

A context-aware approach to construct driving cycles in a volatile urban environment

Asad Yarahmadi^{1,3,4} · Catherine Morency^{1,3,4} · Martin Trepanier^{2,3,4}

¹ Department of Civil, Geological and Mining Engineering, Polytechnique Montréal, Montreal, QC H3T 1J8, Canada

² Department of Mathematical and Industrial Engineering, Polytechnique Montréal, Montreal, QC, Canada

³ Interuniversity Research Centre On Enterprise Networks, Logistics, and Transportation (CIRRELT), Montreal, Canada

⁴ Chair in Transportation Transformation Montréal, Montreal, Canada

Corresponding author: asad.yarahmadi@polymtl.ca

Published in Transportation · Accepted: 8 May 2025 · Published online: 29 May 2025 · DOI:

<https://doi.org/10.1007/s11116-025-10631-8> · © The Author(s), under exclusive licence to Springer

Science+Business Media, LLC, part of Springer Nature 2025

Abstract

Driving cycles are essential for estimating vehicle emissions, as regional driving dynamics are reflected through them. While exogenous factors such as weather conditions, road features, and type of day impact these patterns, this diversity is often neglected by many studies, and limited driving cycles are usually selected. In this study, a multifaceted framework was generated, resulting in a mixed model through which the influence of road and weather factors on driving cycle parameters was examined. A microtrip-based technique and the Markov process were employed on a comprehensive driving dataset consisting of over 10,567,000 trips to construct driving cycles tailored to different conditions, differentiating between weekdays and weekends. Subsequently, to determine the optimal number of driving cycles in volatile conditions, a two-stage method was developed, relying first on a similarity indicator to measure the similarity between cycles, and then on information loss and variability metrics to assess the variability of the driving cycles. It was found that Montreal's driving conditions are best captured by 76 distinct cycles.

Keywords Vehicle emissions · Driving cycle · Information loss · Variability · Weather condition · Road types

Introduction

The challenge of curbing greenhouse gas emissions from global transportation systems has been amplified due to the persistent escalation in passenger activity, which could potentially eclipse the beneficial impacts of all deployed mitigation strategies (Salihi, Demir, and Demir 2023a) (IPCC 2022). According to the International Council on Clean Transportation (ICCT), a significant obstacle in mitigating vehicle emissions is improving the accuracy of effective emissions measures to ensure that all stakeholders in the transportation sector adhere to the stipulated legislative guidelines for emissions control (Yang, Muncrief, and Bandivadekar 2017). The driving cycle serves as a pivotal measure extensively used in the modeling of vehicle emissions.

The prevalent driving cycles are SDC (Standard Driving Cycle) (Knez et al. 2014) and LDC (Local Driving Cycle) (Andrade et al. 2020; Yang Yang et al. 2020). The SDCs represent driving behaviors at the national (e.g., FTP-75 and J10-15) or international level (e.g., WLTC) to certify a vehicle's emissions, while LDCs are used to represent driving patterns at a local scale (city, road type ...). The majority of pre-existing driving cycles, formulated at the urban or national levels, may fail to address the fluctuations in emissions, potentially leading to significant inaccuracies in emission estimations (Mayakuntla and Verma 2018; Liu et al. 2020; Zhang et al. 2021; Kazemi Miyangaskari et al. 2023).

A key challenge in this study is determining the optimal number of driving cycles, as it is crucial to capture the variability of driving behaviors across all types of roads and driving conditions. Recent research has shown

that factors such as precipitation, temperature, road gradient, and speed limits significantly contribute to the variability of driving cycle parameters (Miyangaskary, Keivanpour, and Lamghari 2024; Yarahmadi, Morency, and Trepanier 2023a; Zhang et al. 2021). As a result, driving behavior varies across an entire city road network, changing as weather conditions change.

As shown in Fig. 1, the number of unique driving patterns can range from a single cycle to a count equal to the total number of drivers in an urban area. Selecting a singular driving cycle necessitates a highly aggregated level of driving data, a process that results in substantial information loss and a marked reduction in behavior variability representation. Consequently, the derived driving cycle often fails to accurately represent driving behaviours, engendering a notable discrepancy between actual and projected vehicle emissions. It is worth mentioning that information loss refers to the reduction in the diversity, granularity, and representation of specific driving characteristics when transitioning from the collected driving data to the representative driving cycle(s) (Liu et al. 2020).

Conversely, generating individual driving cycles for each driver increases computational complexity and magnifies both the consumption of time and resources. This approach also increases the risk of encountering outliers and data redundancy. Against this backdrop, the central objective of this study is to devise a novel machine learning methodology with the aim of precisely determining the optimal number of driving cycles within an urban context.



Fig. 1 Driving cycle number spectrum

In this study, optimal refers to a compromise between a set of values. The optimal number of driving cycles is defined as the minimum number of driving cycles that can still capture the maximum variability with the least information loss, effectively representing driving behaviors across different driving conditions, various road types, and in different types of days. The number of driving cycles in an urban setting like Montreal is unlimited and undefined. In fact, trip characteristics are functions of exogenous factors like road, weather, and type of day; therefore, there is an indefinite number of driving cycles in an urban area. As a result, the word 'finite' is used to indicate a practical number of driving cycles that can realistically be generated, considering the limitations of data availability, computational resources, and the importance of minimizing the impact of anomalies. A crucial component of the proposed strategy is its intended adaptability—the methodology is designed to be sufficiently versatile to be applicable across diverse urban areas, irrespective of the size and complexity of a city's road network structure, or its prevailing weather conditions.

A review of existing literature underscores the trend of previous studies predominantly selecting one, or a limited number of driving cycles to represent driving patterns in urban contexts. This study aims to significantly expand upon this base knowledge, proposing an optimized, adaptive solution that is capable of accurately representing the driving behaviors variability in diverse urban contexts.

The present investigation offers several significant contributions to the understanding and analysis of driving cycles within urban environments:

- A methodology is delineated for determining the optimal number of driving cycles for an urban area, contributing to a more accurate representation of driving behaviors.
- An approach is introduced for defining the appropriate level of data amalgamation for the formation of driving cycles, thereby mitigating the challenge of information loss and ensuring the representation of behavior variability.

- Furthermore, a robust model is provided designed to identify and quantify the exogenous factors of most significant influence on driving cycle parameters. This addition to the field enables a more nuanced understanding of the variables that shape driving behaviors and contributes to the precise generation of representative driving cycles.

This study addresses the following critical inquiries:

1. **How many driving cycles are needed to accurately reflect typical driving patterns in Montreal?** This question seeks to determine the optimal number of driving cycles required to capture the typical variations in driving behavior across Montreal, ensuring that the driving cycles developed are representative of real-world conditions.
2. **What are the key external factors that most significantly influence the variability in driving conditions and vehicle emissions in urban environments?** This question aims to identify which external elements—such as weather conditions, road types, or daily patterns—play the most critical role in affecting the variability observed in urban driving behaviors and the corresponding emissions.

In response to the complexities and variability of urban driving conditions, it is hypothesized in this study that a context-aware machine learning methodology can more accurately generate driving cycles that reflect the diverse driving behaviors influenced by different road types, weather conditions, and daily variations. Specifically, it is proposed that this approach will significantly reduce information loss and enhance the behavioral representativeness of driving cycles compared to traditional driving cycle generation methods. The hypothesis is tested by comparing the newly developed driving cycles with existing standard driving cycles in terms of their ability to capture the actual driving patterns observed in real-world urban environments.

The remainder of this paper is structured as follows. First, previous studies that explored the impacts of exogenous factors on driving cycle parameters and different approaches to determine the optimal number of driving cycles are reviewed. Then, the methodology section provides a background about techniques used to conduct this study and describes the steps used to find the optimal number of driving cycles. Finally, the results are discussed, and some perspectives are proposed.

Literature review

The development of a driving cycle can be predominantly delineated into three phases: accumulation of driving data, creation of a driving cycle, and recognition of a representative driving cycle. To enhance the precision of the designed driving cycle, precise data collection methodologies have been leveraged in prior research, ranging from smartphone sensor data (Rechkemmer et al. 2021), to Onboard Recording Devices and OBD-II systems (Ma et al. 2019; Salihu, Demir, and Demir 2023a), connected Vehicle Data (Khattak and Wali 2017), and observation of Naturalistic Driving behaviors (Esser and Rinderknecht 2020). Diverse techniques have also been introduced by previous research, such as Microtrip-based (Yang et al. 2020), Markov Chain-based (Roy and Morency 2020; Xu et al. 2025; Ying Yang et al. 2018), and Fuel-based (Huertas et al. 2018; Quirama et al. 2020) methods to generate driving cycles (Czégé, Vámosi, and Kocsis 2020; José Ignacio Huertas et al. 2019; Miyangaskary, Keivanpour, and Safari 2025).

Although considerable strides have been made in creating driving cycles, the link between fluctuations in driving cycle parameters and external elements like road attributes and weather conditions has scarcely been explored by researchers. As a result, a variety of road segments have been selected for data collection during the route selection phase without considering their impact on driving behavior. The data has typically been gathered without accounting for driving conditions, and in the end, all data has been amalgamated and a single or a few driving cycles have been selected to typify driving behaviors within a city area.

Route selection and road features

The selection of a route is a crucial step that can significantly influence the precision of a driving cycle. During this phase, road links are designated for the data collection process. Therefore, if the chosen segments fail to accurately mirror the characteristics of an urban road network, the driving cycle might not appropriately depict driving behaviors.

A review of related literature indicates that authors commonly have focused on traffic characteristics (Tong and Ng 2021; Yaghoubi et al. 2024; Yang Yang et al. 2020), experiential knowledge (Knez et al. 2014), road functions (Arun et al. 2017), or a limited set of factors such as traffic flow (Keskin, Yıldız, and Arslannur 2024), population density, squares, junctions, and public transport systems (Lee et al. 2025; Pouresmaeili et al. 2018) when identifying representative road segments. However, certain studies, due to their specific case-study contexts, have not considered explicit selection criteria because their routes were predefined. For example, Hao et al. (2024) developed driving cycles for garbage trucks based solely on the routes regularly assigned to these vehicles. Similarly, Tong and Ng (2023) selected a predetermined 4.5 km bus route to collect driving data for cycle generation.

Nevertheless, the consideration of road features, like gradient, curvature, intersection density, road width, and built environment that can influence driving behaviors, is seldom observed (Azam, Bhaskar, and Haque 2022; Cui and Levinson 2018; Faramarzpour et al. 2025).

In their study, Salihu, Demir and Demir, (2023a) analyzed the effect of road gradient on driving cycle parameters, amassing around 10 h of driving data. Their findings revealed that acceleration and deceleration were dependent on the incline of the road. It was concluded that driving cycles vary across road segments with differing slopes, and it was recommended that to enhance the accuracy of vehicle emissions estimates, road segments with gradients ranging from -2 to 2 be included in driving cycle computations. However, the relationship between road gradient and shifts in driving modes was not elaborated.

In another study, the influence of road design features—including road shoulders, their presence, and horizontal geometry—on driving speed and lateral position as indicators of driving behaviors was examined by Abele, Møller and Mette, (2011). Their research demonstrated that participants drove slower and closer to the centerline on curved roads compared to straight ones. The presence of road shoulders was linked with faster speeds and changes in lateral positions on all road segment types. Participants generally displayed higher speeds when road shoulders were available. Additionally, their findings contradicted previous studies (Antonson et al. 2009; Miyangaskary et al. 2025; Sagberg 2003) by suggesting that the presence of trees had no significant effect on driving speed. However, their study was limited as it was conducted in a simulator environment with limited data, and the impact of traffic conditions on driving behavior was not considered. The effect of road geometry on driving behavior, especially in the context of Connected and Autonomous Vehicles (CAVs), was examined by Coropulis et al. (2025). The study analyzed the effects of abrupt geometric transitions—such as straight road segments immediately followed by sharp curves without transitional curves—on driver behavior on two-way, two-lane rural roads. Real-time speed and acceleration data were collected using roadside radar counters. The results revealed a strong correlation between high speeds on straight segments and aggressive braking behavior at sharp curves.

Further research by Wolfermann, Alhajyaseen and Nakamura, (2011) confirmed that signalized intersections significantly influenced driving speed profiles, while Khanfar et al., (2023) demonstrated that flashing green lights encourage more conservative driving behaviors through an analysis of vehicle kinematic parameters. Additionally, Supiyono, Djakfar and Wicaksono, (2022) observed that among all traffic light phases, the yellow light significantly influenced driving behavior.

Collecting driving data and driving conditions

The process of collecting driving data typically involves gathering real-world data over a period that ranges from three days to two weeks in most studies. Alternatively, simulation techniques may be used by researchers to

generate representative driving cycles. In both scenarios, driving conditions may not be adequately considered. Existing literature demonstrates that weather parameters, such as precipitation, temperature, wind speed, and humidity, can impact driving behavior characteristics. It was demonstrated by Faria et al., (2020) that average speed during heavy rain is 22% lower than in clear weather, and acceleration and deceleration are reduced by approximately 10%. Ahmed et al., (2018) further illustrated that weather conditions like snow, fog, and rain compel drivers to significantly reduce their speed compared to clear weather conditions.

The failure to consider weather conditions in the generation of driving cycles may result in driving cycles that do not adequately represent the diversity of driving behaviors. Consequently, the Local Driving Cycles (LDC) may be less representative than the Standard Driving Cycles (SDC) (Gebisa et al. 2021). These limitations have been addressed in some studies by collecting vast driving datasets across various road segments and weather conditions. However, all collected data have typically been aggregated, and either a single or a limited number of driving cycles have been selected. For example, Yuzhou Yang et al. (2024) collected approximately sixty million driving data points from three thousand heavy-duty trucks, yet ultimately aggregated this extensive dataset into only six representative driving cycles.

To improve the accuracy of the LDCs, the aggregation level of data must be carefully considered, and driving cycles should be generated for various driving conditions and road types. Using nearly three million trips collected over two years, Yarahmadi, Morency and Trepanier, (2023b), confirmed that driving cycles under different driving conditions were distinct and proposed considering a separate driving cycle for each driving condition. Furthermore, a driving cycle for different road functions was generated by Zhang et al., (2021), who provided compelling evidence that driving cycle parameters differ among road types. They ultimately suggested that having a road type-based driving cycle could improve the estimation of vehicle emissions.

In conclusion, road characteristics and weather parameters significantly influence driving cycle parameters. However, this aspect has been overlooked in the majority of prior studies. Instead, all data collected under varying road and weather conditions have been aggregated to construct one or a limited number of driving cycles meant to characterize driving behaviors, which may not provide a sufficiently accurate representation. The development of a method to determine the optimal number of driving cycles for various driving conditions and road types could potentially address this gap and enhance the estimation of vehicle emissions.

Comparative analysis with previous studies

Comparative analysis of previous studies with the present study helps to better understand gaps in the literature and highlight differences between them. To achieve this, Table 1 is employed to compare in detail the differences between conducted studies and the present study. As displays, conducted studies can be categorized into two main groups: local driving cycles generated by a small dataset, and local driving cycles produced by a large dataset. In the first group, driving behaviors are mimicked by selecting a limited number of road segments in limited weather conditions, which minimizes the variability of the driving cycle. In the latter group, considerable quantities of driving data are collected; however, like the first group, the collected data are aggregated to construct driving cycles, thereby maximizing information loss. Moreover, in the final step, both groups typically select one or a few driving cycles to mirror driving behaviors in an urban environment, which also minimizes variability. To fill these gaps, in the current study, driving data are filtered based on types of day, road, and weather, and then a driving cycle is constructed for each type. This approach guarantees the availability of a distinct driving cycle for different types.

Table 1 also delineates the distinctions between current research and pre-existing literature. The column labeled "External Factor(s) considered in DC generation" elucidates the exogenous variables taken into account in driving cycle creation by various researchers. The "Aggregation level" column indicates the granularity at which driving data are aggregated for the development of the driving cycle. The column "Number of DCs" presents the number of driving cycles selected to represent driving behaviours in an area. The column "Duration of data

gathering" indicates how much driving data have been collected to construct driving cycles. It should be noted that the acronym "NP" stands for "Not Presented".

In light of the limitations identified in the reviewed studies—particularly the lack of context-aware methodologies and insufficient consideration of data aggregation levels—this study introduces a comprehensive and adaptive framework for generating representative driving cycles. The following section outlines the methodological steps developed to address these research gaps.

Table 1 Comparison of conducted studies and proposed approach

Authors	External factor(s) considered in DC generation	Aggregation level	Number of DCs	Duration of data gathering
Ragione and Giovanni (2016)	Road slope	Road slope	4	NP
Zhang et al. (2021)	Road function	Highway, arterial, and other roads	3	167 h
Tong and Ng (2023)	Road slope	Weekday and weekend	2	16 h
Salihi, Demir, and Demir (2023a)	Road slope	Road slope	18	11 h
Yuzhou Yang et al. (2024)	NP	6 Usage scenarios (logistic, harbor, urban construction, ...)	6	NP
Yarahmadi et al. (2024)	Weather parameters	Weather type	15	19 months
Tharvin et al. (2018)	Traffic conditions	Peak and off-peak hours	1	NP
Arun et al. (2017)	Traffic conditions	Peak and off-peak hours	2	6 months
Chauhan et al. (2020)	Road function / intersections	Arterial urban roads	9	NP
Proposed methodology	Road types and weather types	Road type / weather types / weekday and weekends	76	19 months

Methodology

The primary challenge in determining the optimal number of driving cycles for a region is quantifying the variability of driving behaviors across the entire road network and under all driving conditions. Various steps utilized in this study to develop a methodological framework for identifying the optimal number of driving cycles are illustrated in Fig. 2. As shown in Fig. 2, the first two steps—"Identifying exogenous factors impacting DC including type of road, weather, and days" and "Developing typology approaches to categorize the external factors"—highlighted in gray, were already completed in previous authors' studies. It is worth mentioning that both studies were conducted specifically within the Montreal Island urban environment, and their results were directly applied in this study to determine the optimal number of driving cycles. These road and weather typologies, developed from the findings of the previous studies, are based on Montreal's unique driving conditions, ensuring their relevance for assessing driving cycles in various driving conditions within the same urban context.

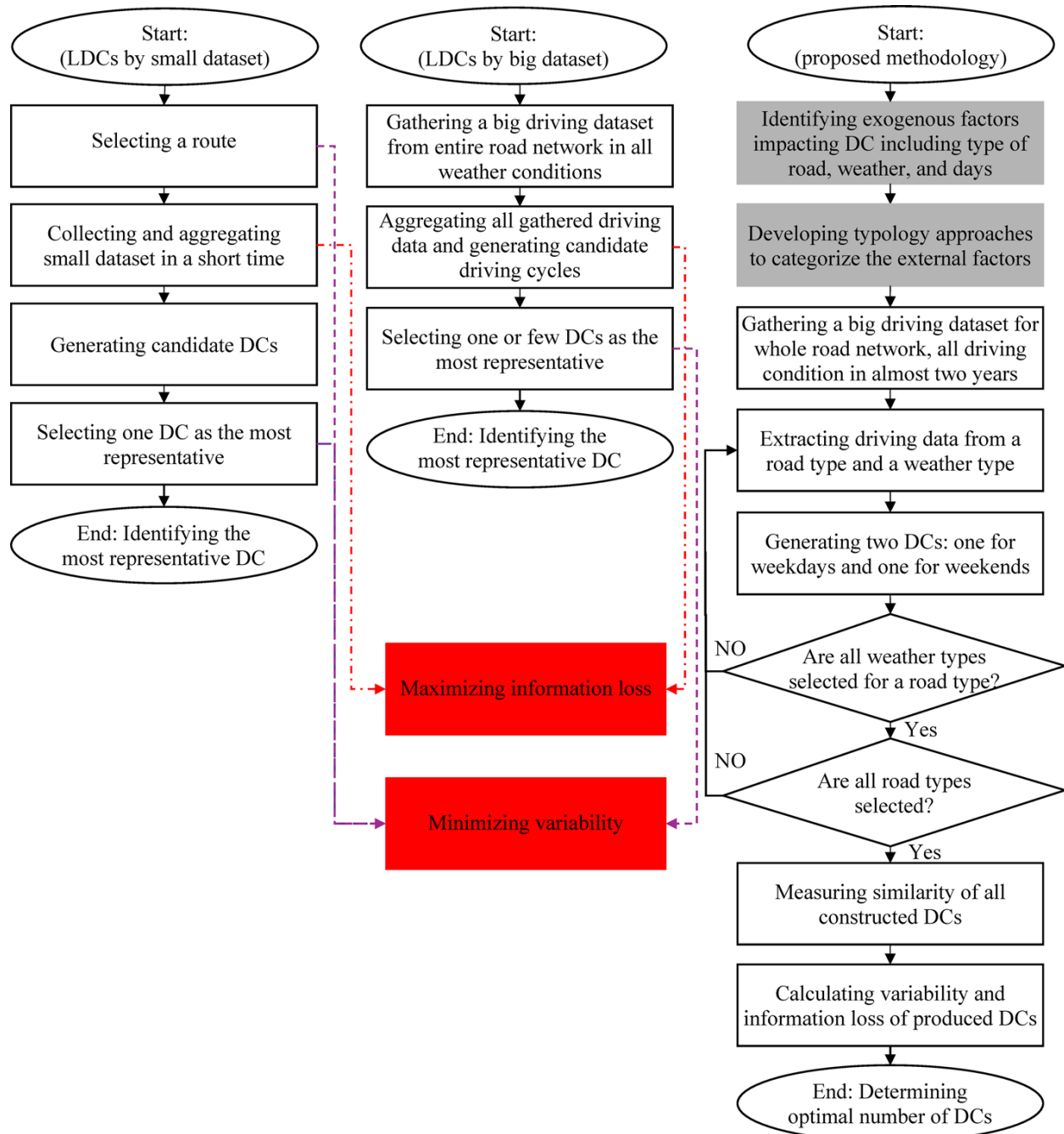


Fig. 2 Comparison of different driving cycle methodologies

In the first step of this study, a road type and a weather type are selected, and subsequently, the driving data generated under these selected conditions are extracted. Next, based on weekdays and weekends, two driving cycles are constructed for the chosen road type and weather condition. This is an iterative process that is repeated for all weather conditions across all road types. Finally, the optimal number of driving cycles is calculated using two indicators: information loss and variability of driving cycles.

New typology of road network

A typology was constructed by Yarahmadi, Morency and Trepanier, (2023b) to categorize road segments based on their impact on driving speed within the Montreal urban environment. Initially, the most significant road features that showed a strong relationship with driving speed were identified. The key features include road grades, curvature, density of signalized intersections, number of road lanes, surrounding land uses, posted speed limits, and road functions. Subsequently, machine learning was utilized to categorize the road segments, leading to the identification of five distinct road types. Their results demonstrated that the newly created typology significantly enhances the explanation of the association between changes in driving speed and road features. In addition, a statistical analysis of speed distributions within each road type revealed that driving speeds were more homogenous within each road type than between different road types.

The developed typology enables the authors to collect driving speed data for each road type, generate driving cycles, and analyze and compare the characteristics of the constructed driving cycles. Table SI 1 presents the distribution of features across five road types. For instance, RT1 (road type 1) shows that 67% of the segments function as tertiary roads, and 4% as secondary roads. Additionally, 56% of the segments in this road type were located in residential areas, and 32% of the segments had a speed limit of 30 km/h.

Weather typology

The significant variability in weather conditions throughout the day and year presents a significant challenge in identifying a representative driving cycle for a specific condition. For instance, hourly weather data from Montreal over one year were analyzed descriptively, exhibiting the following oscillations in factors influencing driving behavior: temperatures range from -25 to 37 degrees, humidity fluctuates between 20 and 100%, wind speed varies from 0 to 70 km/h, and precipitation (snow and rain) ranges from 0 to 28 mm. As a result, to accurately represent driving behaviors throughout the year, a framework capable of comprehending and quantifying the variability of driving conditions must first be developed. An important prerequisite is isolating the impact of weather from road features, given that literature suggests that road features can influence the effects of weather on speed (Chen et al. 2019; Chiou and Chen 2010).

Addressing this, a weather typology was developed by Yarahmadi, Morency and Trepanier, (2023b) using factors affecting speed to quantify the variability of weather conditions. Their findings suggest that weather conditions in Montreal can be categorized into eight distinct weather types, each corresponding to distinct driving behaviors. To validate their results, a road type established in the previous step was selected, and driving cycles were generated for the different weather types. As the road type remained constant while weather conditions varied, it was possible to isolate the impact of weather conditions on driving cycle parameters and ascertain if the generated weather types exert distinct influences on driving behavior. Subsequent statistical analysis revealed that the driving cycles generated under different weather conditions were notably distinct, validating the results of the new weather typology. Moreover, the developed typology provides a better representation of changes in driving cycle parameters.

In this study, the established weather typology is leveraged to generate driving cycles under diverse driving conditions. Table SI 2 presents the features of generated weather types denoted by WT.

Driving cycle generation

The process of generating a driving cycle can be broken down into several stages: collecting driving data, preprocessing the collected data, generating microtrips, clustering the generated microtrips, utilizing a one-step Markov model to construct candidate driving cycles, and eventually assessing the candidate driving cycles and selecting the representative cycle. Further details about the driving cycle generation process are provided in the following subsections.

Collecting driving data

Concerning data collection, the fleet data collection technique is capitalized on in this study, as it offers unique advantages. Firstly, data loggers (i.e., GPS) were permanently installed on a large number of vehicles such as taxis, which, in contrast to other techniques, resulted in the collection of extensive datasets. The time interval for data collection varied from 1 s to 5 min. Trips with recorded intervals exceeding 5 min were eliminated from analysis. In total, data were collected from 6292 taxis, amassing a dataset of approximately 10,567,000 journeys carried out from April 2019 to December 2020 under real-world driving conditions. In addition, secondary factors such as vehicle model, driver's age, or gender were not considered during the data collection process. Moreover, various road types and driving conditions, during peak and off-peak hours, were traversed by a wide array of drivers to transport passengers to their destinations. Consequently, due to the large volume of data that showcases driving behaviors in specific road types and driving conditions, data analysis could be repeated, leading to more robust conclusions. The data were gathered from the entire Montreal Road network, encompassing all road functions (arterial, primary, local, etc.), including approximately 9400 segments spanning a total length of 130 km, under all driving conditions. A detailed overview of the primary characteristics of this driving data is provided in Table 2.

Table 2 Driving data features

Variable	Ride ID	Timestamp	Taxi ID	Speed (km/h)	Longitude	Latitude
Type	Text	Date and Time	Text	Number	Number	Number

Data preprocessing

Upon data collection, map matching was utilized to partition the data according to road types. Following that, the driving data for each road type were subdivided into weekday and weekend groups. These groups were further partitioned into eight subgroups in accordance with weather types.

Then, a range of preprocessing techniques (e.g., outlier analysis and noise removal) were applied to the raw data to make them suitable for analysis. A significant step in this preprocessing was the modification of the timestamp frequency, which was changed from every five seconds to every second. This increase in data granularity was necessary because driving cycle parameters are calculated at a one-second interval, and it allows for a more detailed and accurate analysis of driving behaviors and patterns. To do so, the resample function in the Pandas library was first employed. Then, the KalmanFilter function in the Pykalman library was utilized to predict missing longitude and latitude values. Previous studies indicate that the Kalman filter is an effective method for filling missing values in time series data (Ribeiro 2021). Visualization is a common method used for validating the results of a Kalman filter. To do so, QGIS was employed to conduct a comparative analysis. This process involved loading both the dataset before and after the imputation, which was resampled to a one-second interval, into the QGIS platform. By overlaying these datasets within QGIS, the precision of the imputations could be visually evaluated.

Microtrip generation

As generating the driving cycle from entire trips is challenging, an average driving pattern is defined by analyzing complete trips, which leads to the division of trips into smaller, significant driving patterns called microtrips. The typical definition of a microtrip is the speed profile between two stops, a method commonly used in research. However, this approach has been criticized, particularly for urban driving on local roads, because it often involves short sequences between frequent stops, making the data analysis difficult and potentially biased. In urban areas with heavy traffic, these short microtrips are frequently removed during data preparation, potentially overlooking important congested driving patterns relevant to vehicle emissions estimates.

To address this issue, various microtrip generation techniques were explored by (Nouri 2015; Roy and Morency 2020), such as fixed time interval, fixed speed interval, intersections, event bins based on acceleration, and fixed distance interval. Spatial methods were found to be the most effective for defining microtrips, with the

distance-based method (250 m) yielding the best results (most representative driving cycles) among various spatial definitions. It is worth mentioning that fixed distance microtrips with lengths of 250 m, 500 m, 750 m, and 1000 m were evaluated. By employing thirteen driving cycle assessment parameters, the representativeness of each candidate driving cycle was determined by Nouri based on their similarity to real-world trip behaviors. Her findings concluded that a microtrip length of 250 m produced driving cycles most representative of actual driving patterns. Therefore, this technique was employed to generate microtrips with a length of 250 m. Table SI 3 displays microtrip characteristics, and Fig. 3 illustrates a speed profile segmented into microtrips under various road, weather, and day types.

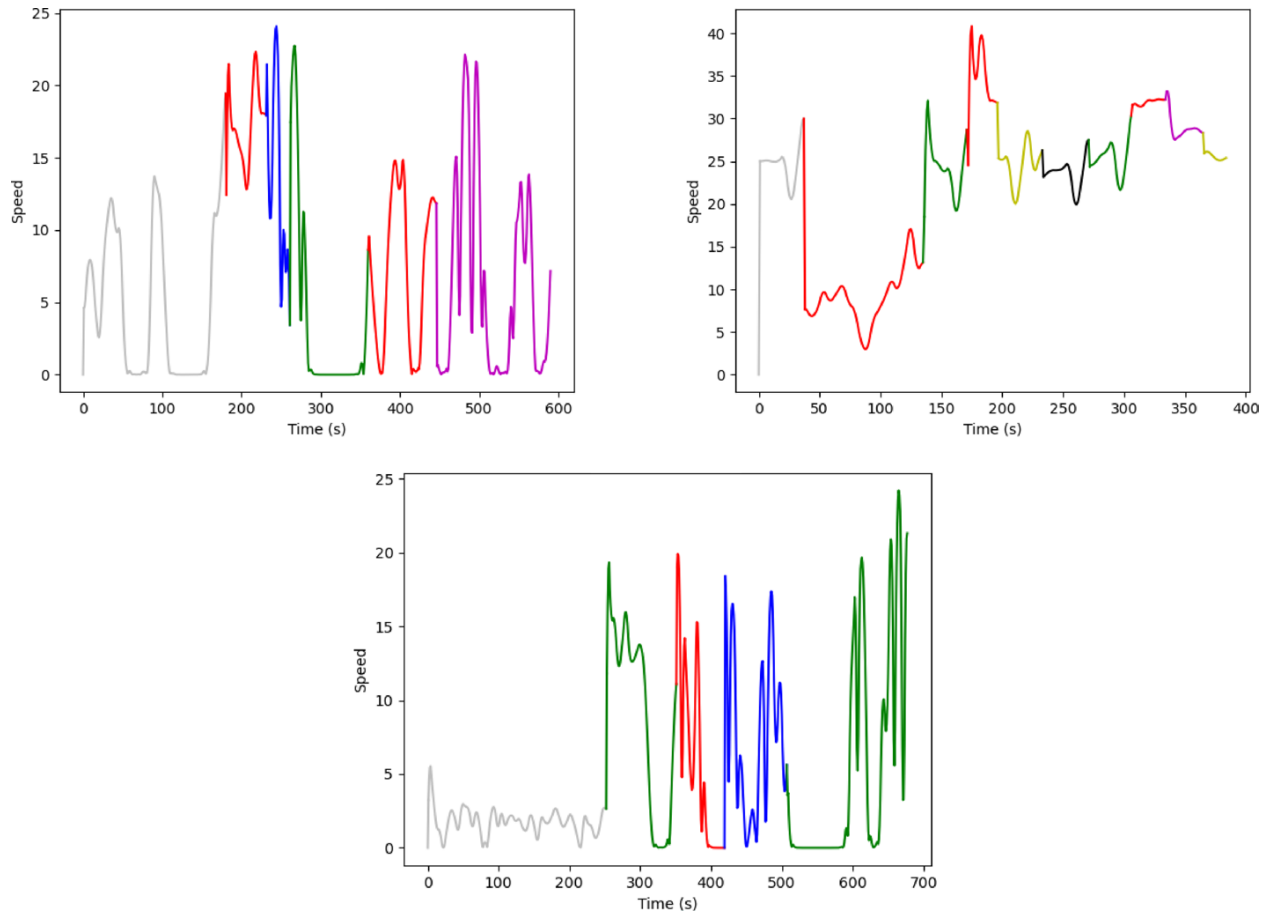


Fig. 3 Speed profile with microtrip segmentation in different types of roads, weather conditions and days

Microtrip clustering

Next, the generated microtrips are clustered, as this is essential for accurately capturing the diversity of driving behaviors and ensuring the representativeness of the driving cycle. In addition, this process reduces the complexity of the dataset and enhances computational efficiency by grouping similar driving patterns together. It also facilitates the identification and analysis of distinct driving patterns, providing valuable insights into various driving conditions. Following k-means clustering, representative microtrips were selected from each cluster based on their proximity to the cluster centroid. Specifically, the microtrip with the smallest Euclidean distance to its cluster centroid was selected as the representative sample for generating driving cycles.

By selecting representative microtrips from each cluster, the final driving cycle remains both balanced and realistic. Among various clustering techniques, k-means has been recognized as one of the most widely used in the literature (Fotouhi and Montazeri-Gh 2012; Yang Yang et al. 2020; Yuhui et al. 2018). However, it requires

the number of clusters to be predefined. A review of the literature shows that the number of clusters has ranged from 2 to 7 (He 2020; Ma et al. 2019; Mayakuntla and Verma 2018). To determine the optimal number of clusters, the Silhouette technique was used, as literature shows it performs better than other techniques (Arbelaitz et al. 2013). The results of the Silhouette analysis for different datasets indicated that the best number of clusters ranged from 7 to 10. Nevertheless, Roy and Morency (2020) suggested selecting a fixed number of clusters to isolate the impact of microtrip definitions and clustering methods on driving cycle performance. Therefore, eight clusters were adopted for all datasets. Figure 4 displays the clustering results.

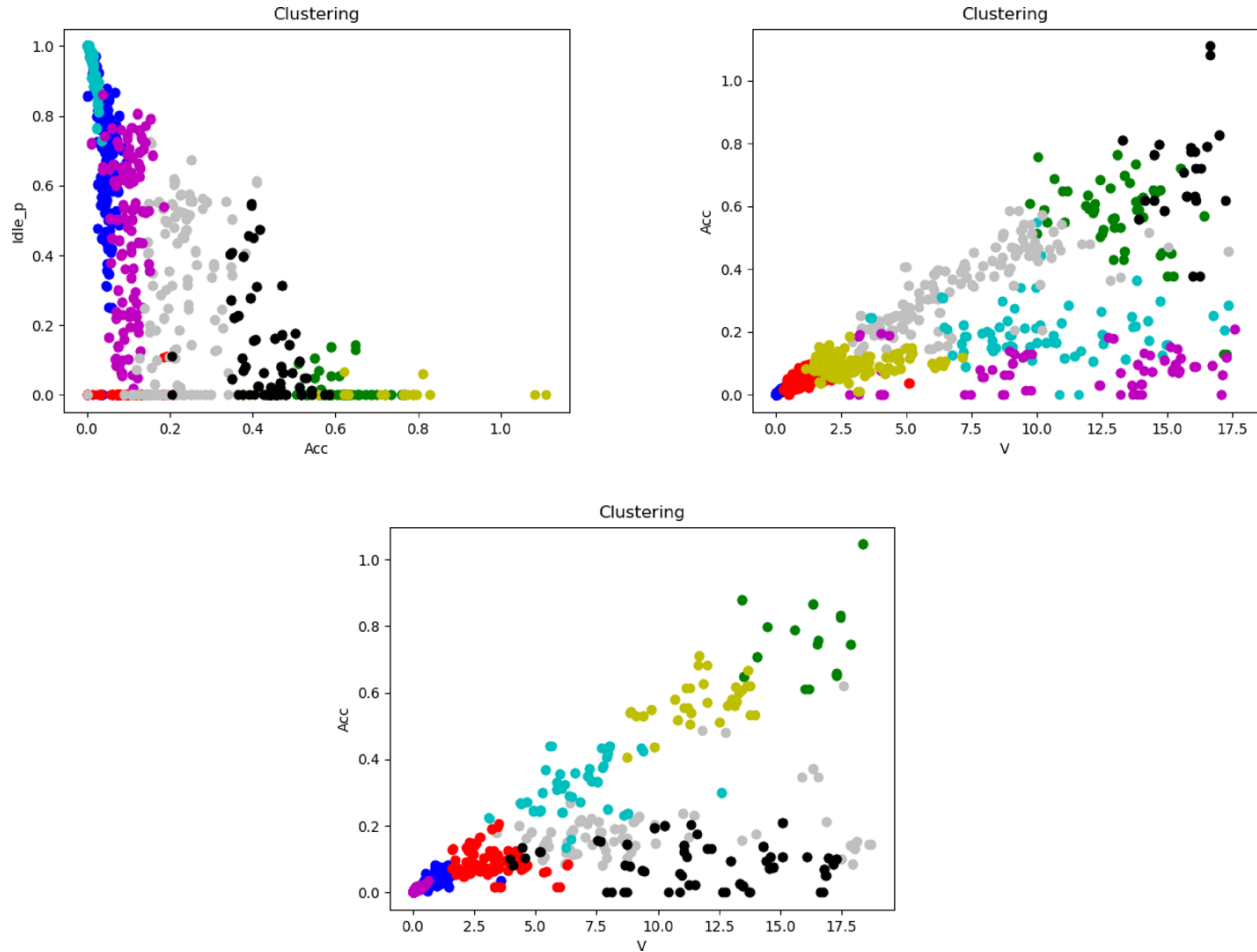


Fig. 4 Clustering results in different weather, road and weather types

Candidate driving cycles construction

To construct a driving cycle, the Markov-chains approach was employed. This process begins by chaining microtrips, initiating with the starting microtrip, and the subsequent microtrip is selected using a one-step Markov model. This method produces a transition matrix that represents the likelihood of transitioning from one cluster of microtrips to another within the dataset. The probabilities are derived from the sequences of microtrips throughout the database. The selection of each new microtrip is determined statistically by the calculated probabilities. Once a microtrip is chosen, the next one is identified based on these transition probabilities.

To perform the Markov model, three parameters must be initialized: delta speed, length of the driving cycle, and the number of iterations. The delta speed parameter defines the maximum acceptable speed difference between the end of one cluster and the beginning of the next one. This parameter ensures continuity in the driving cycle, with the recommended value by the developer being 2 km/h. The next parameter, the number of iterations,

specifies how many candidate driving cycles should be generated to identify the most representative one. The importance of iterations is tied to their effect on the reliability and reproducibility of the driving cycles produced. The selected number of iterations is intended to ensure that the resulting cycles accurately reflect real-world driving behavior. While the default value is 800, it was increased to 1200 due to the number of trips in the dataset.

The final parameter, the length of the driving cycle, is based on the average length of trips. In this study, the trip lengths ranged from 18 to 25 min, as observed between December 2020 and September 2019, respectively.

Driving cycle assessment

Ultimately, as illustrated in Table 3, thirteen parameters are used to evaluate the generated candidate driving cycles and to select the representative one. However, in contrast to the traditional approach in existing literature, this methodology selects two driving cycles to represent driving behaviors under a specific driving condition—one for weekdays and one for weekends. Thus, in total, 16 driving cycles are chosen to reflect driving behaviors for each road type. Consequently, roughly 80 driving cycles are generated, encompassing the five road types and eight weather conditions.

Table 3 Driving cycle assessment criteria

Assessment criteria	Denoted	Unit
Difference in SAFD	NA	(%)
The average speed of the entire driving cycle	V	(km/h)
Average running speed	V _r	(km/h)
Maximum speed	V _m	(km/h)
Velocity standard deviation	V_std	(km/h)
The average acceleration of all acceleration phases	Acc	(m/s ²)
Average deceleration of all deceleration phases	Dcc	(m/s ²)
The average number of acceleration-deceleration changes	Acc_std	(m/s ²)
Root means square acceleration	Acc2	(m/s ²)
Time proportion of idle mode	Idle_p	(%)
Time proportion of acceleration mode	Acc_p	(%)
Time proportion of cruising mode ($-0.1 \text{ m/s}^2 < \text{acceleration} < 0.1 \text{ m/s}^2$; average speed $> 5 \text{ m/s}$)	Cru_p	(%)
Time proportion of deceleration mode ($\text{acceleration} \leq -0.1 \text{ m/s}^2$)	Dcc_p	(%)
Time proportion of creeping mode ($-0.1 \text{ m/s}^2 < \text{acceleration} < 0.1 \text{ m/s}^2$; average speed $< 5 \text{ m/s}$)	Cre_p	(%)

Mixed models

After constructing the driving cycles, it is necessary to examine the interplay between the generated road types, weather types, and parameters of the driving cycles that were constructed. Mixed models, also known as multilevel models, are employed for this purpose. These models blend fixed and random effects to account for hierarchical, complex structures within data. They are particularly useful for handling grouped or clustered data, as they consider the interdependencies typically found in such datasets. Fixed effects are parameters that remain constant across individuals, while random effects allow these parameters to exhibit variation. In mixed-effects models, both types are incorporated, granting a higher degree of flexibility and applicability. This is particularly beneficial when dealing with clustered or grouped data because intra-group correlations are efficiently accounted for by incorporating random effects for the groups.

The model was developed using the Python programming language, utilizing the Statsmodels library. As depicted in Fig. 5, there exists a hierarchical data structure where road types, weather types, and divisions of weekdays and weekends are treated as fixed effects. Drivers in this study are treated as random effects. A total of

30 drivers were randomly selected, and driving cycles were generated for each individual driver under different driving conditions, resulting in the construction of 2400 driving cycles. It should be noted that Table 2 shows that the driving data includes a unique Taxi ID for each taxi. These IDs were used to collect driving data for taxis under various weather conditions and on different road types. Therefore, in this study, the driver—identified by the taxi ID—is considered a random effect. Ultimately, Eq. (1) was developed in the final step to examine the interplay of fixed effects and to determine their influence on the random effects.

$$DC_p = GI_p + \sum_{i=1..n} \beta_{1i,p} RT_i + \sum_{j=1..m} \beta_{2j,p} WT_j + \sum_{k=1..l} \beta_{3k,p} DT_k + \sum_{i,j,k} \beta_{ijk,p} (RT_i \cdot WT_j \cdot DT_k) + \varepsilon_p \quad (1)$$

where: DC_p : Dependent variable representing the p-th driving cycle parameter analyzed in this study (e.g., average speed, acceleration, deceleration). The index p ranges from 1 to 13, referring to the 13 driving parameters listed explicitly in Table 3. n : Total number of road types, m : Total number of weather conditions, l : Total number of day types. GI_p : Global intercept of the chosen parameter. ε_p : Random error term for the p-th parameter. $\beta_{1i,p}$: Coefficient for the effects of the i-th road type on the p-th parameter. $\beta_{2j,p}$: Coefficient for the effects of the j-th weather condition on the p-th parameter. $\beta_{3k,p}$: Coefficient for the effects of the k-th type of day on the p-th parameter. $\beta_{ijk,p}$: Coefficients for the three-way interaction effects among the fixed factors for the p-th parameter. RT_i , WT_j , DT_k : Indicator variables correspond to each level of the categorical fixed effects.

Equation (1) is divided into two components: a summation and a product. The summation is used to examine the impact of external factors on driving cycle parameters, while the product is employed to analyze the interaction between road types and weather conditions during weekdays and weekends.

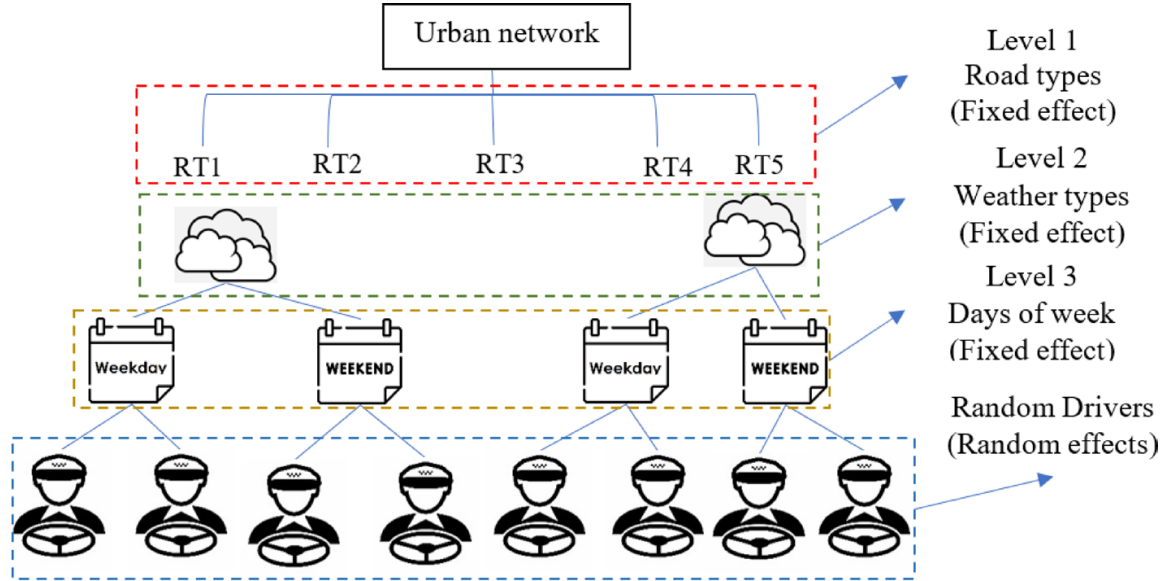


Fig. 5 Hierarchical structure of data of the present study

Optimal number of driving cycles

To decide on the optimal number of driving cycles, a two-stage process was employed in this study. Initially, information loss was used to determine the appropriate level of data aggregation. After that, variability and information loss indicators were utilized to identify the optimal number of driving cycles.

Step 1: Determining aggregation level of driving data

One of the primary hurdles in identifying the optimal number of driving cycles, as discussed in Sect. "Collecting driving data and driving conditions", is determining the aggregation level of driving data. This is accomplished by using information loss as a tool to statistically investigate and contrast the distribution of driving

cycle parameters against the distribution of all driving data. Information loss may occur if the chosen driving cycles do not satisfactorily represent all driving behaviors. Hence, it is critical to confirm that the method proposed for aggregating driving data results in minimal information loss.

Two metrics are utilized in this research to quantify information loss. The first metric, expressed in Eq. (2), was proposed by (Liu et al. 2020). This metric computes the ratio of the range of selected driving cycle parameters to the range of driving cycle parameters for all drivers and is referred to as the ratio of sampled to actual range.

$$\text{Ratio of selected DC to all DCs} = (1/N) \sum_{i=1..N} [1 - R_{dcp}^s / R_{mtp}^a] \quad (2)$$

Where i : set of driving cycle parameter, $i \in \{1, 2, \dots, N\}$. $R_{dcp}^s = |\text{pmax}_i^s - \text{pmin}_i^s|$, range of i -th selected driving cycle parameter (s) in an aggregation level. $R_{mtp}^a = |\text{pmax}_i^a - \text{pmin}_i^a|$, actual range (a) of corresponding microtrip parameter.

The second indicator, Eq. (3), calculates the deviation of quartiles of the selected driving cycle parameters from quartiles of microtrip parameters.

$$\text{Distribution-based information loss} = (1/N) \sum_{i=1..N} [(1/M) \sum_{j=1..M} |PQ_{ij}^s - PQ_{ij}^a| / |PQ_{ij}^a|] \quad (3)$$

where i : set of driving cycle parameter, $i \in \{1, 2, \dots, N\}$. j : set of quarter, $j \in \{1, 2, 3\}$. PQ_{ij}^s = the j -th quartile of i -th selected driving cycle (s) parameter in an aggregation level. PQ_{ij}^a = the j -th quartile of actual range (a) of corresponding driving cycle parameter.

Step 2: Determining optimal number of driving cycles

In the subsequent phase, both variability and information loss indicators are utilized to ascertain the optimal number of driving cycles. To do this, a similarity matrix is first constructed to gauge the likeness between all the created driving cycles. If two driving cycles exhibit a high degree of similarity, it implies that they are exhibiting analogous driving behaviors and therefore should be merged. Hence, the SciPy library in Python is employed to construct a similarity matrix accompanied by a dendrogram. The matrix provides a pairwise comparison of driving cycles, and the dendrogram illustrates the similarity levels of driving cycles within the matrix.

Next, it is crucial to aggregate driving cycles exhibiting the highest similarity and then to measure the impact of this procedure on both the variability of driving cycles and information loss. Variability reflects the diversification of the driving cycles; therefore, a decrease in variability indicates a reduction in the representativeness of the selected driving cycles. In simpler terms, the optimal number of driving cycles is defined as the smallest number that maintains maximum variability while minimizing information loss.

To measure the variability of driving cycles, the Proportional Variability (PV) indicator is employed in this study, as described in Eq. (4) and originally developed by (Heath and Borowski 2013). The domain of PV is defined on a truly proportional scale [0, 1].

$$PV = (\sum_{p=1..M} [2 \cdot \sum_{comb.} D(z_{pi}, z_{pj}) / (n(n-1))]) / M \quad (4)$$

where i, j be indices such that $i, j \in \{1, 2, \dots, n\}$ and $i \neq j$. These indices are used to calculate the relative differences between the parameters of driving cycles. For a given parameter p , z_{pi} represents the value of p in the i -th driving cycle, and z_{pj} represents the value of p in the j -th driving cycle. The function $D(z_{pi}, z_{pj}) = |z_{pi} - z_{pj}| / \text{Max}(z_{pi}, z_{pj})$ is defined to quantify the relative difference between these two values. n is the total number of driving cycles and M is the total number of driving cycles parameters.

The methodology described above was applied to a large-scale driving dataset collected from the Montreal road network. The next section presents the results of the driving cycle construction, statistical validation, and comparative evaluation against traditional approaches.

Result

This section reports the key findings from the application of the proposed methodology. The analysis includes assessment of information loss, similarity between driving cycles, variability across conditions, and evaluation of optimal aggregation levels. It also provides a comparison with standard and local driving cycles to validate the effectiveness of the developed framework.

Mixed models to study the influences of fixed effects on driving cycle parameters

As previously discussed, a mixed model has been formulated to investigate the interaction between external factors. This interaction among fixed effects is intended to represent actual driving conditions, as real-world driving behaviors are influenced by road types, weather conditions, and the day of the week. As a result, considering the interaction between these exogenous factors can contribute to a better understanding of driving behaviors under realistic driving conditions.

Figure 6 presents the results of this model for all driving cycle parameters. The x-axis represents the road types, and the y-axis shows the values of the driving cycle parameters for different weather conditions and weekdays/weekends.

In all graphs, it is evident that the driving cycle parameter values change with variations in fixed effects. This supports the premise that driving cycle parameters are functions of road type, weather type, and weekdays/weekends.

Additionally, the slope of the lines in these graphs indicates the direction and magnitude of the relationship between road type and driving cycle parameters under specific weather conditions and times of day. For instance, a steep upward slope signifies a strong positive relationship, suggesting that the creeping time increases as the road type changes from one category to another. A flat line indicates no relationship, whereas a downward slope signifies a negative relationship.

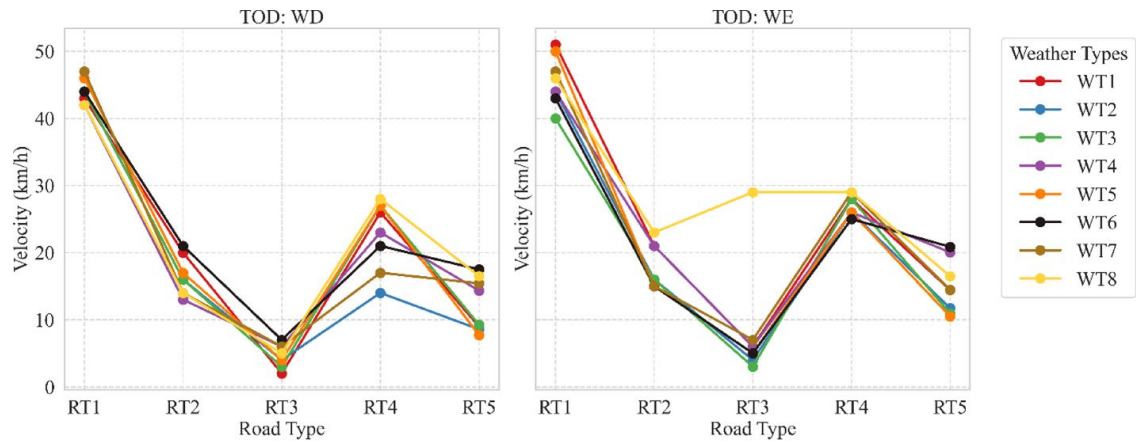
Upon examining the generated graphs, it becomes clear that driving cycle parameters vary across all road types, weather conditions, and days of the week. For example, velocity is observed to decrease significantly from road type 1 to road type 2 under all weather conditions.

Table 4 presents the most significant fixed effect associated with each driving cycle parameter. Notably, the p-values for all factors were found to be less than 0.05, indicating a statistically significant association between the driving cycle parameters and the fixed effects.

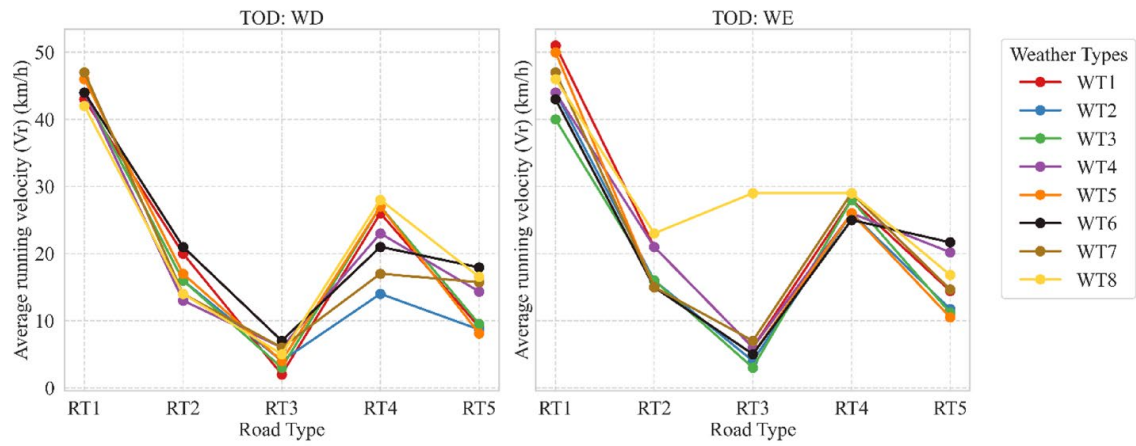
As demonstrated in Eq. (4), the constructed mixed model examines the relationship between fixed effects and driving cycle parameters independently. This enables the model to explore the impact of a single factor while controlling for the effects of others. Furthermore, it assesses the impact of interactions between fixed effects on driving cycle parameters.

The results underscore the importance of considering all fixed effects simultaneously when analyzing driving cycle parameters. For instance, regarding velocity, the most influential isolated factors are RT3, WT7, and WE, while in real driving conditions, the interaction RT3WT8WE has the most significant impact.

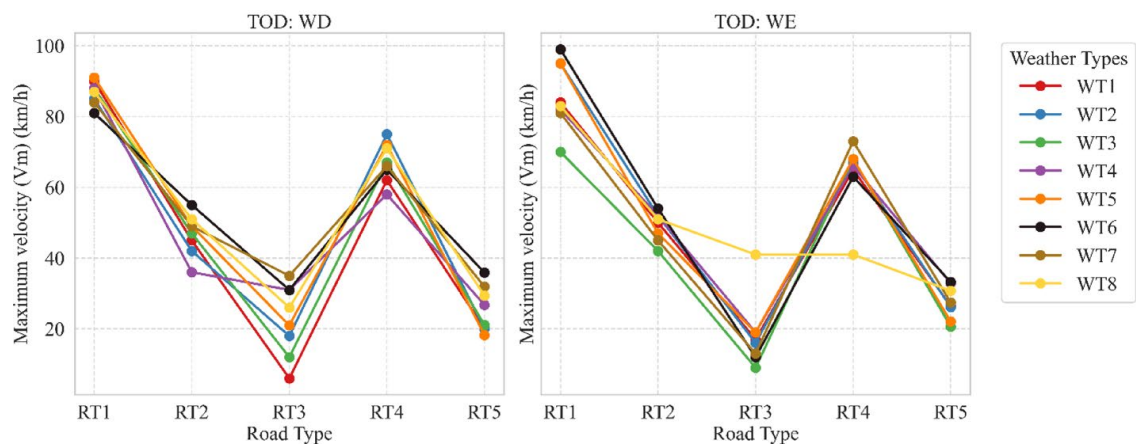
Fig. 6 Changes of driving cycle parameters in different road types, weather types in weekend and weekdays



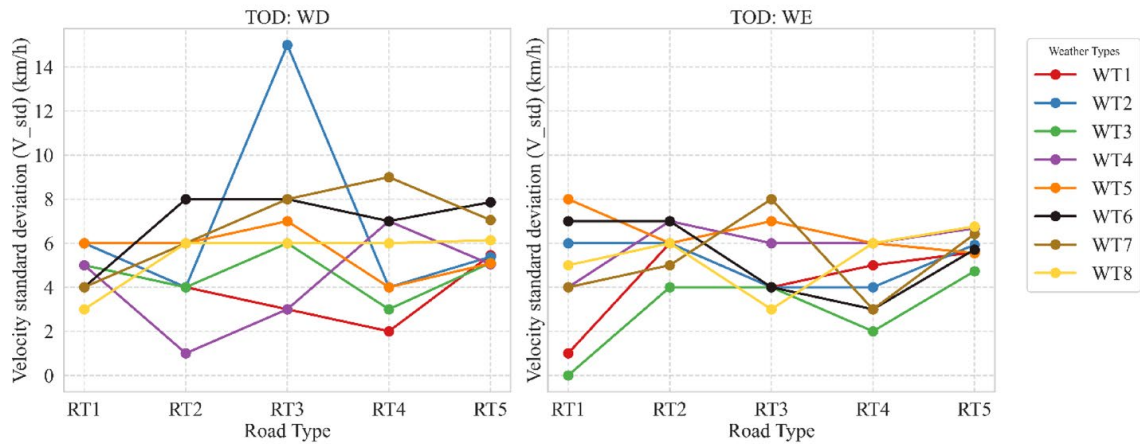
(Velocity)



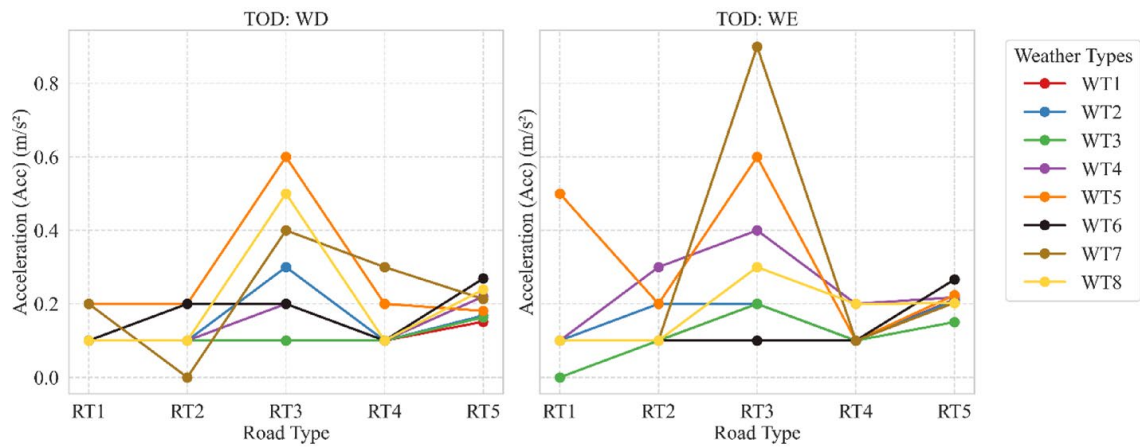
(Average Running Velocity)



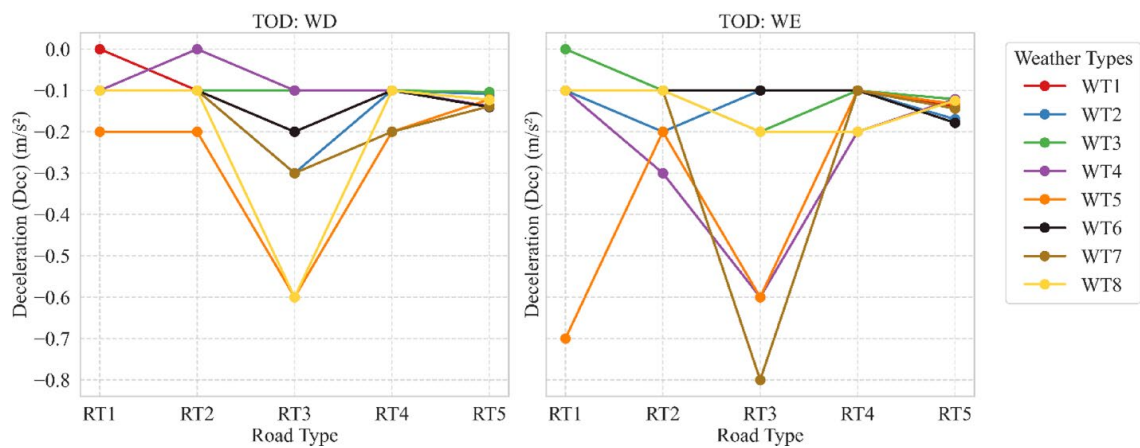
(Maximum velocity)



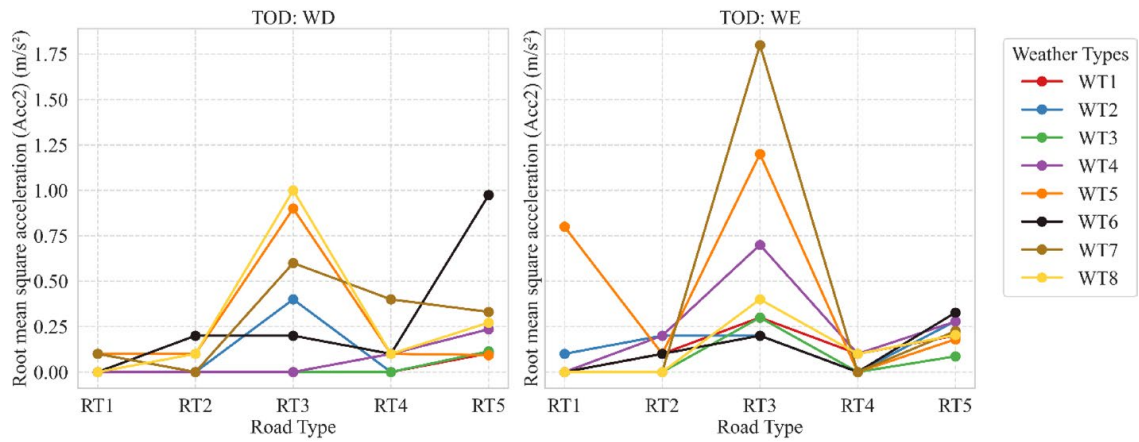
(Standard Deviation of velocity)



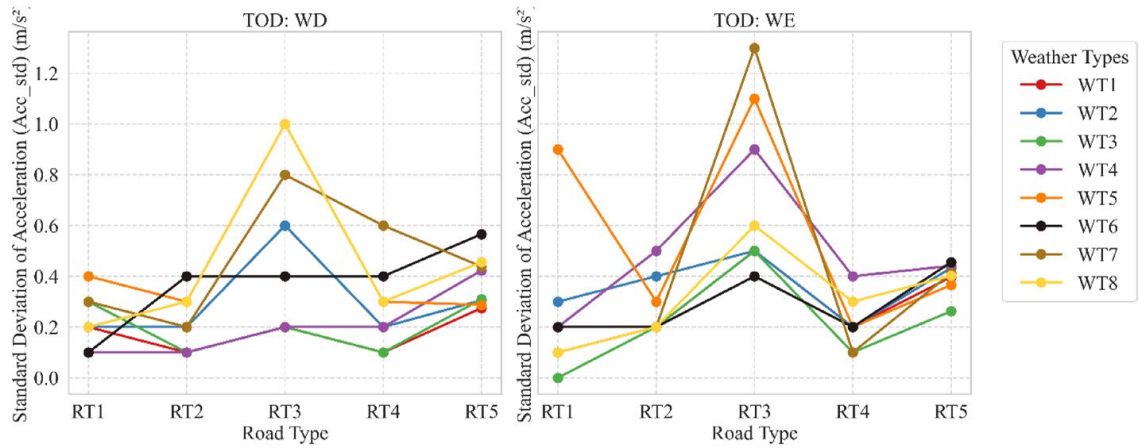
(Acceleration)



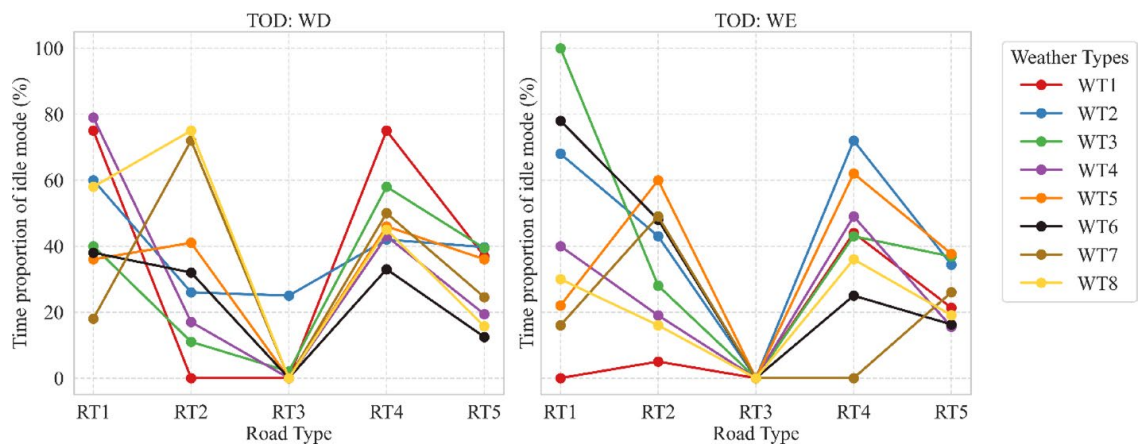
(Deceleration)



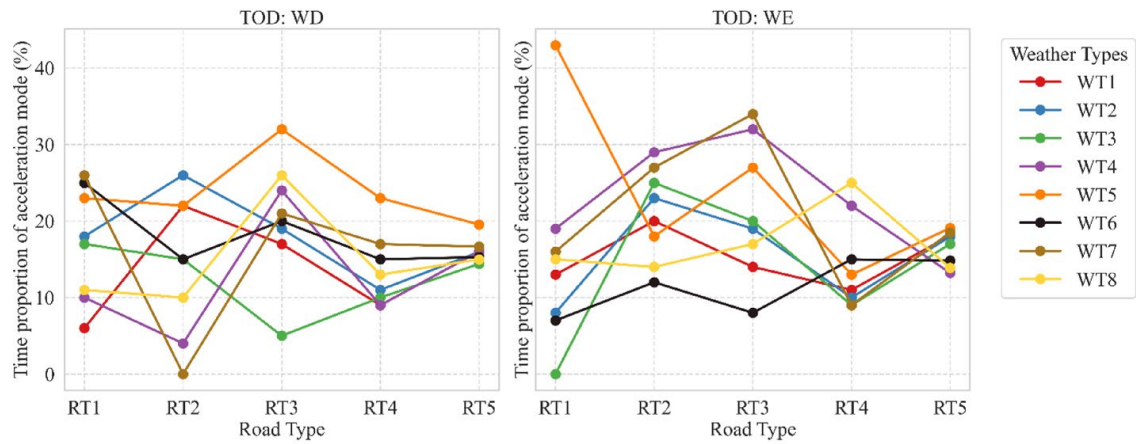
(Root mean square acceleration)



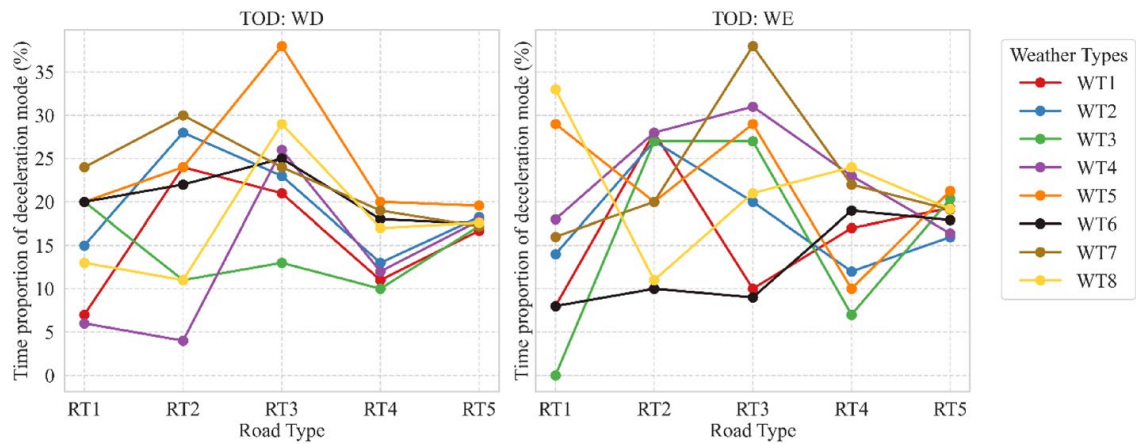
(Standard deviation of acceleration)



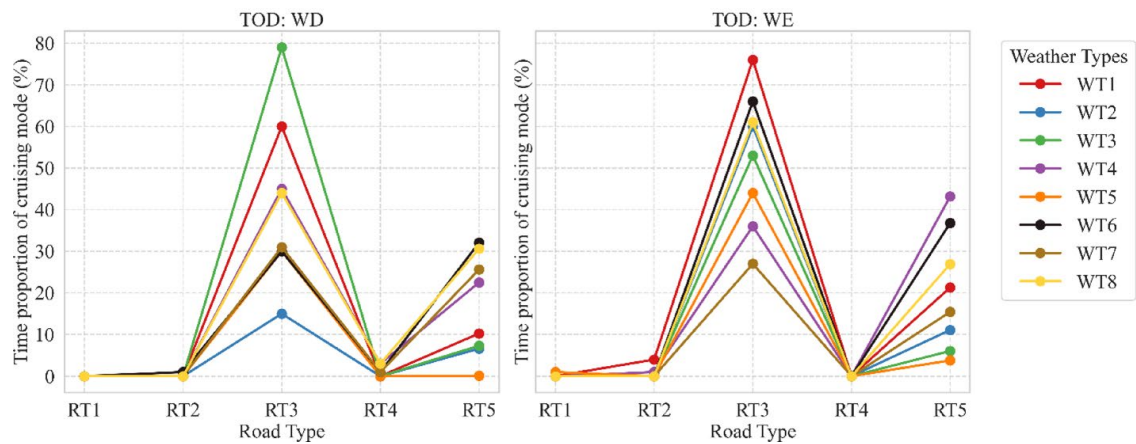
(Time proportion of idle mode)



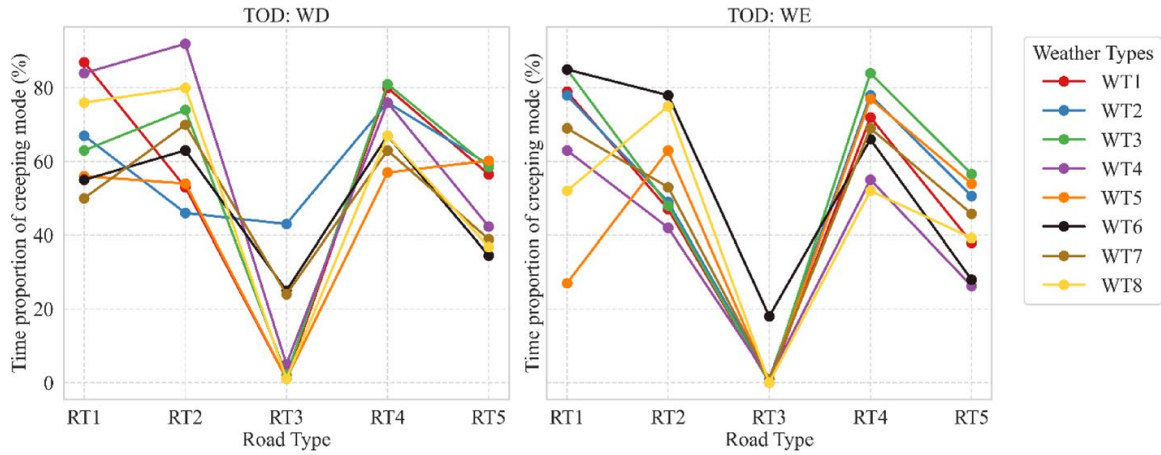
(Time proportion of acceleration mode)



(Time proportion of deceleration mode)



(Time proportion of cruising mode)



(Time proportion of creeping mode)

Table 4 The most important exogenous factors for each driving cycle parameters

Driving cycle parameters	Road type	Weather type	Days of week	Combination of fixed effects (real driving conditions)
V	RT3	WT7	WE	RT3WT8WE
V_std	RT4	WT8	WD	RT3WT2WD
Vm	RT5	WT6	WD	RT3WT6WD
Acc	RT5	WT5/WT7	WE/WD	RT3WT7WE
Dcc	RT3	WT5	WD	RT3WT4WD; RT3WT7WE
Idle_p	RT2, RT3	WT7	WD	RT3WT3WE
Acc_p	RT2	WT7	WE	RT2WT7WE
Dcc_p	RT2	WT7	WE	RT3WT3WE
Cru_p	RT3	WT3	WD	RT3WT3WD
Cre_p	RT3	WT7	WE	RT3WT3WE

Determining aggregation level for the developed methodology

As elaborated in Sect. "Optimal number of driving cycles", the initial step to identify the optimal number of driving cycles involves the use of information loss indicators to ascertain the ideal level of data aggregation. This is done by following the methodology delineated in Sect. "Driving cycle generation" for five different levels of data aggregation. These are:

1. The highest level of aggregation, in which all driving data are amalgamated and only one driving cycle is produced.
2. The second level, where data are grouped according to road types, resulting in five driving cycles.
3. The third level, where data are aggregated based on different driving conditions across various road types, resulting in a total of 40 driving cycles.
4. The fourth level applies the recommended methodology of this study, where data are grouped according to road types, weather conditions, and weekdays or weekends, resulting in 80 driving cycles.
5. The fifth level involves the generation of microtrips for each trip.

Information loss is calculated using Eqs. (3 and 4). Figure 7 presents the variation of information loss for different driving cycle parameters across the defined levels. The information loss values in Fig. 7 range from 0, signifying no information loss, to 1, representing the highest level of information loss. It is evident that the highest and lowest information loss levels are observed for levels L1 and L5, respectively. The results confirm that the methodology presented in this study yields the minimum possible level of information loss. It is worth noting that Fig. 7 displays information loss for indicators Eqs. (3 and 4), as well as the overall information loss of all driving cycle parameters. The latter allows for the determination of average information loss using the mean values of all driving cycle parameters across different levels of aggregation.

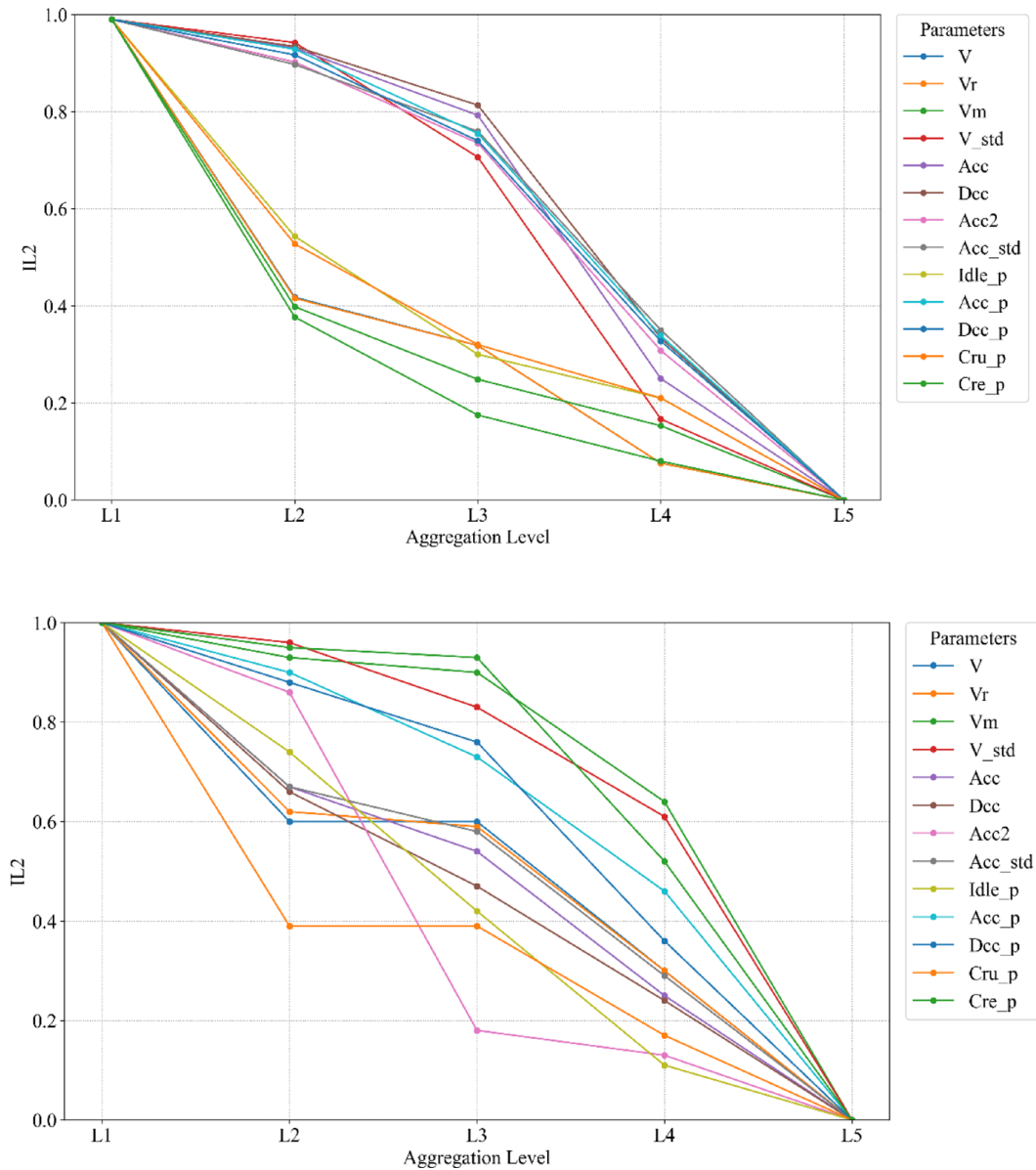


Fig. 7 Comparison of information loss in different aggregation level for information loss indicators

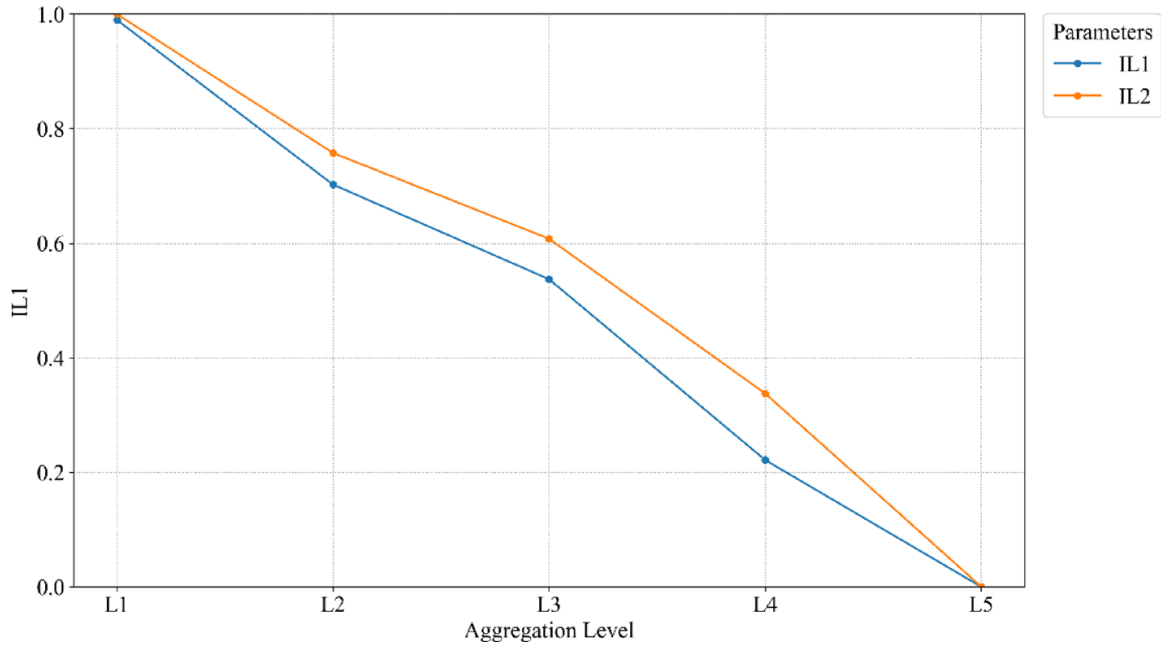


Fig. 7 (continued) Overall information loss across aggregation levels

Step 2: Determining the optimal number of driving cycles

In order to determine the optimal number of driving cycles, it is critical to assess the similarity of the produced driving cycles. This is accomplished by constructing a similarity matrix, where each array represents the similarity between two driving cycles. Figure 8 presents a heatmap of the similarity matrix, with similarity intensifying from dark to light shades. Moreover, the associated dendrogram indicates levels of similarity. The dendrograms on both sides of the matrix are identical and are included to provide a clearer presentation of similarity levels. The value in each cell represents the similarity between a pair of driving cycles. The highest similarity observed is 92%, between RT4WT1WE and RT4WT2WE, while the lowest similarity is 0%. Figure 9 provides a breakdown of the different levels of similarity. Similarity levels of less than 40% account for the largest proportion, indicating that most of the generated driving cycles represent distinct driving behaviors.

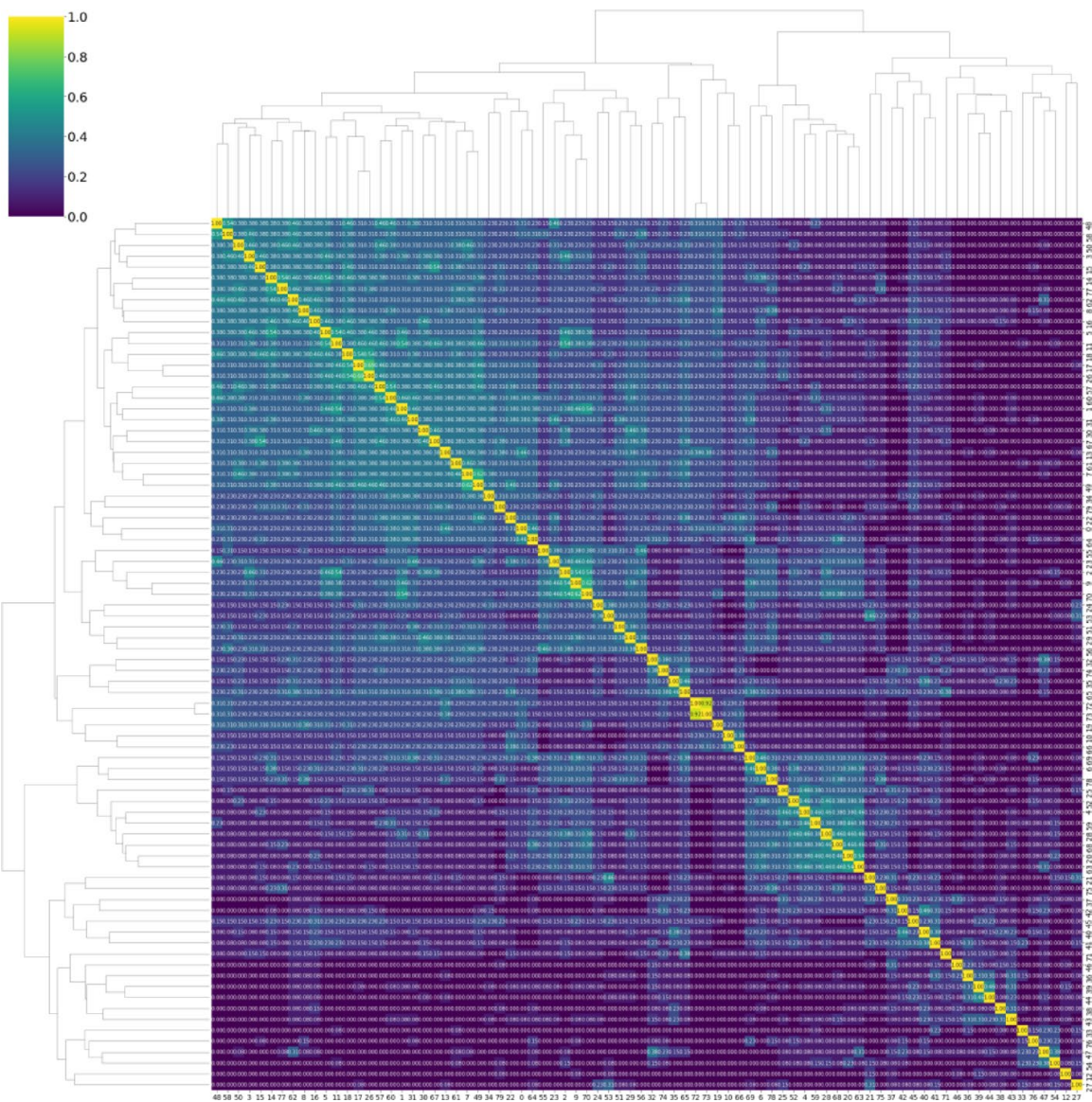


Fig. 8 Similarity matrix of the generated driving cycles

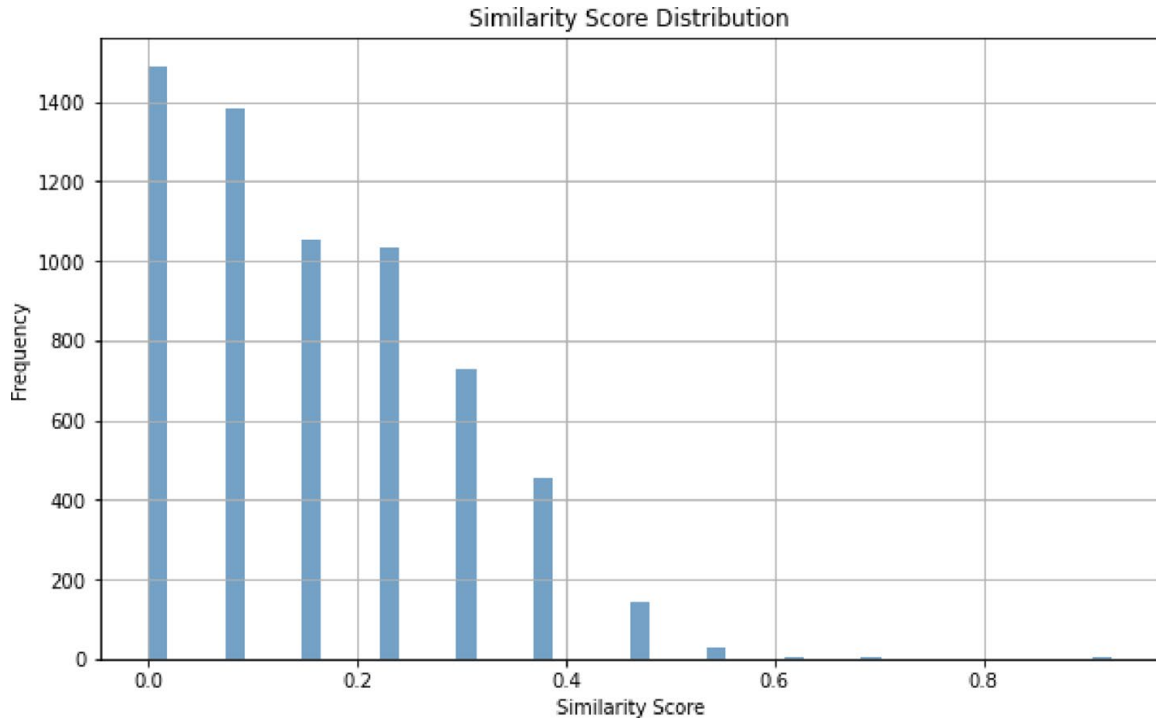


Fig. 9 Share of similarity level in the similarity matrix

To identify the optimal number of driving cycles, the similarity matrix was utilized to pinpoint and merge driving cycles exhibiting the highest similarity at each similarity level. Essentially, within each level of similarity, driving cycles that share the highest degree of similarity were merged into one representative cycle. For instance, at the 69% similarity level, four driving cycles were present; therefore, each pair of similar driving cycles had their data aggregated, and a new driving cycle was calculated as a substitute, yielding a total of two derived cycles. Subsequently, information loss and variability were calculated using Eqs. (3 and 4), respectively. This recursive process continued until variability reached its maximum and information loss reached its minimum.

Figure 10 illustrates the changes in information loss (based on the average of both indicators) and variability across different similarity levels. The blue graph depicts the fluctuation of information loss across various similarity levels. Notably, as the similarity level decreases from 92 to 62%, information loss reaches its minimum. Subsequently, as merging continues, information loss begins to rise.

Moreover, the red graph shows an increase in variability as the similarity level decreases from 92 to 62%, moving from 0.6734 to its peak at 0.6896, and then declining to 0.6667, maintaining this trend at lower similarity levels. The red circle highlights the point of minimum information loss and maximum variability, which occurs at the 62% similarity level.

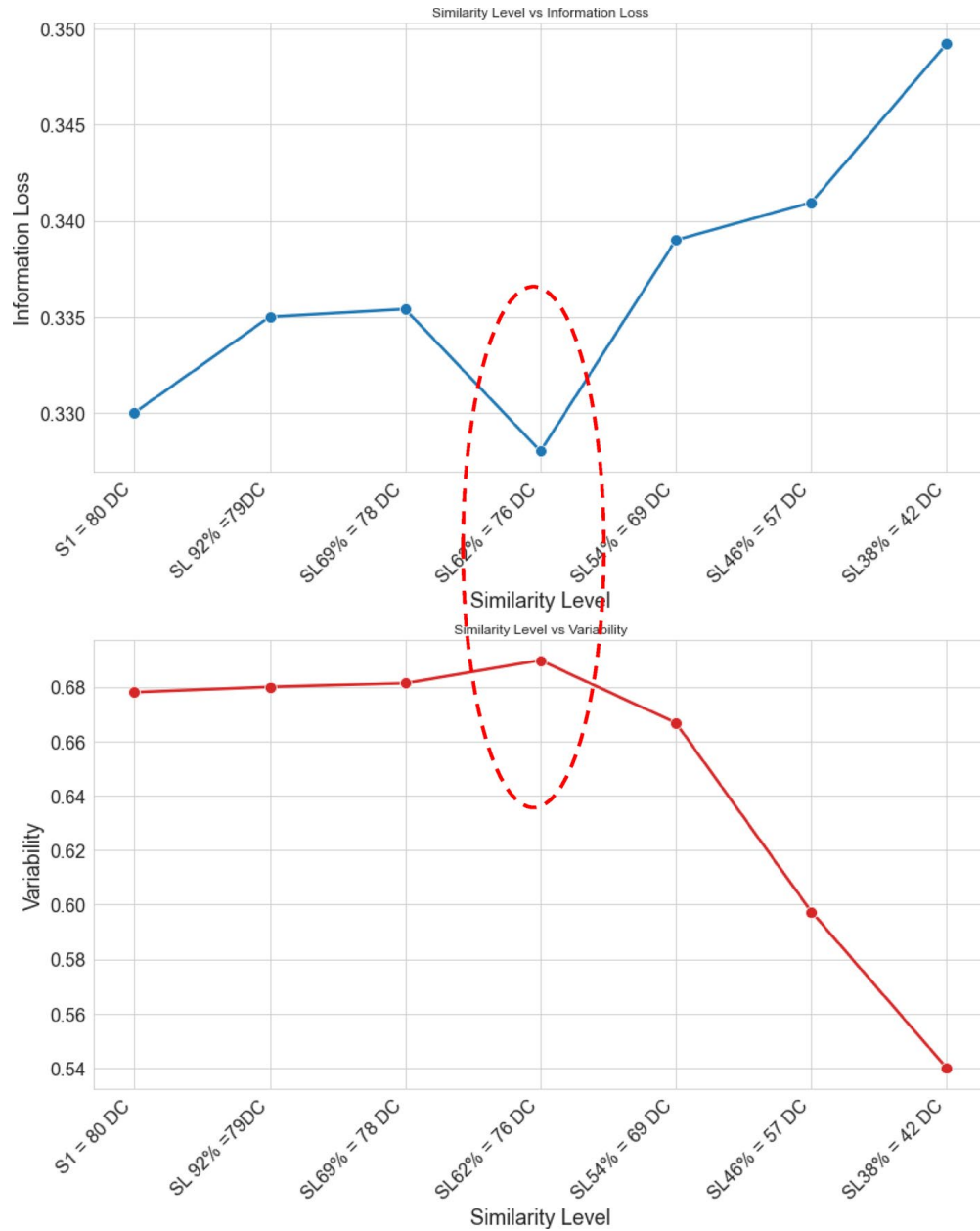


Fig. 10 Similarity level, driving cycle number vs. information loss and variability

In conclusion, by considering the results of both information loss and variability, it is determined that the optimal number of driving cycles required to represent driving behaviors across all road types and driving conditions in Montreal is 76. Table SI 4 presents the parameters of the selected driving cycles.

Comparative evaluation of proposed driving cycles versus traditional approaches

To illustrate the advantages and the accuracy of the newly developed driving cycles generated by our proposed context-aware method, a comparison was conducted against well-established traditional approaches—namely, the local Montreal driving cycle (Nouri and Morency, 2015) and the Standard Driving Cycle (SDC) derived from Canada's 5-cycle test. The primary goal of this comparative evaluation was to determine whether our method, which explicitly accounts for road and weather conditions as well as

weekday/weekend variability, offers improved representation and accuracy in capturing urban driving behavior.

Initially, a descriptive statistical analysis was performed to outline key differences between parameters of the newly developed driving cycles and those of traditional methods. Specifically, key parameters such as average velocity, maximum speed, acceleration/deceleration rates, and proportions of different driving modes (idle, cruising, etc.) were assessed. Preliminary observations revealed notable deviations; for example, velocity ranges captured by our method spanned broader values (approximately 26 to 48 km/h) compared to the narrower ranges reported by traditional approaches (approximately 25 to 34 km/h). This indicates that conventional driving cycles may lack the representational capacity required for reflecting real-world urban driving variability influenced by diverse road and weather conditions.

To statistically confirm whether the observed differences were significant, a One Sample t-test was employed, evaluating each driving cycle parameter against those from the traditional cycles. The null hypothesis (H_0) asserted that there is no statistically significant difference between the mean values of each parameter in our generated driving cycles (μ) and those reported in existing local and standard driving cycles (μ_0). The alternative hypothesis (H_1) suggested a significant difference between these mean values. The hypotheses are thus defined as:

$$H_0: \mu = \mu_0, H_1: \mu \neq \mu_0$$

The test statistic is computed using the standard t-test formulation, where s represents the standard deviation and n the number of newly constructed driving cycles.

Table 5 summarizes the outcomes of the t-tests, clearly illustrating significant differences across all driving cycle parameters when compared to both the local Montreal driving cycle and Canada's SDC. All observed p-values were consistently below 0.05, reinforcing that the proposed approach significantly diverges from previous methods.

These findings underscore two important conclusions:

1. **Enhanced Representativeness:** The broader range and significant differences in parameter values indicate that traditional cycles do not sufficiently reflect the variability observed under actual diverse driving conditions.
2. **Context-Aware Advantage:** By explicitly incorporating road types, weather conditions, and temporal variations (weekday/weekend distinctions), the new methodology significantly enhances the precision of driving cycle representation, thereby offering a more robust basis for accurate emissions estimation.

Overall, this comparative analysis clearly demonstrates the superiority and enhanced accuracy of the proposed method over conventional approaches, highlighting its suitability and reliability for estimating vehicle emissions in complex urban environments.

Table 5 Differences between the developed driving cycles and local and standard driving cycles

		V	Vr	Vm	V_std	Acc	Dcc	Acc2	Acc_std	Idle_p	Acc_p	Dcc_p	Cru_p	Cre_p
Proposed model	Min-WDs	13	13	17	1	0	-0.6	0	0.1	0%	0%	1%	0%	1%
	Max-WDs	58	58	91	15	0.8	0	1.5	1.2	79%	32%	38%	79%	92%
	Min-WE	12	12	16	0	0	-0.8	0	0	0%	7%	7%	0%	0%
	Max-WE	50	50	99	8	0.9	0	1.8	1.3	78%	43%	38%	76%	85%
	μ	30	38	57	5.25	0.2	-0.2	0.17	0.33	31%	16%	17%	11%	47%
	s	16.52	12	22.92	2.29	0.16	0.16	0.33	0.35	25%	8%	8%	22%	10%
(Nouri 2015) (μ_o)		23	32	104	—	0.43	-0.5	—	0.29	19.9%	34.5%	12.5%	31.2%	1.6%
5-cycle test (city) (μ_o)		34	—	90	—	5.3	—	—	—	18%	—	—	—	—
One sample T-test (Nouri 2015)	T-statistic	2.62	4.37	-21.70	—	-10.29	16.84	—	2.27	6.63	-19.21	4.73	-6.89	42.16
	P-value	0.01	0.00	0.00	—	0.00	0.00	—	0.03	0.00	0.00	0.00	0.00	0.00
One sample T-test 5-cycle test (city)	T-statistic	-3.64	—	-15.80	—	-240.9	—	—	—	7.43	—	—	—	—
	P-value	0.00	—	0.00	—	0.00	—	—	—	0.00	—	—	—	—

Conclusion

This paper aimed to address three key gaps in the literature on driving cycle generation: (1) determining the optimal number of driving cycles, (2) proposing a novel methodology for the aggregation of driving data, and (3) developing a mixed model to examine the influence of exogenous factors on driving cycle parameters.

Firstly, a driving cycle is a dynamic construct with parameters that change over time and location. However, in the existing literature, typically one or a few driving cycles have been used to represent driving behaviors. In response to this, a novel methodological framework was developed in this study to create and determine the optimal number of driving cycles that adequately represent driving behaviors under all driving conditions across an urban street network. The results suggest that 76 driving cycles are sufficient to capture the maximum variability in driving behaviors, thereby minimizing the uncertainty arising from information loss.

Secondly, data aggregation can cause discrepancies between real driving behavior and the generated driving cycle, which may introduce uncertainty into the driving cycle generation process. Many studies have not accounted for the impact of data aggregation on information loss and the accuracy of the generated driving cycles, often aggregating all driving data collected from various road types and weather conditions. This study proposed an approach that uses two indicators to determine the ideal level of data aggregation for generating driving cycles. This methodology aims to minimize information loss and enhance the representativeness of the generated driving cycles.

Thirdly, this study pioneered the development of a multilevel mixed model to examine the influence of road types and weather conditions on driving cycle parameters. The mixed model effectively demonstrated these intricate relationships, contributing to a deeper understanding of the dynamic nature of driving cycles.

While this study focuses on the impact of route, weather, and day type on driving cycles, the significant influence of traffic composition on driving behavior is acknowledged. However, due to constraints related to the availability of traffic flow data, it was not feasible to incorporate traffic composition variables such as vehicle type diversity and overall traffic density in the current analysis. Future research could benefit from including these factors to further improve the accuracy and representativeness of the driving cycles. Therefore, the findings of this study should be interpreted in light of these limitations.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1007/s11116-025-10631-8>.

Author's contribution

Asad Yarahmadi: Conceptualization, Data curation, Formal analysis, Methodology, Writing—original draft, Writing—review & editing, Funding acquisition. **Catherine Morency:** Conceptualization, Supervision, Resources, Funding acquisition, Reviewing & editing. **Martin Trépanier:** Conceptualization, Supervision, Funding acquisition, Reviewing & Editing.

Funding

The analysis presented in this paper was conducted as part of the research program of the Chair in the Transformation of Transportation (CTT), which is mainly financed by Quebec's Ministry of Economics and Innovation (MEI).

Data availability

Data will be available upon request.

Declarations

Competing interests: The authors declare no competing interests.

References

- Abele, L., and Mette M.: "The Relationship between road design and driving behavior." In 3rd International Conference on Road Safety and Simulation, 17 (2011).
- Ahmed, M. et al.: *Driver Performance and Behavior in Adverse Weather Conditions: An Investigation Using the SHRP2 Naturalistic Driving Study Data—Phase 2.* (2018)
- Al-Samari, A.: Study of emissions and fuel economy for parallel hybrid versus conventional vehicles on real world and standard driving cycles. *Alex. Eng. J.* 56(4), 721–726 (2017)
- Andrade, G.M., Araújo, F.W., Santos, M.P., Magnani, F.S.: Standardized comparison of 40 local driving cycles: energy and kinematics. *Energies* 13(20), 5434 (2020)
- Antonson, H., Mårdh, S., Wiklund, M., Blomqvist, G.: Effect of surrounding landscape on driving behaviour: a driving simulator study. *J. Environ. Psychol.* 29(4), 493–502 (2009)
- Arbelaitz, O., et al.: An extensive comparative study of cluster validity indices. *Pattern Recogn.* 46(1), 243–256 (2013)
- Arun, N.H., Mahesh, S., Ramadurai, G., Shiva Nagendra, S.M.: Development of driving cycles for passenger cars and motorcycles in Chennai, India. *Sustain. Cities Soc.* 32, 508–512 (2017)
- Chauhan, B.P., Joshi, G.J., Parida, P.: Development of candidate driving cycles for an urban arterial corridor of Vadodara city. *Int. J. Transp. Econ. Eng. Law* 4(75), 16 (2020)
- Chen, C., et al.: Influence of adverse weather on drivers' perceived risk during car following based on driving simulations. *J. Mod. Transp.* 27(4), 282–292 (2019)
- Chiun, C.Y., Chen, T.C.: Direct and indirect factors affecting emissions of cars and motorcycles in Taiwan. *Transportmetrica* 6(3), 215–233 (2010)
- Coropolis, S., Berloco, N., Intini, P., Ranieri, V.: Road design influence on driving behaviors: the influence of curve design, a case study. *Transp. Econ. Manag.* 3, 104–116 (2025)
- Cui, M., Levinson, D.: Accessibility and the Ring of Unreliability. *Transp. Transp. Sci.* 14(1–2), 4–21 (2018)
- Czégé, L., Vámosi, A., Kocsis, I.: Review on construction procedures of driving cycles. *Int. J. Eng. Manag. Sci.* 5(2) (2020). <https://doi.org/10.21791/IJEMS.2020.2.31>
- Esser, A., Rinderknecht, S.: "Process for the validation of using synthetic driving cycles based on naturalistic driving data sets." In 2020 IEEE 23rd International Conference on Intelligent Transportation Systems, ITSC 2020 (2020).
- Faramarzpour, H., et al.: Specification of a greenhouse in cold climate condition, mathematical model and optimization. *J. Building Phys.* 48(4), 579–608 (2025). <https://doi.org/10.1177/17442591241240798>
- Faria, M.V., Baptista, P.C., Farias, T.L., Pereira, J.M.S.: Assessing the impacts of driving environment on driving behavior patterns. *Transportation* 47(3), 1311–1337 (2020)
- Fotouhi, A., Montazeri-Gh, M.: Tehran driving cycle development using the K-means clustering method. *Scientia Iranica* 20(2), 286–293 (2012)
- Gebisa, A., Gebresenbet, G., Gopal, R., Nallamotheu, R.B.: Driving cycles for estimating vehicle emission levels and energy consumption. *Future Transp.* 1(3), 615–638 (2021)
- Hao, Y., et al.: Developing driving cycles for garbage trucks to estimate fuel consumption. *Transp. Res. Part D Transp. Environ.* 136, 104469 (2024)
- He, Y.: "Research on the Construction Method of Vehicle Driving Cycle Based on Mean Shift Clustering." *arXiv* (2020).
- Heath, J.P., Borowski, P.: Quantifying proportional variability. *PLoS ONE* 8(12), e84074 (2013)
- Huertas, J.I., Giraldo, M., Quirama, L.F., Díaz, J.: Driving cycles based on fuel consumption. *Energies* 11(11), 3064 (2018)

- Huertas, J.I., Quirama, L.F., Giraldo, M., Díaz, J.: Comparison of three methods for constructing real driving cycles. *Energies* 12(4), 665 (2019)
- IPCC: 2022: Transport. In *The Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Climate Change* (2022).
- Kazemi Miyangaskari, M., Mehrregan, M.R., Safari, H., Keivanpour, S., Dehghan, N.M.: Designing a closed-loop supply chain multi-objective optimization model with the aim of supplier selection and order allocation (case study: iran's retail stores for protein products). *Indus. Manag. Stud.* 21(69), 1–42 (2023)
- Keskin, F., Yıldız, M., Arslannur, B.: A data-driven framework for driving cycle generation and analysis. *Transp. Res. Rec.* 2679(1), 2199–2212 (2025)
- Khanfar, N.O., et al.: Driving behavior classification at signalized intersections using vehicle kinematics: application of unsupervised machine learning. *Int. J. Inj. Contr. Saf. Promot.* 30(1), 34–44 (2023)
- Khattak, A.J., Wali, B.: Analysis of volatility in driving regimes extracted from basic safety messages transmitted between connected vehicles. *Transp. Res. Part C Emerging Technol.* 84, 48–73 (2017)
- Knez, M., Muneer, T., Jereb, B., Cullinane, K.: The estimation of a driving cycle for Celje and a comparison to other European cities. *Sustain. Cities Soc.* 11, 56–60 (2014)
- Lee, G., Yeon, J., Kim, N., Park, S.: A comprehensive methodology for developing and evaluating driving cycles for electric vehicles using real-world data. *eTransportation* 1(24), 100409 (2025)
- Liu, J., Khattak, A., Han, L., Yuan, Q.: How much information is lost when sampling driving behavior data? Indicators to quantify the extent of information loss. *J. Intell. Connect. Vehicles* 3(1), 17–29 (2020)
- Liu, X., et al.: Study on driving cycle synthesis method for city buses considering random passenger load. *J. Adv. Transp.* 2020(1), 3871703 (2020)
- Ma, R., et al.: Real-world driving cycles and energy consumption informed by large-sized vehicle trajectory data. *J. Clean. Prod.* 223, 564–574 (2019)
- 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)
- Miyangaskary, M.K., Keivanpour, S., Lamghari, A.: "Optimizing closed-loop supply chain in the electric vehicle battery industry: a fully fuzzy approach." In *Product Lifecycle Management. Leveraging Digital Twins, Circular Economy, and Knowledge Management for Sustainable Innovation*, eds. Christophe Danjou et al. Cham: Springer Nature, 38–49 (2024).
- Miyangaskary, M.K., Keivanpour, S., Safari, H.: A multi-objective optimization model of a closed-loop supply chain for supplier selection and order allocation under uncertainty: a case study of retail stores for protein products. *Transp. Res. Proc.* 82, 323–341 (2025)
- Miyangaskary, M.K., et al.: Enhancing sustainability performance of a closed-loop supply chain for protein products using a fully fuzzy multi-objective optimisation. *Int. J. Syst. Sci. Operat. Logist.* 12(1), 2450608 (2025). <https://doi.org/10.1080/23302674.2025.2450608>
- Nouri, P.: "Enhancing the Gasoline Vehicles' CO₂ Emissions Estimation in Montreal." Polytechnique Montreal (2015).
- Pouresmaeili, M.A., Aghayan, I., Taghizadeh, S.A.: Development of mashhad driving cycle for passenger car to model vehicle exhaust emissions calibrated using on-board measurements. *Sustain. Cities Soc.* 36, 12–20 (2018)
- Quirama, L.F., Giraldo, M., Huertas, J.I., Jaller, M.: Driving cycles that reproduce driving patterns, energy consumptions and tailpipe emissions. *Transp. Res. Part D Transp. Environ.* 1(82), 102294 (2020)
- Ragione, L.D., Meccariello, G.: "The influence of road gradient in an integrated approach of real driving cycles and emissions factors model." In *Transportation Research Procedia*, Elsevier B.V., 3179–88 (2016).
- Rechkemmer, S.K., et al.: Utilization of smartphone data for driving cycle synthesis based on electric two-wheelers in Shanghai. *IEEE Trans. Intell. Transp. Syst.* 22(2), 876–886 (2021)
- Ribeiro, S.M.: *Imputation by Decomposition and by Time Series Pattern: Novel Imputation Methods for Missing Data in Time Series* (2021).

- Roy, F., Morency, C.: Comparing driving cycle development methods based on Markov chains. *Transp. Res. Rec.* 2675(3), 212–221 (2021)
- Sagberg, F.: Institute of Transport Economics. *Influencing Driver Behaviour and Safety by Road System Improvements*. <https://trid.trb.org/view/697621> (2003)
- Salihu, F., Demir, Y.K., Demir, H.G.: Effect of road slope on driving cycle parameters of urban roads. *Transp. Res. Part D Transp. Environ.* 118, 103676 (2023a)
- Salihu, F., Demir, Y.K., Demir, H.G.: Effect of road slope on driving cycle parameters of urban roads. *Transp. Res. Part D: Transp. Environ.* 118, 103676 (2023b)
- Shafiul, A.M., Bhaskar, A., Haque, M.M.: Driving behaviour modelling in the context of heterogeneous traffic and poor lane discipline conditions: the state of the art and beyond. *Transp. Transport Sci.* 18(3), 367–434 (2022)
- Supiyono, L.D., Wicaksono, A.: "Drivers behavior at signalized intersections with countdown timer reduces air pollution by gender." In *IOP Conference Series Earth and Environmental Science*, Institute of Physics (2022).
- Tharvin, R. et al.: "Development of driving cycle for passenger car under real world driving conditions in Kuala Lumpur, Malaysia." In *IOP Conference Series Materials Science and Engineering*, Institute of Physics Publishing (2018).
- Tong, H.Y., Ng, K.: Development of bus driving cycles using a cost effective data collection approach. *Sustain. Cities Soc.* 1(69), 102854 (2021)
- Tong, H.Y., Ng, K.W.: Developing electric bus driving cycles with significant road gradient changes: a case study in Hong Kong. *Sustain. Cities Soc.* 1(98), 104819 (2023)
- Wolfermann, A., Alhajyaseen, W.K.M., Nakamura, H.: "Modeling Speed Profiles of Turning Vehicles at Signalized Intersections." 2004_Velocity_Planning (September 2016): 1–17 (2011).
- Xu, N., Liu, Q., Zhang, Z., Li, J.: Development of multiple driving cycles with equivalent energy consumption to expand a standard driving cycle. *Energy* 319, 134951 (2025)
- Yaghoubi, R., Mahboubi, A., Miyangaskary, M.K., Keivanpour, S.: Developing a Robust Multi-objective optimization model for reverse logistics of electric vehicle batteries. In: *IFIP International Conference on Advances in Production Management Systems* (pp. 492–507). Cham: Springer Nature Switzerland (2024).
- Yang, Z., Muncrief, R., Bandivadekar, A.: *Global Baseline Assessment of Compliance and Enforcement Programs for Vehicle Emissions and Energy Efficiency*. www.theicct.org (2017).
- Yang, Y., Li, T., Zhang, T., Yu, Q.: Time dimension analysis: comparison of nanjing local driving cycles in 2009 and 2017. *Sustain. Cities Soc.* 1(53), 101949 (2020)
- Yang, Y. et al.: "Markov chain-based approach of the driving cycle development for electric vehicle application." In *Energy Procedia*, Elsevier Ltd, pp. 502–7 (2018).
- Yang, Y., et al.: A novel heavy-duty truck driving cycle construction framework based on big data. *Transp. Res. Part D Transp. Environ.* 127, 104077 (2024)
- Yarahmadi, A., Morency, C., Trepanier, M.: Identifying optimal number of driving cycles to represent diverse driving conditions. *Int. J. Sustain. Transp.* 18(8), 704–726 (2024). <https://doi.org/10.1080/15568318.2024.2397647>
- Yarahmadi, A., Morency, C., Trepanier, M.: "Identifying optimal number of driving cycles in different driving conditions using machine learning." In *World Conference on Transport Research—WCTR 2023*, www.sciencedirect.com / www.elsevier.com/locate/procedia 2352–1465 (2023a).
- Yarahmadi, A., Morency, C., Trepanier, M.: New data-driven approach to generate typologies of road segments. *Transport. A Transport Sci.* 20(2), 2163206 (2024b)
- Yuhui, P., Yuan, Z., Huibao, Y.: Development of a representative driving cycle for urban buses based on the K-means cluster method. *Clust. Comput.* 22, 6871–6880 (2018)
- Zhang, L., et al.: Road type-based driving cycle development and application to estimate vehicle emissions for passenger cars in Guangzhou. *Atmos. Pollut. Res.* 12(8), 101138 (2021)