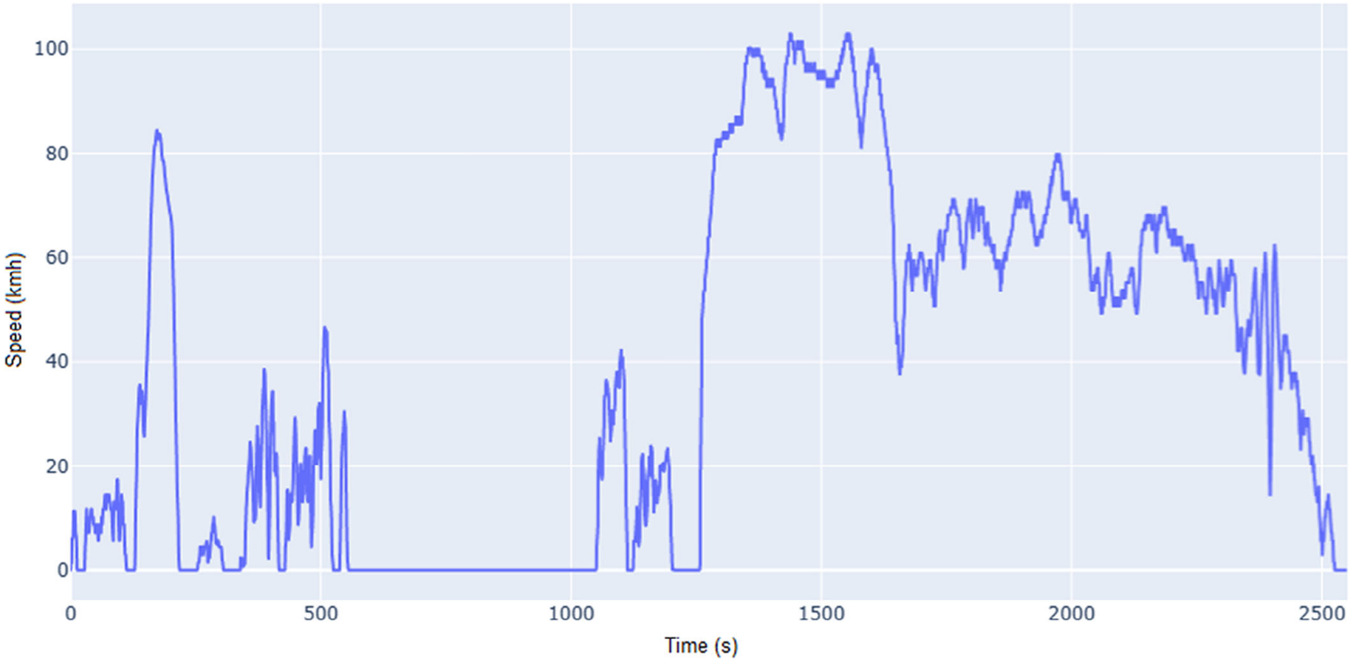


**FIGURE 4** GPS driving cycle.

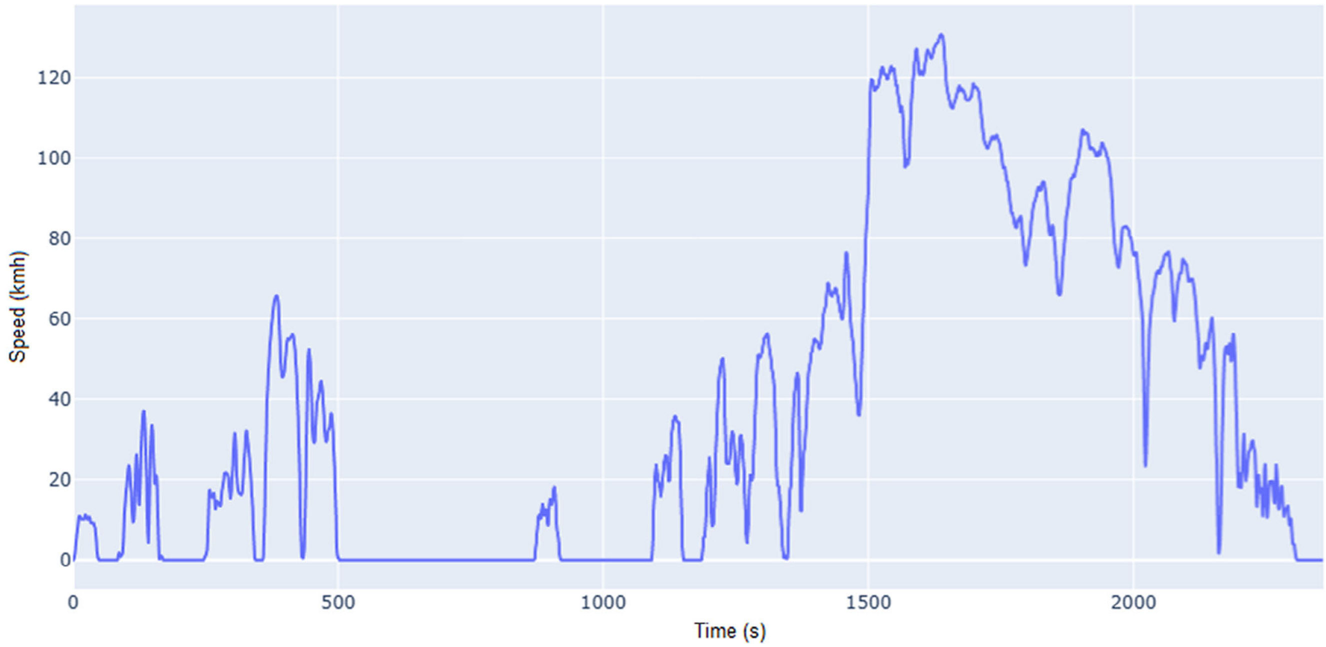


**FIGURE 5** GPS smoothed driving cycle.

see how well each DC represents the accurate OBD data. This measurement helps determine if GPS data is a suitable, cheap alternative for OBD data. Table 4 compares the relative error of the final DCs compared to the OBD data. Of course, the relative error of the OBD cycle is the same as reported in Table 3 because the cycle was built from OBD data.

Table 4 provides interesting results to the issue of this paper. Using the chosen validation method of comparing CPs in the DC to those in the dataset, the raw GPS DC represents the driving behaviour captured by the OBD better than the OBD cycle. Even though the GPS DC was built using only GPS data, it represents the average CPs of the OBD data better than the OBD cycle. The GPS smoothed final DC contains almost the same amount of relative error as the OBD cycle, too. This can lead to one of two conclusions: either the GPS data is a good substitute for OBD data and captures the same data almost exactly

**TABLE 3** Characteristic parameters of final cycles and average of constructed cycles.

| | Final cycle | | | Average of 1000 cycles | | |
|---|---|---|---|---|---|---|
| **Parameter** | **OBD** | **GPS** | **GPS smoothed** | **OBD** | **GPS** | **GPS smoothed** |
| Rel error | 2.06 | 1.16 | 1.94 | 3.45 | 3.61 | 3.44 |
| Mean speed (km/h) | 47.56 | 37.77 | 37.31 | 46.02 | 36.95 | 40.08 |
| Mean speed >0 (km/h) | 49.92 | 43.91 | 42.81 | 47.05 | 39.18 | 41.45 |
| Std speed (km/h) | 33.95 | 34.14 | 40.59 | 30.65 | 29.57 | 29.94 |
| Mean accel (m/s$^2$) | 0.34 | 0.47 | 0.34 | 0.32 | 0.42 | 0.32 |
| Max accel (m/s$^2$) | 3.32 | 4.40 | 2.21 | 2.64 | 3.95 | 2.01 |
| Std accel (m/s$^2$) | 0.40 | 0.52 | 0.37 | 0.37 | 0.45 | 0.35 |
| Mean decel (m/s$^2$) | −0.36 | −0.44 | −0.36 | −0.34 | −0.40 | −0.34 |
| Max decel (m/s$^2$) | −2.95 | −5.80 | −2.61 | −2.22 | −3.32 | −2.21 |
| Std decel (m/s$^2$) | 0.39 | 0.50 | 0.37 | 0.39 | 0.43 | 0.34 |
| Idle time (%) | 4.73 | 13.99 | 12.84 | 2.48 | 7.04 | 4.16 |
| Cruise time (%) | 40.77 | 30.21 | 35.27 | 48.00 | 41.59 | 47.21 |
| Accel time (%) | 28.49 | 28.29 | 26.75 | 26.53 | 26.59 | 25.58 |
| Decel time (%) | 26.01 | 27.51 | 25.14 | 22.98 | 24.77 | 23.04 |

**TABLE 4** Relative errors of final DCs compared to OBD dataset.

| **Final cycle** | **Relative error compared to OBD data** |
|---|---|
| OBD | 2.06 |
| GPS | 1.86 |
| GPS smoothed | 2.08 |

or there is an issue with the common method of validating DCs that simplifies the diversity of driving behaviour into a few CPs. By looking at the rest of the data and how much the raw GPS data varies from the other two sets in nearly all CPs, it can be asserted that the latter conclusion is far more likely, and the validation methods used to create DCs need to be reviewed and revised.

## 5 | CONCLUSIONS

Although this study worked with limited data and simple construction methods, the results provide a guideline for the development of future DCs in the research programme.

First, the results confirm that modern day technology allows for cheap and accessible data to be used in the creation of DCs. The GPS equipped on all modern smartphones, which themselves are universal, can provide a more than reasonable estimate of a vehicle's speed throughout a trip. From there, simple data processing techniques such as rolling averages can reduce the noise and dropped data points to create a dataset that very closely resembles the vehicle's real speed. Likewise, the OBD dongle used for data collection was a relatively cheap and commercially available option. Cheaper data collection devices and faster computers likely means future DCs can be built from larger datasets than in previous decades.

Second, the results show that the state-of-the-art methodology used to create DCs today most likely have some inherent flaws that likely lead to DCs not accurately capturing driving behaviour. A few kinematic CPs and relative error measurements can lead to a misrepresentation of driving behaviour in the final DC. This study points to the usefulness of validating DCs based on an accuracy threshold applied to all CPs rather than selecting the lowest total relative error from candidate cycles. Additionally, it will be necessary to look beyond kinematic CPs to CPs that capture the intent of the final DC better, such as energy use or speed-acceleration frequency distributions.

The scope of this study was set within the known limitations, and the conclusions were drawn using a robust methodology and analysis of a real-world driving data. The difficulties in collecting real-world data cannot be underestimated, both in the context of the technical challenges but also owner concerns about privacy and tracking. It would be useful if other researchers could confirm the findings presented here using real-world data collected in their cities.

The results of this paper provide confidence for past and future DCs built from GPS data. The GPS data is able to capture speeds trends well without many issues of missing or noisy data, especially once smoothed. It calls into question other steps of DC construction such as choosing CPs which future research could examine. The unreliability of common CPs to describe driving behaviour can have major effects on emissions tests and other uses of DCs.

## AUTHOR CONTRIBUTIONS

**Harry Smith II**: Conceptualization; data curation; formal analysis; investigation; methodology; visualization;

writing—original draft; writing—review and editing. **Suhail Akhtar**: Writing—review and editing. **Brian Caulfield**: Supervision; writing—review and editing. **Margaret O'Mahony**: Supervision; writing—review and editing.

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## ORCID

*Harry Smith II* https://orcid.org/0000-0002-2969-2464
*Suhail Akhtar* https://orcid.org/0000-0002-8790-199X
*Brian Caulfield* https://orcid.org/0000-0003-3877-475X
*Margaret O'Mahony* https://orcid.org/0000-0002-8647-0091

## REFERENCES

1. Xiao, Z., Dui-jia, Z., Jun-min, S.: Synthesis of methodologies and practices for developing driving cycles. Energy Proc. 16, 1868–1873 (2012). https://doi.org/10.1016/j.egypro.2012.01.286
2. Wang, L., Ma, J., Zhao, X., Li, X.: Development of a typical urban driving cycle for battery electric vehicles based on kernel principal component analysis and random forest. IEEE Access 9, 15053–15065 (2021). https://doi.org/10.1109/ACCESS.2021.3052820
3. Mayakuntla, S.K., Verma, A.: A novel methodology for construction of driving cycles for Indian cities. Transp. Res. Part D: Transp. Environ. 65, 725–735 (2018). https://doi.org/10.1016/j.trd.2018.10.013
4. Zhao, X., Zhao, X., Yu, Q., Ye, Y., Yu, M.: Development of a representative urban driving cycle construction methodology for electric vehicles: a case study in Xi'an. Transp. Res. Part D: Transp. Environ. 81, 102279 (2020). https://doi.org/10.1016/j.trd.2020.102279
5. Ashtari, A., Bibeau, E., Shahidinejad, S.: Using large driving record samples and a stochastic approach for real-world driving cycle construction: Winnipeg driving cycle. Transp. Sci. 48(2), 170–183 (2014)
6. Galgamuwa, U., Perera, L., Bandara, S.: Developing a general methodology for driving cycle construction: comparison of various established driving cycles in the world to propose a general approach. J. Transp. Technol. 05(04), 191–203 (2015). https://doi.org/10.4236/jtts.2015.54018
7. Morey, J.E., Limanond, T., Niemeier, D.A.: Validity of chase car data used in developing emissions cycles. J. Transp. Stat. 3(2), 15–28 (2009)
8. Amirjamshidi, G., Roorda, M.J.: Development of simulated driving cycles for light, medium, and heavy duty trucks: case of the Toronto Waterfront Area. Transp. Res. Part D: Transp. Environ. 34, 255–266 (2015). https://doi.org/10.1016/j.trd.2014.11.010
9. Lin, J.: A Markov process approach to driving cycle development. Ph.D. Dissertation, University of California, Davis (2002)
10. Berzi, L., Delogu, M., Pierini, M.: Development of driving cycles for electric vehicles in the context of the city of Florence. Transp. Res. Part D: Transp. Environ. 47, 299–322 (2016). https://doi.org/10.1016/j.trd.2016.05.010
11. Komorska, I., et al.: Adaptive driving cycles of EVs for reducing energy consumption. Energies 14(9), 2592 (2021). https://doi.org/10.3390/en14092592
12. Arun, N.H., Mahesh, S., Ramadurai, G., Shiva Nagendra, S.M.: Development of driving cycles for passenger cars and motorcycles in Chennai, India. Sustainable Cities Soc. 32, 508–512 (2017). https://doi.org/10.1016/j.scs.2017.05.001
13. Nouri, P., Morency, C.: Evaluating microtrip definitions for developing driving cycles. Transp. Res. Rec. 2627, 86–92 (2017). https://doi.org/10.3141/2627-10
14. OBD Solutions, LLC: OBDLINK MX+. https://www.obdlink.com/products/obdlink-mxp/ (2023).
15. Svistunov, S.: Car scanner ELM OBD2. https://www.carscanner.info/ (2023).
16. Han, D.S., Choi, N.W., Cho, S.L., Yang, J.S., Kim, K.S., Yoo, W.S., Jeon, C.H.: Characterization of driving patterns and development of a driving cycle in a military area. Transp. Res. Part D: Transp. Environ. 17(7), 519–524 (2012). https://doi.org/10.1016/j.trd.2012.06.004

**How to cite this article:** Smith, H., Akhtar, S., Caulfield, B., O'Mahony, M.: Validity of GPS data in driving cycles. IET Intell. Transp. Syst. 1–7 (2024). https://doi.org/10.1049/itr2.12574